

# Modeliranje uspjeha studenata rudarenjem podataka iz hibridnog okruženja za učenje uz primjenu analize vremenskih nizova u frekvencijskoj domeni

---

Gamulin, Jasna

Doctoral thesis / Disertacija

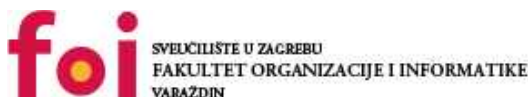
2018

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Organization and Informatics Varaždin / Sveučilište u Zagrebu, Fakultet organizacije i informatike Varaždin**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:211:117290>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-12-26**



Repository / Repozitorij:

[Faculty of Organization and Informatics - Digital Repository](#)





Sveučilište u Zagrebu

Fakultet organizacije i informatike

Jasna Gamulin

**MODELIRANJE USPJEHA STUDENATA  
RUDARENJEM PODATAKA IZ HIBRIDNOG  
OKRUŽENJA ZA UČENJE UZ PRIMJENU  
ANALIZE VREMENSKIH NIZOVA U  
FREKVENCIJSKOJ DOMENI**

DOKTORSKI RAD

Varaždin, 2018.

# PODACI O DOKTORSKOM RADU

## I. AUTOR

Ime i prezime	Jasna Gamulin
Datum i mjesto rođenja	29. rujna 1964., Zagreb
Naziv fakulteta i datum diplomiranja na VII/I stupnju	Ekonomski fakultet Sveučilišta u Zagrebu, 15. veljače 1989.
Naziv fakulteta i datum diplomiranja na VII/II stupnju	
Sadašnje zaposlenje	Medicinski fakultet Sveučilišta u Zagrebu

## II. DOKTORSKI RAD

Naslov	Modeliranje uspjeha studenata rudarenjem podataka iz hibridnog okruženja za učenje uz primjenu analize vremenskih nizova u frekvencijskoj domeni
Broj stranica, slika, tabela, priloga, bibliografskih podataka	128 stranica, 32 slike, 124 tablice, 59 stranica priloga, 105 bibliografskih izvora
Znanstveno područje i polje iz kojeg je postignut doktorat znanosti	Društvene znanosti, Informacijske i komunikacijske znanosti
Mentori ili voditelji rada	Prof. dr. sc. Dragutin Kermek Prof. dr. sc. Diana Šimić
Fakultet na kojem je obranjen doktorski rad	Fakultet organizacije i informatike Sveučilišta u Zagrebu
Oznaka i redni broj rada	142

## III. OCJENA I OBRANA

Datum sjednice Fakultetskog vijeća na kojoj je prihvaćena tema	20. srpnja 2016.
Datum predaje rada	26. veljače 2018.
Datum sjednice Fakultetskog vijeća na kojoj je prihvaćena pozitivna ocjena rada	19. lipnja 2018.
Sastav povjerenstva koje je rad ocijenilo	Izv. prof. dr. sc. Jasminka Dobša, predsjednica Prof. dr. sc. Marijan Krašna, član Izv. prof. dr.sc. Nina Begičević Ređep, članica
Datum obrane doktorskog rada	10. srpnja 2018.
Sastav povjerenstva pred kojim je rad obranjen	Izv. prof. dr. sc. Jasminka Dobša, predsjednica Prof. dr. sc. Marijan Krašna, član Izv. prof. dr.sc. Nina Begičević Ređep, članica
Datum promocije	



Sveučilište u Zagrebu

Fakultet organizacije i informatike

Jasna Gamulin

**MODELIRANJE USPJEHA STUDENATA  
RUDARENJEM PODATAKA IZ HIBRIDNOG  
OKRUŽENJA ZA UČENJE UZ PRIMJENU  
ANALIZE VREMENSKIH NIZOVA U  
FREKVENCIJSKOJ DOMENI**

DOKTORSKI RAD

Mentori: Prof. dr. sc. Dragutin Kermek

Prof. dr. sc. Diana Šimić

Varaždin, 2018.



University of Zagreb

Faculty of Organization and Informatics

Jasna Gamulin

**MODELING STUDENT SUCCESS THROUGH  
DATA MINING ON BLENDED LEARNING  
ENVIRONMENT DATA USING FREQUENCY  
DOMAIN TIME SERIES ANALYSIS**

DOCTORAL DISSERTATION

Supervisors: Professor Dragutin Kermek, Ph.D.  
Professor Diana Šimić, Ph.D.

Varaždin, 2018

## **ZAHVALE**

Zahvaljujem mojim mentorima - prof. dr. sc. Dragutinu Kermeku i prof. dr. sc. Diani Šimić na vrijednim smjernicama, na njihovom vremenu i trudu uloženom u stvaranje ovog rada.

Zahvaljujem dekanu Fakulteta i dugogodišnjem voditelju Studija - prof. dr. sc. Nevenu Vrčeku na razumijevanju, brizi i pomoći.

Zahvaljujem mužu - Ozrenu na ljubavi i potpori tijekom ovog dugog niza naših zajedničkih godina, na svim trenucima kada smo se zajedno smijali i plakali.

Zahvaljujem mami - Slavici, na djelotvornoj pomoći i ljubavi koju mi, bez puno riječi, tiho i nenametljivo cijeli život pruža.

Zahvaljujem tati, Josipu †, koji mi je uvijek bio i ostao nadahnuće.

Zahvaljujem djeci koja su nam darovana – Lovri, Luki, Luciji, Franu na ljubavi i toplini koju od njih primam i koja me pokreće.

## Sažetak

Svaki pristup studenta sistemu za upravljanje učenjem (engl. *learning management system*, LMS) generira podatke o tome tko je pristupio, kada, s koje IP (engl. *Internet protocol*, IP) adrese i koja je aktivnost provedena. Na taj način prikuplja se mnoštvo podataka o studentima i to tako da se studente pri tom ne ometa u procesu učenja. Dostupni podaci mogu se iskoristiti i tako da se pristupi studenata određenom predmetu na LMS-u promatraju kao vremenski nizovi. U ovom radu pokazano je da promatrani vremenski nizovi nose informaciju o ponašanju studenata koja je indikativna za njihov uspjeh na završnom ispitu. Primjenom Fourierove transformacije (FT) na kreirane vremenske nizove, tj. prebacivanjem vremenskih nizova u frekvencijsku domenu, otvorena je mogućnost da se količina podataka smanji, a proces obrade podataka postane operativniji. Drugi razlog primjene FT je otkriti skrivene periodičnosti koje su ključ modeliranja i predikcije, a treći razlog je smanjenje šuma te postizanje bolje točnosti predikcije uspjeha studenata.

Hipoteze ovog rada ispitane su na skupovima podataka o studentima upisanim na 2 fakulteta, na ukupno 4 različita predmeta tijekom dvije akademske godine. Uzorak se sastojao od podataka o približno 1400 studenata.

U ovom istraživanju je pokazano da je moguće kreirati model predikcije uspjeha studenata rudarenjem podataka iz hibridnog okruženja za učenje primjenom analize vremenskih nizova pristupa studenata LMS-u u frekvencijskoj domeni te da takav model postiže točnost predikcije koja je jednaka ili veća od točnosti predikcije pomoću sličnih modela opisanih u literaturi. Razvijena je tehnika za sažimanje vremenskih nizova pristupa studenata LMS-u u frekvencijskoj domeni uz uklanjanje šuma, što znači da je korištenjem smanjenog broja koeficijenata postignuta bolja točnost predikcije u odnosu na modele temeljene na punom skupu podataka. Nadalje, pokazano je da je moguće postići točnost predikcije uspješnosti koja je veća od 75%, a temeljem podataka o prve 2/3 ukupnog trajanja predmeta.

Temeljni doprinos rada je primjena spektralne analize, odnosno analize vremenskih nizova u frekvencijskoj domeni na skup podataka iz hibridnog okruženja za učenje.

**Ključne riječi:** hibridno okruženje za učenje, analiza vremenskih nizova, LMS logovi, frekvencijska domena, rudarenje edukacijskih podataka

## **Abstract**

Each student log to the learning management system (LMS) generates data on who accessed the LMS, when, from which IP address, and which activity is performed. It is the way to collect the large amount of data on students and not to disturb the process of learning while the data are collected. The available data could be also used in the way that student logs to a certain course on the LMS are regarded as time series. In this paper it is proven that the observed time series carry the information on student behavior which is indicative for the final exam success. By applying the Fourier transform (FT) to the created time series, i.e. by switching time series to the frequency domain, the opportunity occurs to reduce the amount of data and for the data processing to become more operational. The second reason to apply FT is to reveal the hidden periodicities which are the key to modeling and prediction, and the third reason is to reduce the noise and to achieve better accuracy of student success prediction.

The hypotheses of this study are tested on data sets on students enrolled in two schools, in four different courses, and during two academic years. The sample consists of data on approximately 1,400 students.

This research has demonstrated that it is possible to create the student success prediction model through data mining on blended learning environment data using frequency domain time series analysis of student logs to LMS. The proposed model achieves the prediction accuracy which is equal or bigger of the prediction accuracies achieved by similar models described in the literature. The technique for compressing time series of student logs to LMS is developed. The technique removes the noise, meaning that using the reduced number of coefficients gives better prediction accuracy than using models based on the full amount of data. Furthermore, it is proven that it is possible to achieve the success prediction with the accuracy more than 75% based on data for the first 2/3 of the total length of the course.

The fundamental contribution of the paper is the application of spectral analysis, i.e. analysis of time series in the frequency domain to the data set from the hybrid learning environment.

**Keywords:** blended learning environment, time series analysis, LMS logs, frequency domain, educational data mining



## SADRŽAJ

Popis slika .....	X
Popis tablica .....	XIV
Popis upotrijebljenih kratica .....	XXVI
1. Uvod.....	1
1.1. Problem istraživanja .....	1
1.2. Ciljevi istraživanja.....	3
1.3. Hipoteze istraživanja .....	4
1.4. Metodologija istraživanja .....	4
1.5. Struktura rada .....	7
2. Pregled prethodnih istraživanja.....	9
2.1. Rudarenje edukacijskih podataka .....	9
2.2. Studentski modeli .....	12
2.3. Studentski modeli s vremenskom segmentacijom atributa.....	12
3. Skup podataka i metodološki okvir.....	16
3.1. Skup podataka.....	16
3.2. Vremenski nizovi.....	17
3.3. Vrijeme kumulacije .....	19
3.4. Fourierova transformacija.....	19
3.5. Iz vremenske u frekvencijsku domenu .....	20
3.6. Sažimanje podataka i uklanjanje šuma .....	20
3.7. Klasifikacija.....	21
3.8. Analiza glavnih komponenti.....	21
4. Modeliranje .....	23
4.1. Prethodna obrada podataka.....	23
4.2. Algoritmi i softver .....	32
4.2.1. Umjetna neuronska mreža.....	34
4.2.2. Metoda potpornih vektora .....	36
4.2.3. Metoda $k$ najbližih susjeda .....	38
4.2.4. Naivni Bayesov klasifikator .....	39

4.3.	Model.....	40
4.4.	Evaluacija .....	44
4.4.1.	Točnost.....	44
4.4.2.	Vrednovanje .....	46
4.4.3.	Test statističke značajnosti .....	46
5.	Rezultati .....	48
5.1.	Prethodni rezultati.....	48
5.2.	Modeliranje korištenjem podataka u frekvencijskoj domeni .....	50
5.2.1.	Usporedba točnosti predikcije po algoritmima .....	52
5.2.2.	Usporedba točnosti predikcije po dužini promatranog vremenskog niza .....	60
5.2.3.	Usporedba točnosti predikcije po vremenu kumulacije .....	64
5.3.	Rezultati testiranja hipoteza.....	69
5.3.1.	Prva hipoteza.....	69
5.3.2.	Druga hipoteza .....	78
5.3.3.	Treća hipoteza .....	106
6.	Sinteza rezultata .....	109
6.1.	Sinteza rezultata testiranja hipoteza .....	109
6.2.	Doprinosi i ograničenja istraživanja.....	114
7.	Etički aspekti istraživanja .....	117
8.	Zaključak.....	117
9.	Smjernice za buduća istraživanja.....	118
	Literatura.....	120
	Prilozi.....	129
	Životopis .....	185
	Popis radova.....	186

## POPIS SLIKA

Slika 1: Tablica iz LMS-ove baze podataka koja sadrži podatke o svim studentima i njihovim logovima na određeni predmet.....	5
Slika 2: Jednostavni ciklički model.....	18
Slika 3: Vremenski niz logova za određenog studenta.....	29
Slika 4: Primjerak periodograma, odnosno Fourierovih koeficijenata za jednog studenta.....	30
Slika 5: Topologija modela umjetne neuronske mreže.....	34
Slika 6: Umjetna neuronska mreža s propagacijom unaprijed i skrivenim slojem.....	35
Slika 7: Temeljni pojmovi za SVM algoritam – granica, margina i kazna.....	36
Slika 8: Klase koje se ne mogu linearno razdvojiti.....	37
Slika 9: Transformacija u klase koje se mogu linearno razdvojiti.....	37
Slika 10: Osnovni proces .....	41
Slika 11: Petlja za računanje modela s postupnim povećavanjem broja atributa koje uzima u obzir.....	41
Slika 12: Postupak modeliranja i validacije unutar operatora za optimizaciju.....	42
Slika 13: Validacija .....	42
Slika 14: Broj klikova na LMS i datumi kolokvija i ispita .....	48
Slika 15: PC prikaz faktorskih bodova PC1 i PC2 komponente vremenskog niza logova za studente upisane u predmet Napredne WEB tehnologije i servisi. Uspješnost studenata u 3 kategorije: vrlo dobri (zelene zvjezdice), prosječni (plavi kvadratići) i loši (crveni trokutići). Elipse na slici definiraju granice pouzdanosti od 95%.....	50

Slika 16: a) vremenski nizovi logova za dva slučajno odabrana studenta upisana na predmet NWTIS <sub>sat</sub> 2012; b) DFT periodogrami izračunati iz vremenskih nizova prikazanih na slici a). Zbog bolje preglednosti prikazana je samo donja polovina periodograma.....	51
Slika 17: Točnost za punu dužinu predmeta, vrijeme kumulacije 1 sat: crvene linije – točnost za smanjeni broj koeficijenata; crne linije - točnost za puni broj koeficijenata (pune linije - 2 kategorije; isprekidane linije - 3 kategorije).....	81
Slika 18: Broj koeficijenata za punu dužinu predmeta, vrijeme kumulacije 1 sat: crvene linije – smanjeni broj koeficijenata; crne linije - puni broj koeficijenata (pune linije - 2 kategorije; isprekidane linije - 3 kategorije).....	82
Slika 19: Točnost za punu dužinu predmeta, vrijeme kumulacije 1 dan: crvene linije – točnost za smanjeni broj koeficijenata; crne linije - točnost za puni broj koeficijenata (pune linije - 2 kategorije; isprekidane linije - 3 kategorije).....	85
Slika 20: Broj koeficijenata za punu dužinu predmeta, vrijeme kumulacije 1 dan: crvene linije – smanjeni broj koeficijenata; crne linije - puni broj koeficijenata (pune linije - 2 kategorije; isprekidane linije - 3 kategorije).....	86
Slika 21: Usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata, puna dužina trajanja predmeta, po promatranim algoritmima, vrijeme kumulacije 1 sat, 2 kategorije.....	87
Slika 22: Usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata, puna dužina trajanja predmeta, po promatranim algoritmima, vrijeme kumulacije 1 sat, 3 kategorije.....	88
Slika 23: Usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata, puna dužina trajanja predmeta, po promatranim algoritmima, vrijeme kumulacije 1 dan, 2 kategorije.....	88

Slika 24: Usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata, puna dužina trajanja predmeta, po promatranim algoritmima, vrijeme kumulacije 1 dan, 3 kategorije.....	89
Slika 25: Točnost za prve 2/3 trajanja predmeta, vrijeme kumulacije 1 sat: crvene linije – točnost za smanjeni broj koeficijenata; crne linije - točnost za puni broj koeficijenata (pune linije - 2 kategorije; isprekidane linije - 3 kategorije).....	95
Slika 26: Broj koeficijenata za prve 2/3 trajanja predmeta, vrijeme kumulacije 1 sat: crvene linije – smanjeni broj koeficijenata; crne linije - puni broj koeficijenata (pune linije - 2 kategorije; isprekidane linije - 3 kategorije).....	96
Slika 27: Točnost za prve 2/3 trajanja predmeta, vrijeme kumulacije 1 dan: crvene linije – točnost za smanjeni broj koeficijenata; crne linije - točnost za puni broj koeficijenata (pune linije - 2 kategorije; isprekidane linije - 3 kategorije).....	99
Slika 28: Broj koeficijenata za prve 2/3 trajanja predmeta, vrijeme kumulacije 1 dan: crvene linije – smanjeni broj koeficijenata; crne linije - puni broj koeficijenata (pune linije - 2 kategorije; isprekidane linije - 3 kategorije).....	100
Slika 29: Usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata, prve 2/3 trajanja predmeta, po promatranim algoritmima, vrijeme kumulacije 1 sat, 2 kategorije.....	101
Slika 30: Usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata, prve 2/3 trajanja predmeta, po promatranim algoritmima, vrijeme kumulacije 1 sat, 3 kategorije.....	102
Slika 31: Usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata, prve 2/3 trajanja predmeta, po promatranim algoritmima, vrijeme kumulacije 1 dan, 2 kategorije.....	102

Slika 32: Usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata, prve 2/3 trajanja predmeta, po promatranim algoritmima, vrijeme kumulacije 1 dan, 3 kategorije.....103

## POPIS TABLICA

Tablica 1: Pregled ispitanih predmeta i inicijalnih skupova podataka za akademsku godinu 2012./2013.....	16
Tablica 2: Pregled ispitanih predmeta i inicijalnih skupova podataka za akademsku godinu 2013./2014.....	17
Tablica 3: Usporedni podaci o elementima i aktivnostima na promatranim predmetima, njihovom broju u broju dnevničkih zapisa za akademsku godinu 2012./2013.....	25
Tablica 4: Usporedni podaci o elementima i aktivnostima na promatranim predmetima, njihovom broju u broju dnevničkih zapisa za akademsku godinu 2013./2014.....	26
Tablica 5: Matrica zabune za 2 klase.....	44
Tablica 6: Točnost - 4 algoritma, zavisna varijabla diskretizirana na 2 kategorije, vrijeme kumulacije je 1 sat, na vremenskom nizu pune dužine predmeta, primijenjena tehnika sažimanja .....	52
Tablica 7: Točnost - 4 algoritma, zavisna varijabla diskretizirana na 2 kategorije, vrijeme kumulacije je 1 dan, na vremenskom nizu pune dužine predmeta, primijenjena tehnika sažimanja .....	53
Tablica 8: Točnost - 3 algoritma, zavisna varijabla diskretizirana na 3 kategorije, vrijeme kumulacije je 1 sat, na vremenskom nizu pune dužine predmeta, primijenjena tehnika sažimanja.....	54
Tablica 9: Točnost - 3 algoritma, zavisna varijabla diskretizirana na 3 kategorije, vrijeme kumulacije je 1 dan, na vremenskom nizu pune dužine predmeta, primijenjena tehnika sažimanja.....	55
Tablica 10: Točnost - 4 algoritma, zavisna varijabla diskretizirana na 2 kategorije, vrijeme kumulacije je 1 sat, na vremenskom nizu za 2/3 trajanja predmeta, primijenjena tehnika sažimanja.....	56

Tablica 11: Točnost - 4 algoritma, zavisna varijabla diskretizirana na 2 kategorije, vrijeme kumulacije je 1 dan, na vremenskom nizu za 2/3 trajanja predmeta, primijenjena tehnika sažimanja.....	57
Tablica 12: Točnost - 3 algoritma, zavisna varijabla diskretizirana na 3 kategorije, vrijeme kumulacije je 1 sat, na vremenskom nizu za 2/3 trajanja predmeta, primijenjena tehnika sažimanja.....	58
Tablica 13: Točnost - 3 algoritma, zavisna varijabla diskretizirana na 3 kategorije, vrijeme kumulacije je 1 dan, na vremenskom nizu za 2/3 trajanja predmeta, primijenjena tehnika sažimanja.....	59
Tablica 14: Točnosti predikcije algoritma ANN za predmete koji traju jedan semestar, dvije kategorije, vrijeme kumulacije 1 sat – za različite dužine vremenskih nizova, primijenjena tehnika sažimanja.....	60
Tablica 15: Točnosti predikcije algoritma ANN za predmete koji traju jedan semestar, dvije kategorije, vrijeme kumulacije 1 dan – za različite dužine vremenskih nizova, primijenjena tehnika sažimanja.....	61
Tablica 16: Točnosti predikcije algoritma ANN za predmete koji traju jedan semestar, tri kategorije, vrijeme kumulacije 1 sat – za različite dužine vremenskih nizova, primijenjena tehnika sažimanja.....	61
Tablica 17: Točnosti predikcije algoritma ANN za predmete koji traju jedan semestar, tri kategorije, vrijeme kumulacije 1 dan – za različite dužine vremenskih nizova, primijenjena tehnika sažimanja.....	62
Tablica 18: Točnosti predikcije algoritma ANN za predmet koji traje jedan tjedan, dvije kategorije, vrijeme kumulacije 1 sat – za različite dužine vremenskih nizova, primijenjena tehnika sažimanja.....	62
Tablica 19: Usporedba točnosti predikcije po dužini promatranog vremenskog niza i grupama skupova podataka, p-vrijednost Friedmanovog testa (bijela polja=nema razlike, siva polja=ima razlike).....	63



Tablica 20: Točnost predikcije po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 sat, 2 kategorije, primijenjena tehnika sažimanja.....	64
Tablica 21: Točnost predikcije po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 sat, 3 kategorije, primijenjena tehnika sažimanja.....	64
Tablica 22: Točnost predikcije po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 dan, 2 kategorije, primijenjena tehnika sažimanja.....	65
Tablica 23: Točnost predikcije po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 dan, 3 kategorije, primijenjena tehnika sažimanja.....	65
Tablica 24: Rezultati Wilcoxonovog testa usklađenih parova – usporedba točnosti po algoritmima između rezultata temeljenih na skupovima podataka s vremenom kumulacije 1 sat i 1 dan, puna dužina trajanja predmeta, p-vrijednost (bijelo polje=nema razlike, sivo polje=ima razlike).....	66
Tablica 25: Točnost predikcije po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 sat, 2 kategorije, primijenjena tehnika sažimanja.....	66
Tablica 26: Točnost predikcije po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 sat, 3 kategorije, primijenjena tehnika sažimanja.....	67
Tablica 27: Točnost predikcije po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 dan, 2 kategorije, primijenjena tehnika sažimanja.....	67
Tablica 28: Točnost predikcije po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 dan, 3 kategorije, primijenjena tehnika sažimanja.....	67
Tablica 29: Rezultati Wilcoxonovog testa usklađenih parova – usporedba točnosti po algoritmima između rezultata temeljenih na skupovima podataka s vremenom kumulacije 1 sat i 1 dan, prve 2/3 trajanja predmeta, p-vrijednost (bijelo polje=nema razlike, sivo polje=ima razlike).....	68
Tablica 30: Točnosti predikcije za modele opisane u ovom radu, a temeljem podataka u punoj dužini trajanja sva četiri predmeta, zavisna varijabla diskretizirana na 2 kategorije, po	

algoritmima, za 2012. godinu, primijenjena tehnika sažimanja. Crveno je označena najniža vrijednost, a zeleno najviša vrijednost za pojedini algoritam.....	70
Tablica 31: Točnost i broj koeficijenata po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 sat, 2 kategorije, primijenjena tehnika sažimanja.....	78
Tablica 32: Točnost i broj koeficijenata po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 sat, 2 kategorije, pun broj koeficijenata.....	79
Tablica 33: Točnost i broj koeficijenata po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 sat, 3 kategorije, primijenjena tehnika sažimanja.....	79
Tablica 34: Točnost i broj koeficijenata po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 sat, 3 kategorije, pun broj koeficijenata.....	80
Tablica 35: Točnost i broj koeficijenata po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 dan, 2 kategorije, primijenjena tehnika sažimanja.....	83
Tablica 36: Točnost i broj koeficijenata po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 dan, 2 kategorije, pun broj koeficijenata.....	83
Tablica 37: Točnost i broj koeficijenata po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 dan, 3 kategorije, primijenjena tehnika sažimanja.....	84
Tablica 38: Točnost i broj koeficijenata po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 dan, 3 kategorije, pun broj koeficijenata.....	84
Tablica 39: Rezultati Wilcoxonovog testa usklađenih parova – usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova za sve skupove podataka u punoj dužini trajanja predmeta i po promatranim algoritmima s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata (0=nema razlike, 1=ima razlike).....	87
Tablica 40: Usporedba točnosti (1=točnost uz smanjen broj koeficijenata je veća od točnosti uz puni broj koeficijenata, 0=točnost uz smanjen broj koeficijenata je manja od točnosti uz puni broj koeficijenata).....	91

Tablica 41: Točnost i broj koeficijenata po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 sat, 2 kategorije, primijenjena tehnika sažimanja.....	92
Tablica 42: Točnost i broj koeficijenata po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 sat, 2 kategorije, pun broj koeficijenata.....	93
Tablica 43: Točnost i broj koeficijenata po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 sat, 3 kategorije, primijenjena tehnika sažimanja.....	93
Tablica 44: Točnost i broj koeficijenata po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 sat, 3 kategorije, pun broj koeficijenata.....	94
Tablica 45: Točnost i broj koeficijenata po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 dan, 2 kategorije, primijenjena tehnika sažimanja.....	97
Tablica 46: Točnost i broj koeficijenata po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 dan, 2 kategorije, pun broj koeficijenata.....	97
Tablica 47: Točnost i broj koeficijenata po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 dan, 3 kategorije, primijenjena tehnika sažimanja.....	97
Tablica 48: Točnost i broj koeficijenata po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 dan, 3 kategorije, pun broj koeficijenata.....	98
Tablica 49: Rezultati Wilcoxonovog testa usklađenih parova – usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova za sve skupove podataka, za prve 2/3 trajanja predmeta i po promatranim algoritmima s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata (0=nema razlike, 1=ima razlike).....	101
Tablica 50: Usporedba točnosti (1=točnost uz smanjen broj koeficijenata je veća od točnosti uz puni broj koeficijenata, 0=točnost uz smanjen broj koeficijenata je manja od točnosti uz puni broj koeficijenata), za modele temeljene na podacima o prve 2/3 trajanja predmeta.....	105

Tablica 51: Točnost predikcije za modele temeljene na podacima o prve 2/3 trajanja predmeta, primijenjena tehnika sažimanja.....	107
Tablica 52: Točnosti - ANN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	129
Tablica 53: Minimalan broj koeficijenata za maksimalnu točnost - ANN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	129
Tablica 54: Točnosti - ANN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova – puni broj koeficijenata.....	129
Tablica 55: Puni broj koeficijenata - ANN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova.....	130
Tablica 56: Točnosti - ANN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	130
Tablica 57: Minimalan broj koeficijenata za maksimalnu točnost - ANN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	130
Tablica 58: Točnosti - ANN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - puni broj koeficijenata.....	131
Tablica 59: Puni broj koeficijenata - ANN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova.....	131
Tablica 60: Točnosti - ANN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	131
Tablica 61: Minimalan broj koeficijenata za maksimalnu točnost - ANN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	132

Tablica 62: Točnosti - ANN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata.....	132
Tablica 63: Puni broj koeficijenata - ANN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova.....	132
Tablica 64: Točnosti - ANN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	133
Tablica 65: Minimalan broj koeficijenata za maksimalnu točnost - ANN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	133
Tablica 66: Točnosti - ANN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - puni broj koeficijenata.....	133
Tablica 67: Puni broj koeficijenata - ANN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova.....	134
Tablica 68: Točnosti - ANN – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	134
Tablica 69: Minimalan broj koeficijenata za maksimalnu točnost - ANN – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	134
Tablica 70: Točnosti - ANN – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata.....	134
Tablica 71: Puni broj koeficijenata - ANN – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova.....	135
Tablica 72: Točnosti - SVM – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	135

Tablica 73: Minimalan broj koeficijenata za maksimalnu točnost - SVM – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	135
Tablica 74: Točnosti - SVM – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata.....	136
Tablica 75: Puni broj koeficijenata - SVM – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova.....	136
Tablica 76: Točnosti - SVM – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	136
Tablica 77: Minimalan broj koeficijenata za maksimalnu točnost - SVM – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	137
Tablica 78: Točnosti - SVM – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - puni broj koeficijenata.....	137
Tablica 79: Puni broj koeficijenata - SVM – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova.....	137
Tablica 80: Točnosti - SVM – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	138
Tablica 81: Minimalan broj koeficijenata za maksimalnu točnost - SVM – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	138
Tablica 82: Točnosti - SVM – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata.....	138
Tablica 83: Puni broj koeficijenata - SVM – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova.....	138

Tablica 84: Točnosti - kNN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	139
Tablica 85: Minimalan broj koeficijenata za maksimalnu točnost - kNN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	139
Tablica 86: Točnosti - kNN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata.....	139
Tablica 87: Puni broj koeficijenata - kNN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova.....	140
Tablica 88: Točnosti - kNN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	140
Tablica 89: Minimalan broj koeficijenata za maksimalnu točnost - kNN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	140
Tablica 90: Točnosti - kNN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - puni broj koeficijenata.....	141
Tablica 91: Puni broj koeficijenata - kNN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova.....	141
Tablica 92: Točnosti - kNN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	141
Tablica 93: Minimalan broj koeficijenata za maksimalnu točnost - kNN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	142
Tablica 94: Točnosti - kNN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata.....	142

Tablica 95: Puni broj koeficijenata - kNN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova.....	142
Tablica 96: Točnosti - kNN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	143
Tablica 97: Minimalan broj koeficijenata za maksimalnu točnost - kNN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	143
Tablica 98: Točnosti - kNN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - puni broj koeficijenata.....	143
Tablica 99: Puni broj koeficijenata - kNN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova.....	144
Tablica 100: Točnosti - kNN – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	144
Tablica 101: Minimalan broj koeficijenata za maksimalnu točnost - kNN – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	144
Tablica 102: Točnosti - kNN – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata.....	144
Tablica 103: Puni broj koeficijenata - kNN – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova.....	145
Tablica 104: Točnosti - NB – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	145
Tablica 105: Minimalan broj koeficijenata za maksimalnu točnost - NB – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	145



Tablica 106: Točnosti - NB – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata.....	146
Tablica 107: Puni broj koeficijenata - NB – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova.....	146
Tablica 108: Točnosti - NB – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	146
Tablica 109: Minimalan broj koeficijenata za maksimalnu točnost - NB – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	147
Tablica 110: Točnosti - NB – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - puni broj koeficijenata.....	147
Tablica 111: Puni broj koeficijenata - NB – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova.....	147
Tablica 112: Točnosti - NB – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	148
Tablica 113: Minimalan broj koeficijenata za maksimalnu točnost - NB – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	148
Tablica 114: Točnosti - NB – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata.....	148
Tablica 115: Puni broj koeficijenata - NB – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova.....	149
Tablica 116: Točnosti - NB – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	149

Tablica 117: Minimalan broj koeficijenata za maksimalnu točnost - NB – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	149
Tablica 118: Točnosti - NB – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - puni broj koeficijenata.....	150
Tablica 119: Puni broj koeficijenata - NB – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova.....	150
Tablica 120: Točnosti - NB – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	150
Tablica 121: Minimalan broj koeficijenata za maksimalnu točnost - NB – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja.....	150
Tablica 122: Točnosti - NB – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata.....	151
Tablica 123: Puni broj koeficijenata - NB – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova.....	151
Tablica 124: Usporedba točnosti predikcije za različite dužine vremenskih nizova (primijenjena tehnika sažimanja) po algoritmima i grupama predmeta – Friedmanov test, sumarni rezultati u tablici 19.....	152

## Popis upotrijebljenih kratica

Ac	točnost (engl. <i>accuracy</i> )
ANN	umjetna neuronska mreža (engl. <i>artificial neural network</i> )
ARIMA	autoregresivni integrirani pomični prosjek (engl. <i>autoregressive integrated moving average</i> )
ARMA	autoregresivni pomični prosjek (engl. <i>autoregressive moving average</i> )
AUC	površina ispod ROC krivulje (engl. <i>area under the ROC curve</i> )
BP	povratna propagacija pogreške (engl. <i>back propagation</i> )
b-učenje	hibridno/ mješovito okruženje za učenje (engl. <i>blended learning/ b-learning</i> )
CAD	program (engl. <i>computer aided design</i> )
CBA	ispitivanje potpomognuto računalom (engl. <i>computer based assessment</i> )
CMC	kombinacija više klasifikatora (engl. <i>combination of multiple classifiers, CMC</i> )
CRISP-DM	metoda (engl. <i>cross-industry standard process for data mining</i> )
DFT	diskretna Fourierova transformacija (engl. <i>discrete Fourier transform</i> )
EDM	edukacijsko rudarenje podataka/ dubinska analiza edukacijskih podataka (engl. <i>educational data mining</i> )
e-učenje	elektroničko učenje/ online učenje (engl. <i>e-learning</i> )
FFT	brza Fourierova transformacija (engl. <i>fast Fourier transform</i> )
FIB	Fizika i biofizika
FMD	Fizika medicinske dijagnostike
FT	Fourierova transformacija
IKT	informacijske i komunikacijske tehnologije
kNN	metoda k najbližih susjeda (engl. <i>k-nearest neighbors</i> )
LAK	analitika učenja i znanja (engl. <i>learning analytics and knowledge</i> )
LMS	sistem za upravljanje učenjem (engl. <i>learning management system</i> )
LON-CAPA	engl. <i>learning online network with computer-assisted personalized approach</i>
MIL	učenje pomoću više instanci (engl. <i>multiple instance learning</i> )
MMLA	više modalna analitika učenja (engl. <i>multimodal learning analytics</i> )

Moodle	dinamičko okruženje za učenje orijentirano na modularne objekte (engl. <i>modular object-oriented dynamic learning environment</i> )
NB	naivni Bayesov klasifikator (engl. <i>naïve Bayes</i> )
NN	neuronska mreža (engl. <i>neural network</i> )
NWTIS	Napredne WEB tehnologije i servisi
PCA	analiza glavnih komponenti (engl. <i>principal component analysis</i> )
Pr	preciznost (engl. <i>precision</i> )
QoI	kvaliteta interakcije (engl. <i>quality of interaction</i> )
Re	opoziv (engl. <i>recall</i> )
ROC	krivulja (engl. <i>receiver operating characteristic curve</i> )
SEDM	sistem za edukacijsko rudarenje podataka (engl. <i>system for educational data mining</i> )
SVM	metoda potpornih vektora (engl. <i>support vector machines</i> )
Sp	specifičnost (engl. <i>specificity</i> )
WDIP	WEB dizajn i programiranje

## 1. UVOD

Uvodni dio sadrži prikaz problema istraživanja, ciljeve i hipoteze istraživanja, korištenu metodologiju te strukturu rada.

### 1.1. Problem istraživanja

Hibridno (mješovito, engl. *blended learning/ b-learning*) okruženje za učenje rašireno je u visokom obrazovanju u Hrvatskoj i svijetu, posebice u Europi. U Hrvatskoj je hibridno učenje vrlo učestali i prihvaćeni način primjene informacijskih i komunikacijskih tehnologija (IKT) u visokom obrazovanju ako sudimo prema objavljenim strategijama naših sveučilišta. U Strategiji studija i studiranja Sveučilišta u Zagrebu 2014.- 2025., (dio vezan uz e-učenje str. 32, cilj 2.6) navedeno je da se na Sveučilištu u Zagrebu u pravilu primjenjuje mješoviti oblik e-učenja koji omogućava korištenje najbolje prakse poučavanja u fizičkom i virtualnom okruženju. U Strategiji Sveučilišta u Splitu 2015-2020 kao jedan od ciljeva navedeno je da bi svi kolegiji svih studijskih programa trebali biti pokriveni barem 25% nastavnim sadržajima e-učenja.

Diana Laurillard (2004) definira online učenje (e-učenje) kao korištenje bilo koje nove tehnologije ili aplikacije, za učenje ili kao pomoć u učenju. Poblize, b-učenje je takva vrsta okruženja za učenje gdje se tradicionalna predavanja u učionici i komunikacija licem-u-lice između nastavnika i studenata isprepliće s učenjem potpomognutim računalom (Bubaš i Kermek, 2004). Specifičnosti hibridnog učenja je u tome da integrira tehnologiju, ali istovremeno koristi i prednosti klasičnih oblika nastave (Garrison i Kanuka, 2004). Prema Laurillard i sur., (2009, str. 304), potrebno je naći pravu mješavinu za svaki predmet i okruženje jer nema univerzalnog recepta te stoga treba biti sposoban naći pravi omjer.

Ulaskom studenta, nastavnika ili administratora u neki od sistema za upravljanje učenjem (LMS) u tom sistemu nastaje zapis o provedenoj aktivnosti. Svaki pristup LMS-u (klik, log, dnevnički zapis) generira podatke o tome tko je pristupio, kada, s koje IP adrese i koja je aktivnost provedena. Na taj način prikuplja se mnoštvo podataka o studentima i to tako da ih se pri tom ne ometa u procesu učenja.

Kako iskoristiti to mnoštvo podataka koje nam se samo nudi?

Ovdje je predložen jedan od mogućih načina: promatrati pristupe studenata određenom predmetu na LMS-u kao vremenske nizove. Zatim provjeriti nose li ti vremenski nizovi informacije o

ponašanju studenata koje bi mogle biti indikativne za njihov uspjeh na završnom ispitu. Potom, primijeniti Fourierovu transformaciju (FT) na kreirane vremenske nizove kako bi se smanjila količina podataka i učinila proces obrade podataka operativnim. Problem količine podataka ovdje nije bezazlen. Kao što će biti opisano kasnije, neki izračuni nisu bili gotovi ni nakon nekoliko dana. Drugi razlog primjene FT je otkriti skrivene periodičnosti koje su ključ modeliranja i predikcije, a treći razlog je smanjenje šuma te postizanje bolje točnosti predikcije uspjeha studenata. Autorica nije pronašla rad u kojem je spektralna analiza, odnosno analiza vremenskih nizova u frekvencijskoj domeni korištena na skupu podataka iz hibridnog okruženja za učenje i to drži temeljnim doprinosom ovog rada. Koja je bila motivacija da se Fourierova transformacija koristi za analizu vremenskih logova studenata? Promatranjem broja pristupa koji se generira u LMS-u, bilo ukupnog, po studentu ili nastavnom materijalu, uočeno je da postoji periodičnost. Ta periodičnost poklapala se s terminima seminara, vježbi i kolokvija, pa se tako pojavila ideja da se u toj periodičnosti nalazi obrazac povezan s različitim kategorijama studenata. Najzanimljivijom se činila povezanost studentskog uspjeha s ponašanjem na LMS-u i to je bio motiv za istraživanje. Vremenski niz studentskih logova može se shvatiti kao signal dobiven mjerenjem npr. zvuka ili signala dobivenog magnetskom rezonancijom. Takav signal najlakše je proučiti pomoću FT jer će istaknuti frekvencije, odnosno periodičnosti od kojih se signal sastoji. Pretpostavljeno je da FT koeficijenti povezani s određenim frekvencijama mogu ukazivati na obrasce ponašanja promatranih kategorija studenata.

Kako bi se preliminarno došlo do saznanja ima li cijeli ovaj plan smisla, primijenjena je analiza glavnih komponenti (engl. *principal component analysis*, PCA) na jedan skup podataka. Radi se o podacima o vremenskom nizu pristupa studenata LMS-u za jedan od kasnije analiziranih predmeta (Napredne WEB tehnologije i servisi u ak. g. 2012./2013., oko 100000 logova). Postizanje grupiranja podataka s obzirom na uspjeh studenata na završnom ispitu, pokazatelj je da vremenski nizovi podataka o pristupu studenata LMS-u nose informaciju o uspješnosti studenta. U literaturi postoje autori koji su se bavili otkrivanjem vremenskog uzorka aktivnosti studenata na LMS-u u svrhu predikcije, npr. Lara i sur. (2014), Cobo i sur. (2011) i Delgado i sur. (2006), no ne na ovakav način.

Hipoteze ovog rada ispitane su na skupovima podataka o studentima upisanim na 2 fakulteta, na ukupno 4 različita predmeta, tijekom dvije akademske godine, ukupno na približno 1400 studenata. Osam osnovnih skupova podataka (za svaki od 4 predmeta tijekom dvije akademske

godine) je dodatno ispitivano s obzirom na vrijeme kumulacije i s obzirom na diskretiziranje zavisne varijable na 2 i na 3 kategorije. Izračuni su provedeni pomoću 4 algoritma. Predložen je postupak iterativnog smanjivanja broja koeficijenata (atributa, engl. *attribute/ feature/ variable*) u modelima, kako bi se smanjio šum i povećala točnost. Uspoređivani su modeli kreirani pomoću postupka sažimanja broja koeficijenata s osnovnim modelima, koji sadrže pun broj koeficijenata. Napravljeno je više od 100.000 izračuna.

U istraživanju su ispitane osnovne postavke modela: je li predloženi model uopće moguće napraviti te postiže li točnost predikcije zbog koje bi vrijedilo dalje raditi na usavršavanju modela u smislu razvijanja tehnike sažimanja vremenskih nizova i u smislu praktične primjene. To su potrebni koraci kako bi se u nekim budućim istraživanjima moglo pristupiti proučavanju utjecaja pojedinih Fourierovih koeficijenata i njihovom povezivanju s ritmom kolegija, kako bi se mogli uspoređivati rezultati modeliranja na različitim tipovima kolegija te kako bi se moglo pristupiti traženju najprikladnijeg algoritma i podešavanju postavki algoritma. Ukratko, predloženi model predikcije uspjeha studenata rudarenjem podataka iz hibridnog okruženja za učenje primjenom analize vremenskih nizova pristupa studenata LMS-u u frekvencijskoj domeni pruža nov uvid i pridonijet će boljem razumijevanju procesa podučavanja u hibridnom okruženju te slijedno tome boljem planiranju i odabiru intervencija.

## **1.2. Ciljevi istraživanja**

Cilj je ovog istraživanja kreirati model predikcije uspjeha studenata rudarenjem podataka iz hibridnog okruženja za učenje primjenom analize vremenskih nizova pristupa studenata LMS-u u frekvencijskoj domeni. Razviti tehniku za sažimanje vremenskih nizova pristupa studenata LMS-u u frekvencijskoj domeni uz uklanjanje šuma, što znači postizanje bolje točnosti predikcije uz korištenje manjeg broja atributa. Nadalje, cilj je ovog istraživanja ispitati je li moguće postići točnost predikcije uspješnosti koja je veća od 75%, a temeljem podataka o prve 2/3 ukupnog trajanja predmeta.

### 1.3. Hipoteze istraživanja

H1: Točnost predikcije uspjeha studenata (na konačnom ispitu iz određenog predmeta) pomoću modela izgrađenog rudarenjem podataka iz hibridnog okruženja za učenje, a primjenom analize vremenskih nizova logova studenata u LMS-u u frekvencijskoj domeni, jednaka je ili veća od točnosti predikcije pomoću sličnih modela opisanih u literaturi.

H2: Primjenom tehnike sažimanja vremenskih nizova postiže se bolja točnost predikcije u odnosu na modele temeljene na punom skupu podataka.

H3: Točnost predikcije modela je jednaka ili veća od 75% kada se modelira temeljem podataka o prve 2/3 trajanja predmeta.

### 1.4. Metodologija istraživanja

Korištena je CRISP-DM metoda (engl. *cross-industry standard process for data mining*). Metoda ima sljedeće faze (Larose, 2014, p 6): razumijevanje poslovnih (obrazovnih) procesa, razumijevanje podataka, priprema podataka, modeliranje, evaluacija i primjena.

Prva faza tj. razumijevanje procesa pokazano je kroz pregled prethodnih istraživanja, točnije kroz opis studentskih modela s vremenskom segmentacijom atributa (dio 2.3.).

Druga faza, tj. razumijevanje podataka pokazano je pobliže u dijelu koji opisuje skup podataka (dio 3.1.) Uzorak je odabrani, prigodni (dostupni) uzorak. Za potrebe ovog rada obrađen je veliki broj podataka iz LMS-a na Sveučilištu u Zagrebu, s Medicinskog fakulteta te s Fakulteta organizacije i informatike. Na Medicinskom fakultetu obrađeni su podaci prikupljeni tijekom jednog predmeta prve studijske godine (1 semestar) te jednog predmeta četvrte godine (jedan tjedan) tijekom dvije akademske godine. S Fakulteta organizacije i informatike obrađeni su podaci jednog jednosemestralnog predmeta preddiplomskog studija i jednog jednosemestralnog predmeta diplomskog studija, također tijekom dvije akademske godine. Veličina ukupnog uzorka je oko 1400 studenata. Radi se o različitim tipovima studenata na dva fakulteta, s različitim frekvencijom pristupa. Jedino što je zajedničko je postojanje i korištenje LMS-a u klasičnoj nastavi. Studenti medicine dolaze s odličnim ocjenama iz srednje škole, posebice iz fizike, kemije i biologije, a studenti informatike su više tehnološki orijentirani, a traže se bolje ocjene iz



matematike. Očekivano, studenti Fakulteta organizacije i informatike imaju puno veću frekvenciju pristupa LMS-u od studenata medicine.

U dnevničkim zapisima (log) LMS-a čuvaju se izvještaji o aktivnostima sudionika. Studentski log je izvještaj o aktivnosti na Moodle-u: kojim stranicama je i kada student pristupao, s koje IP adrese i koje su bile aktivnosti (pogledaj, dodaj, osvježi i izbriši) kao što je prikazano na slici 1.

	id	time	jmbag	module	action	url	info
<input type="checkbox"/>	4822931	2011-09-25 09:43:40	Jedinstveni identifikacijski brojevi studenata	course	view	view.php?id=44	44
<input type="checkbox"/>	4823491	2011-09-25 10:39:23		course	view	view.php?id=44	44
<input type="checkbox"/>	4823935	2011-09-25 11:16:57		course	view	view.php?id=44	44
<input type="checkbox"/>	4823976	2011-09-25 11:20:21		course	view	view.php?id=44	44
<input type="checkbox"/>	4824018	2011-09-25 11:23:13		resource	view	view.php?id=3979	1098
<input type="checkbox"/>	4825026	2011-09-25 12:30:36		course	view	view.php?id=44	44
<input type="checkbox"/>	4825030	2011-09-25 12:30:41		forum	view discussion	discuss.php?d=3864	3864
<input type="checkbox"/>	4825032	2011-09-25 12:30:50		course	view	view.php?id=44	44
<input type="checkbox"/>	4825044	2011-09-25 12:31:13		course	view	view.php?id=44	44
<input type="checkbox"/>	4825048	2011-09-25 12:31:17		resource	view	view.php?id=16028	6667
<input type="checkbox"/>	4825137	2011-09-25 12:34:58		course	view	view.php?id=44	44
<input type="checkbox"/>	4825138	2011-09-25 12:35:01		forum	view discussion	discuss.php?d=3864	3864
<input type="checkbox"/>	4825143	2011-09-25 12:35:04		course	view	view.php?id=44	44
<input type="checkbox"/>	4825263	2011-09-25 12:44:12		course	view	view.php?id=44	44
<input type="checkbox"/>	4825280	2011-09-25 12:45:12		course	view	view.php?id=44	44
<input type="checkbox"/>	4827264	2011-09-25 14:50:18	course	view	view.php?id=44	44	
<input type="checkbox"/>	4827265	2011-09-25 14:50:24	forum	view discussion	discuss.php?d=3864	3864	
<input type="checkbox"/>	4827476	2011-09-25 15:07:24	course	view	view.php?id=44	44	
<input type="checkbox"/>	4827477	2011-09-25 15:07:28	resource	view	view.php?id=16028	6667	
<input type="checkbox"/>	4827527	2011-09-25 15:11:28	course	view	view.php?id=44	44	
<input type="checkbox"/>	4827530	2011-09-25 15:11:32	forum	view discussion	discuss.php?d=3864	3864	
<input type="checkbox"/>	4828729	2011-09-25 16:28:54	course	view	view.php?id=44	44	
<input type="checkbox"/>	4828730	2011-09-25 16:28:57	forum	view discussion	discuss.php?d=3864	3864	
<input type="checkbox"/>	4828732	2011-09-25 16:29:07	course	view	view.php?id=44	44	
<input type="checkbox"/>	4829482	2011-09-25 17:15:17	course	view	view.php?id=44	44	
<input type="checkbox"/>	4829717	2011-09-25 17:30:43	course	view	view.php?id=44	44	
<input type="checkbox"/>	4829720	2011-09-25 17:30:46	forum	view discussion	discuss.php?d=3864	3864	
<input type="checkbox"/>	4829727	2011-09-25 17:30:58	course	view	view.php?id=44	44	
<input type="checkbox"/>	4829820	2011-09-25 17:35:41	course	view	view.php?id=44	44	
<input type="checkbox"/>	4829821	2011-09-25 17:35:45	resource	view	view.php?id=16028	6667	

Slika 1: Tablica iz LMS-ove baze podataka koja sadrži podatke o svim studentima i njihovim logovima na određeni predmet

Svaka aktivnost je zabilježena kao jedan klik. Riječi klik, log, dnevnički zapis i pristup ovdje se koriste naizmjenično i s istim značenjem. U analizu dnevničkih zapisa uključeni su svi zapisi o aktivnostima studenata. Unutar tog skupa nalaze se i tzv. neproduktivni zapisi, a to su oni koji se odnose na prolazak kroz izbornike u LMS-u (engl. *course-view*). Za svakog studenta kreira se vremenski niz koji sadrži broj i vremenski slijed njegovih pristupa određenom predmetu na LMS-u za vrijeme trajanja tog predmeta.

Model je konstruiran tako da je zavisna varijabla uspjeh studenta na završnom ispitu iz određenog predmeta. Numerički podaci za zavisnu varijablu podijeljeni su u kategorijske klase u svrhu povećanja interpretabilnosti i razumijevanja (Romero, Ventura i García, 2008) i to na 2 načina: u dvije i u tri kategorije. Kategorije su formirane na sljedeći način: dvije kategorije (Dobri – ocjene 3,4,5 i Loši – ocjene 1 i 2), a tri kategorije (Vrlo dobri – ocjene 4 i 5, Prosječni –

ocjena 3, Loši – ocjene 1 i 2). Kategorije su formirane na opisani način kako bi distribucija ispitanika po kategorijama bila što ujednačenija. S druge strane, takvo formiranje kategorija može predstavljati izvjesno ograničenje o kojem treba voditi računa kod diskusije rezultata. Nezavisne varijable (vremenski nizovi pristupa LMS-u) su iz vremenske domene transformirane u frekvencijsku domenu.

Vrijeme kumulacije pristupa (period u kojem se signal mjeri) je 1 sat i 1 dan. Promatrajući ukupni broj pristupa studenata za vrijeme trajanja kolegija uočeno je da se broj pristupa kreće oko 100 000. Kad se taj broj podijeli s brojem upisanih studenata, može se zaključiti da je najkraće smisljeno vrijeme akumulacije signala 1 sat. Za neko kraće vrijeme akumulacije broj promatranih točaka u vremenu dramatično bi se smanjio, a u većini točaka ne bi bilo ni jednog pristupa što bi onemogućilo analizu. Zanimljivo je također da li produživanje vremena akumulacije mijenja rezultate analize. Kako bi se to provjerilo, bilo je logično kao drugo vrijeme kumulacije odrediti 1 dan.

Treća faza tj. priprema podataka ima veliku težinu u ovom radu jer upravo sažimanje podataka, uklanjanje šuma i otkrivanje periodičnosti predstavlja srž ovog rada. Faza pripreme podataka je opisana u dijelu 4.1. U fazi pretprocesiranja podataka obrađeni su podaci iz tablice LMS-ove baze pristupa studenata određenim resursima promatranog predmeta tako da su dobiveni vremenski nizovi po pojedinom studentu. Oni su zatim transformirani u periodograme po studentu (frekvencijska domena), odabrano je 20 najintenzivnijih koeficijenata po studentu i zatim izrađena zajednička tablica najintenzivnijih Fourierovih koeficijenata koja je korištena za modeliranje.

Četvrta faza, tj. modeliranje opisano je pobliže u dijelu 4.3. Kreirani su modeli predikcije za 4 predmeta, tijekom 2 akademske godine, a zavisna varijabla diskretizirana je na 2 i 3 klase, s vremenom kumulacije 1 sat i 1 dan. Budući da jedan od promatranih predmeta traje samo 5 radnih dana, za taj predmet je promatrano samo vrijeme kumulacije 1 sat. Za taj predmet su ocjene prošao/pao te je zavisna varijabla diskretizirana u samo 2 klase). Stoga je promatrano ukupno 26 osnovnih skupova podataka.

Peta faza, tj. evaluacija opisana je u dijelu 4.3. Za svaki od navedenih skupova, izračunata je točnost predikcije pomoću 4 uobičajena algoritma. Kao mjera uspješnosti modela predikcije korištena je točnost, tj. postotak točno klasificiranih ishoda (pozitivnih i negativnih) u odnosu na ukupan broj ishoda.

Točnost predikcije dobivena rudarenjem vremenskih nizova logova studenata u LMS-u u frekvencijskoj domeni uspoređena je s točnošću predikcije pomoću sličnih modela opisanih u literaturi.

Iterativno je smanjivan broj Fourierovih koeficijenata kako bi se uočio trend kretanja točnosti predviđanja u odnosu na broj korištenih Fourierovih koeficijenata te kako bi se definirala točka u kojoj je točnost predikcije maksimalna, a broj korištenih koeficijenata minimalan (ako postoje lokalni minimumi). Dobivena maksimalna točnost je uspoređena s točnošću predikcije modela konstruiranim korištenjem punog broja tj. svih koeficijenata.

Dužina vremenskog niza smanjena je na prve 2/3 trajanja predmeta kako bi se uočilo je li točnost predikcije od barem 75%.

Zbog vrlo velikog broja modela za ocjenu točnosti rezultata korištena je metoda jednostrukog vrednovanja u omjeru 70% uzorka za učenje i 30% za testiranje. Pri usporedbi točnosti predikcije korišten je neparametrijski statistički test. Kako se radi o usporedbi više algoritama na istim skupovima podataka, primijenjen je Friedmanov test (Demšar, 2006; Čehovin i Bosnić, 2010). U slučajevima gdje Friedmanov test pokazuje da postoji statistički značajna razlika primijenjena je *post hoc* Nemenyi test (Nemenyi, 1963) za višestruka testiranja kako bi se utvrdilo između kojih parova modela/metoda postoji statistički značajna razlika. Wilcoxonov test usklađenih parova (Wilcoxon, 1945) (engl. *Wilcoxon matched pairs test*) je korišten za usporedbu točnosti na modelima s punim brojem koeficijenata i modelima sa smanjenim brojem koeficijenata, tj. prije i poslije transformacije predložene u ovom radu.

Šesta faza, tj. primjena u ovom radu nije obrađena jer je istraživanje eksploratorno i potrebna su dodatna istraživanja da bi se došlo do faze primjene.

### **1.5. Struktura rada**

Rad započinje pregledom prethodnih istraživanja koji uključuje kratki osvrt na najvažnija područja istraživanja vezana uz e-učenje i b-učenje. Slijedi opis razvoja i djelokruga edukacijskog rudarenja podataka (engl. *educational data mining*, EDM) kao novog istraživačkog područja te pregled znanstvenih istraživanja na području predviđanja uspjeha studenata. Konačno, napravljen je pregled radova koji se bave studentskim modelima s vremenskom segmentacijom atributa.

U narednom dijelu opisan je skup podataka, definirani su vremenski nizovi, vrijeme kumulacije i Fourierove transformacije. Opisan je razlog prebacivanja podataka iz vremenske u frekvencijsku domenu. Definirani su pojmovi klasifikacije i analize glavnih komponenti.

U sljedećem dijelu detaljno je opisan proces prethodne obrade podataka. Naveden je kratki prikaz rada algoritama korištenih za predikciju. Opisan je proces modeliranja. Definirani su pojmovi točnosti i vrednovanja i opisani korišteni testovi statističke značajnosti.

Nadalje, opisani su prethodni rezultati vezani uz primjenu analize glavnih komponenti na jednom skupu podataka. Navedene su usporedbe točnosti predikcije po algoritmima, po dužini vremenskog niza i po vremenu kumulacije. Spomenute usporedbe važne su za dio koji slijedi, a to su rezultati testiranja hipoteza.

Naredno poglavlje sadrži sintezu rezultata testiranja hipoteza te doprinose i ograničenja istraživanja.

Sljedeći dio daje kratak osvrt na etičke aspekte istraživanja.

Na kraju rada nalazi se zaključak i smjernice za buduća istraživanja.

## **2. PREGLED PRETHODNIH ISTRAŽIVANJA**

Primjena e-učenja i posebice, b-učenja u praksi nema dugu povijest što znači da su otvorena mnoga pitanja u vezi s njihovom primjenom. Radi se o vrlo dinamičnom, ali i relativno neistraženom području. Stoga su prethodna istraživanja, odnosno istraživačka pitanja kao i korištena metodologija, vrlo raznolika.

Primjerice, primjena tehnologije otvara mogućnost široke primjene formativnog e-ispitivanja. Utjecaj formativnog e-ispitivanja na konačni rezultat i stav studenata prema takvom načinu ispitivanja proučavali su Wilson i sur. (2011), Chetchumlong (2010), Cruz i sur. (2011). Chen i sur. (2007) nastoje doći do dijagnoze e-učenja dubinskom analizom profila studenta. Rad autora Terzis i Economides (2010) nastavlja se na cijelo područje koje se bavi izgradnjom modela s temeljnim konceptima koji utječu na bihevioralnu namjeru studenata da koriste ispitivanje potpomognuto računalom (engl. *computer based assessment*, CBA). Amoroso (2005) istražuje način uporabe online sredstva za ispitivanje namijenjenog velikim skupinama, a Deneui i Dodge (2006) istražuju vezu između učestalosti korištenja raznih resursa na LMS-u i uspjeha studenta u hibridnom okruženju za učenje.

Integracija i revizija teorija o stilovima učenja u cilju definiranja uloge klasičnog načina podučavanja u odnosu na multimediju (Kliček i Zekić Sušac, 2003), kao i razvoj sistema za učenje koji se prilagođava stilu učenja pojedenog studenta i načinu njegovog korištenja materijala za učenje (Ocepek i sur., 2013) vrlo su zanimljiva područja istraživanja. Feldman i sur. (2014) istražuju kako detektirati stil percepcije studenta pomoću igara.

Dráždilová i sur. (2010) nastoje pronaći latentne studentske društvene mreže temeljem podataka pohranjenih u Moodle-ovim logovima, a iz kojih dolaze do zaključaka o sličnom ponašanju pojedinih grupa studenata

Uza svu raznolikost, svim ovim radovima je zajedničko da koriste brojne podatke o korištenju LMS-a i/ili raznih oblika online procjene znanja tijekom trajanja predmeta te da ti podaci daju brojne informacije o ponašanju studenata i o rezultatima studenata. To je pridonijelo razvoju novih istraživačkih područja: rudarenja edukacijskih podataka te analitike učenja i znanja.

### **2.1. Rudarenje edukacijskih podataka**

U rudarenju edukacijskih podataka (engl. *educational data mining*, EDM) kao i u analitici učenja i znanja (engl. *learning analytics and knowledge*, LAK) kao novim istraživačkim područjima,

jedna od popularnijih zadaća je predviđanje uspjeha studenata (Romero i Ventura, 2007; Baker i Yacef, 2009; Siemens i Baker, 2010; Gasevic i sur., 2011). Polazište tih istraživačkih područja je da prikupljanje, analiziranje i izvještavanje temeljeno na mnoštvu podataka koji nastaju iz međudjelovanja studenata i informacijsko komunikacijskih tehnologija, može pružiti nov uvid u procese učenja i podučavanja (Dawson i sur., 2015) te utjecati na učenje (Divjak i Maretić, 2015). Nešto šire, novo, istraživačko područje je više modalna analitika znanja (engl. *multimodal learning analytics*) koja predstavlja skup tehnika koje koriste brojne izvore podataka (video, logovi, rukotvorine studenata, audio, geste, biosenzori) za proučavanje učenja u realnom, ekološki ispravnom, multimedijalnom okruženju za učenje (Blikstein i Worsley, 2016).

Prema Zafra i sur. (2011) rudarenje podataka je proces ekstrahiranja uzoraka (engl. *pattern*) iz podataka, a EDM daje dodatni uvid u ponašanje predavača, studenata, menadžera i drugog nastavnog osoblja i djeluje kao aktivna i automatska pomoć za donošenje boljih odluka o nastavnim aktivnostima.

U hrvatskom jeziku nalazimo prijevod s engleskog *data mining*, i kao rudarenje podataka (Velić, 2014) i kao dubinska analiza podataka (Vranić, 2009). U ovom radu će se koristiti obje inačice prijevoda. Međunarodno društvo za dubinsku analizu edukacijskih podataka definira EDM kao novo istraživačko područje koje se bavi razvojem metoda za istraživanje jedinstvenih vrsta podataka dobivenih iz edukacijskog okruženja i korištenjem tih metoda za bolje razumijevanje studenata i okruženja u kojem uče (Siemens i Baker, 2010).

Rudarenje edukacijskih podataka razvijeno je preslikavanjem ideje iz e-trgovine u edukacijsko okruženje. U e-trgovini dubinska analiza se koristi kako bi se procijenio interes klijenata i povećala prodaja (Romero i Ventura, 2010). Rudarenje edukacijskih podataka razlikuje se od rudarenja u e-trgovini po sudionicima, odnosno zainteresiranim stranama i po specifičnim ciljevima. Sudionici su studenti, predavači, autori predmeta, sveučilišta i fakulteti te sistem administratori (Romero i Ventura, 2010). Specifični ciljevi rudarenja edukacijskih podataka su unaprjeđenje procesa podučavanja, vođenje učenja te razumijevanje obrazovnih fenomena. Karakteristika je svih tih ciljeva da ih je teško kvantificirati i mjeriti (Romero i Ventura, 2007) što ujedno predstavlja i razliku u odnosu na e-trgovinu i opravdava razvoj EDM kao zasebnog područja istraživanja.

Uobičajeni koraci EDM su prikupljanje podataka, prethodna obrada podataka, primjena rudarenja podataka te interpretiranje i vrednovanje rezultata (Romero, Ventura i García, 2008). U

svim spomenutim koracima ima prostora za dodatna istraživanja: odabir i konstrukcija atributa, zavisnih varijabli, algoritama i metoda.

Prema Baker i Yacef, (2009) tijekom perioda od 1995-2005 najučestaliji oblik rudarenja edukacijskih podataka bila je dubinska analiza ili rudarenje veza (engl. *relationships*), a predviđanje je bilo na drugom mjestu po učestalosti istraživanja. Tijekom 2008-2009 dogodio se preokret, pa je predviđanje postalo najučestaliji oblik dubinske analize podataka, a drugo mjesto su zauzela istraživanja pomoću modela. Pojava istraživanja pomoću modela objašnjava se uključivanjem istraživača koji se bave psihometrijom i studentskim modeliranjem u EDM zajednicu. Pri istraživanju pomoću modela, razvija se model neke pojave, a zatim se taj model koristi kao dio neke druge analize, kao što je predviđanje ili rudarenje veza.

Prema autorima Sison i Shimura (1998) studentski model je približan, moguće i djelomičan, a prvenstveno kvalitativan prikaz studentskog znanja o određenoj domeni, odnosno temi ili vještini unutar te domene, koji nam može potpuno ili djelomično objasniti određene vidove studentskog ponašanja. Studentski model nam daje podatke o studentu koji koristi neki sistem za učenje temeljen na računalu, a sa svrhom da se sistem što bolje prilagodi studentu. Gradnja modela uključuje prethodnu obradu podataka, optimiziranje parametara i odabir atributa.

Modeliranje studenta i predviđanje uspješnosti u sistemima za e-učenje (Kotsiantis i sur., 2004; Lykourantzou i sur., 2009; Cobo i sur., 2011; Jovanović i sur., 2012; Dorça i sur., 2013; Lara i sur., 2014; Romero i Ventura 2017) te u tutorskim i adaptabilnim sistemima za učenje (Thai-Nghe i sur., 2010; Chen i sur., 2007) razlikuje se od modeliranja u hibridnim okruženjima za učenje (Dias i Diniz, 2013) budući da studentski modeli u hibridnim okruženjima za učenje u pravilu raspolažu s manje podataka. Radovi koji se odnose na hibridna okruženja za učenje nisu tako brojni i, očekivano, ne temelje se tako često na metodama strojnog učenja (Delgado i sur., 2006 –umjetna neuronska mreža; Deneui i Dodge, 2006 – korelacija učestalosti korištenja LMS-a i uspjeha; Thai-Nghe i sur., 2007 – stabla odlučivanja i Bayesove mreže; Dekker i sur., 2009 – naivni Bayesov klasifikator; Divjak i Oreški, 2009 - diskriminacijska analiza). Bez obzira na to, hibridna okruženja za učenje sve su češća u visokom obrazovanju u Hrvatskoj te stoga trebaju biti bolje istražena.

## 2.2. Studentski modeli

Atributi koji se obično koriste u modeliranju studenata su npr. broj riješenih zadataka, broj pristupa kvizu, broj kvizova koje je student položio, broj kvizova koje student nije položio, broj poruka na forumu koje je student napisao, broj poruka na forumu koje je student pročitao, vrijeme utrošeno za zadatke, vrijeme utrošeno za kvizove, vrijeme utrošeno na forumima (Jovanović i sur., 2012) kao i sociološki i demografski pokazatelji te pokazatelji percepcije studenata o studiranju. Često se kao zavisna varijabla za modeliranje uzima završna ocjena za promatrani predmet (Romero, Ventura i García, 2008.; Jovanović i sur., 2012; Thai Nghe i sur., 2007; Nasiri i sur., 2012), zatim postotak odustajanja od studija (Lykourentzou i sur., 2009; Dekker i sur., 2009; Jadrić i sur.2010). Autorice Divjak i Oreški (2009) koristile su više od 30 atributa (demografski i percepcija studenata o studiju) za podatke prikupljene upitnikom, a kako bi predvidjele studentski uspjeh koristeći diskriminacijsku analizu. Autori Jadrić i sur. (2010) bave se predviđanjem odustajanja od studija u visokom obrazovanju koristeći regresiju, stabla odlučivanja i umjetnu neuronsku mrežu (engl. *artificial neural network*, ANN). Na uzorku od 715 studenata upisanih na Ekonomski fakultet u Splitu kao prediktorske varijable koriste demografske podatke te podatke o upisanim predmetima, ispitima kao i postignutom rangu na razredbenom ispitu.

Moguće su i varijante gdje je zavisna varijabla kvaliteta interakcije s LMS-om (Dias i Diniz, 2013) kao i distribucija vjerojatnosti određenog stila učenja (Dorça i sur., 2013). Istraživanje Dorça i sur. (2013) provedeno je kao simulacija, a ne na stvarnom skupu podataka. Dekker i sur. (2009) kao prediktorske varijable koriste broj do sada upisanih predmeta, njihovu prosječnu ocjenu, broj upisanih prirodnih predmeta, broj upisanih matematičkih predmeta i prosječnu ocjenu matematičkih predmeta. Kao ciljna varijabla za modeliranje uzeta je stopa odustajanja od studija.

## 2.3. Studentski modeli s vremenskom segmentacijom atributa

Postoje razni modeli predviđanja uspjeha/ odustajanja od studija, međutim pristupi studenata LMS-u, promatrani kao vremenski niz, rijetko čine prediktorsku varijablu u hibridnim okruženjima za učenje. Očekivano, tutorski i adaptabilni sistemi za učenje na tom su području puno više istraživani. Postavlja se pitanje je li moguće predvidjeti uspjeh studenata obradom samo dnevničkih zapisa u LMS-u. Pristupi čine vrstu podataka koja se automatski prikuplja na



svim kolegijima na visokim učilištima gdje se LMS koristi. U preliminarnim istraživanjima je uočeno da sam broj pristupa nekom kolegiju kao ni vrijeme iskorišteno za određenu aktivnost nemaju značajniju prediktivnu snagu pri modeliranju uspjeha studenata (Gamulin i sur., 2013). Međutim, ako se pristupi promatraju kao vremenski nizovi, možda je moguće, uz određene obrade, modeliranje uspjeha na konačnom ispitu iz određenog predmeta.

S obzirom da se u Hrvatskoj i Europi većinom koristi ili se nastoji koristiti hibridno okruženje za učenje (misije većine naših sveučilišta) izrada takvog robusnog modela koji ne zahtijeva dodatno prikupljanje podataka osim već postojećih, ima svoje opravdanje.

Radovi koji se bave otkrivanjem vremenskog uzorka ponašanja na LMS-u su relativno rijetki. Cobo i sur. (2011) su kao ciljnu varijablu uzeli studentsku aktivnost na online forumima, podijeljenu u sljedeće kategorije: aktivni učenik, promatrač i zabušant. Radilo se o virtualnom okruženju za učenje. Skup podataka temeljen je na 55 studenata upisanih u predmet Elektronički krugovi na Otvorenom sveučilištu Katalonije u Španjolskoj. Predloženo je korištenje aglomerativnog hijerarhijskog klaster algoritma. U radu nema rezultata jer je riječ o preliminarnom izvještaju.

Delgado i sur. (2006) modeliraju završne ocjene na kraju predmeta. Istraživanje je provedeno na uzorku od 240 studenata upisanih u predmet Metodologija i tehnologija programiranja Sveučilišta u Cordobi u Španjolskoj. Atributi su: broj službenih registracija u predmet, broj provjera znanja, ocjena, ukupan broj pristupa LMS-u tijekom trajanja predmeta, ukupan broj pregleda resursa, postotak pregleda resursa u odnosu na ukupan broj pristupa, broj pregleda resursa po različitim kategorijama (teorija, primjeri), postotak broja pregleda resursa po kategorijama, segmentacija broja pristupa po mjesecima, segmentacija postotka pristupa po mjesecima.

Dias i Diniz (2013) bave se modeliranjem studenata i predavača na ciljnu varijablu kvaliteta interakcije (engl. *quality of interaction*, QoI) s LMS-om u hibridnom okruženju za učenje. Podaci se odnose na studente upisane na Tehničko sveučilište Lisabon u Portugalu. U modeliranju je korišten neizraziti (engl. *fuzzy*) sistem sa 600 neizrazitih ekspertnih pravila i skup od 110 LMS metrika. Korišteni su podaci o 5 hibridnih predmeta, tijekom 51 tjedna, a istraživanje je uključivalo 75 profesora i 1037 studenata, odnosno oko 611000 interakcija s LMS-om. Kao vrijeme kumulacije interakcija korišten je jedan tjedan. Oblici interakcije: gledaj (engl. *view*, V) – Journal, Wiki, blog obrazac, forum, diskusija, chat, predaj, izvještaj, kviz,

povratna informacija, stranica predmeta; dodaj (engl. *addition*, AD) – modul, post, aktivnost, resurs, zadatak, oznaka; mijenjanje (engl. *alteration*, AL) – upload, osvježi, dodijeli, uredi/ izbriši. Ideja da vremenski uzorak logova na LMS nosi informaciju o studentima opisana je u radu Lara i sur. (2014) no u okruženju isključivo online učenja te predviđajući postotak odustajanja. Lara i sur. (2014) su predložili sistem koji se temelji na analizi podataka dobivenih iz interakcije studenata s LMS-om. Konstruirali su nekoliko kategorija interakcija, a analiza stvarnih podataka je pokazala da studenti uglavnom koriste radnju “pogledaj”. S obzirom na to, u ovoj disertaciji podaci će biti generalizirani tako da će se promatrati isključivo pristupi bez obzira o kakvoj se aktivnosti zapravo radi. E-učenje, kako je opisano u radu Lara i sur. (2014), specifično je po vremenskom i prostornom jazu između predavača i studenata kao i po visokoj stopi odustajanja (70-80%). Kao nezavisne varijable korišteni su: broj radnji na LMS-u u tjednu za pojedinog studenta, broj dana u tjednu u kojima je bilo radnji na LMS-u, broj pregleda tjedno po studentu po određenom materijalu na LMS-u. Mjerena je euklidska udaljenost atributa svakog studenta od atributa historijskog modela. Skup podataka za učenje bio je 100 studenata, a skup za validaciju 50 studenata po svakom od 4 promatrana predmeta. Radilo se o podacima za 4 predmeta koja traju po 20 tjedana. Koristeći Sistem za edukacijsko rudarenje podataka (engl. *system for educational data mining*, SEDM), koji autori predlažu, na polovici trajanja predmeta postignuta je točnost predikcije od 94%. U tu svrhu bili su potrebni povijesni podaci o dvije generacije studenata kako bi se model mogao testirati na trećoj generaciji od 50 studenata. Model za predmet s prve godine studija ima puno nižu točnost predikcija od modela koji se odnose na predmete s ostale tri godine studija.

Dobashi (2015) koristi analizu vremenskih nizova u hibridnom okruženju za učenje. Nastava na promatranom predmetu traje tijekom 15 srijeda u zimskom semestru 2014. godine. Studenti su upisani u predmet Uvod u analizu društvenih podataka na Aichi sveučilištu u Japanu. Studentima je dostupno 124 PDF-a na LMS-u (Moodle-u). Podaci se odnose na 61 studenta, odnosno na približno 30000 logova. Vrijeme kumulacije je 5 minuta. Pomoću Excel tablica je praćeno pristupaju li studenti temama koje se obrađuju pravovremeno (onako kako je profesor sugerirao) ili sa zakašnjenjem ili pak pristupaju nasumično temama, nevezano uz redoslijed predavanja. Prikupljene informacije poslužit će za unaprjeđenje budućih predavanja.

Autori Xie i sur. (2014) koriste analizu vremenskih nizova kako bi procijenili uspješnost procesa dizajniranja pri korištenju CAD programa (engl. *computer aided design*). Rad se temelji na

analizi vremenskih nizova koje se prikupljaju neprimjetno, za vrijeme dok studenti rade, bez ometanja procesa učenja. Uzorak se sastojao od samo 20 učenika srednje škole u Massachusettsu, SAD, od kojih je 16 bilo muških i 4 djevojke. Razmatran je broj akcija u minuti, broj upotrijebljenih elemenata te odnos broja akcija koje se odnose na gradnju i broja akcija koje se odnose na popravak. Za analizu su korišteni vizualni prikazi i elementarna statistika.

Khan i sur. (2012) kategoriziraju studente temeljem prosječnog intervala između dvije sesije na online forumima i duljine trajanja njihovih sesija na online forumima. Istraživanje je provedeno tijekom dvije akademske godine na preddiplomskom studiju Projektnog menadžmenta na poslovnoj školi Brunel u Londonu, Ujedinjeno kraljevstvo. Uzorak se sastojao od podataka o 303 studenta, a studenti su komunicirali putem foruma na Blackboard-u. Temeljem podataka o aktivnosti na online forumu provedena je hijerarhijska klaster analiza kako bi se identificirale podgrupe među studentima. Autori zaključuju da aktivnost na online forumu ima određenu vrijednost zbog ostvarivanja društvene interakcije no dobiveni pokazatelji o ponašanju studenata nisu dovoljni da bi se moglo prosuđivati o njihovoj uspješnosti.

### 3. SKUP PODATAKA I METODOLOŠKI OKVIR

#### 3.1. Skup podataka

Uzorak je odabrani, prigodni (dostupni) uzorak. Obraden je veliki broj podataka iz LMS-a na Sveučilištu u Zagrebu, s Medicinskog fakulteta te s Fakulteta organizacije i informatike. Na Medicinskom fakultetu obrađeni su podaci prikupljeni tijekom jednog predmeta prve studijske godine (1 semestar) te jednog predmeta četvrte godine (jedan tjedan) tijekom akademskih godina 2012./2013. i 2013./2014. S Fakulteta organizacije i informatike obrađeni su podaci jednog predmeta preddiplomskog studija i jednog predmeta diplomskog studija, također tijekom spomenute dvije akademske godine. Oba predmeta s Fakulteta organizacije i informatike traju jedan semestar. Veličina ukupnog uzorka je oko 1400 studenata.

Podaci o svim predmetima i inicijalnim skupovima podataka nalaze se tablici 1 (za 2012./2013.) i tablici 2 (za 2013./2014.).

*Tablica 1: Pregled ispitanih predmeta i inicijalnih skupova podataka za akademsku godinu 2012./2013.*

Naziv predmeta	Broj studenata 2012/2013	Duljina trajanja predmeta (dani)	Broj logova po predmetu	Broj kategorija zavisne varijable	Broj ispitanika po kategorijama	Vrijeme kumulacije	Skup varijabli
					D/L i VD/P/L		
Fizika i biofizika	308	120	106465	2	213/95	T <sub>sat</sub>	FIB 2K <sub>sat</sub>
						T <sub>dan</sub>	FIB 2K <sub>dan</sub>
				3	157/56/95	T <sub>sat</sub>	FIB 3K <sub>sat</sub>
						T <sub>dan</sub>	FIB 3K <sub>dan</sub>
Fizika medicinske dijagnostike	255	5	20149	2	197/58	T <sub>sat</sub>	FMD 2K <sub>sat</sub>
Napredne WEB tehnologije i servisi	64	120	101988	2	36/28	T <sub>sat</sub>	NWTIS 2K <sub>sat</sub>
						T <sub>dan</sub>	NWTIS 2K <sub>dan</sub>
				3	16/20/28	T <sub>sat</sub>	NWTIS 3K <sub>sat</sub>
						T <sub>dan</sub>	NWTIS 3K <sub>dan</sub>
WEB dizajn i programiranje	80	120	114621	2	41/39	T <sub>sat</sub>	WDIP 2K <sub>sat</sub>
						T <sub>dan</sub>	WDIP 2K <sub>dan</sub>
				3	15/26/39	T <sub>sat</sub>	WDIP 3K <sub>sat</sub>
						T <sub>dan</sub>	WDIP 3K <sub>dan</sub>

Tablica 2: Pregled ispitanih predmeta i inicijalnih skupova podataka za akademsku godinu 2013./2014.

Naziv predmeta	Broj studenata 2013/2014	Duljina trajanja predmeta (dani)	Broj logova po predmetu	Broj kategorija zavisne varijable	Broj ispitanika po kategorijama	Vrijeme kumulacije	Skup varijabli
					D/L i VD/P/L		
Fizika i biofizika	310	120	99323	2	212/98	T <sub>sat</sub>	FIB 2K <sub>sat</sub>
						T <sub>dan</sub>	FIB 2K <sub>dan</sub>
				3	156/56/98	T <sub>sat</sub>	FIB 3K <sub>sat</sub>
						T <sub>dan</sub>	FIB 3K <sub>dan</sub>
Fizika medicinske dijagnostike	269	5	21156	2	196/73	T <sub>sat</sub>	FMD 2K <sub>sat</sub>
Napredne WEB tehnologije i servisi	51	120	75916	2	37/14	T <sub>sat</sub>	NWTIS 2K <sub>sat</sub>
						T <sub>dan</sub>	NWTIS 2K <sub>dan</sub>
				3	8/29/14	T <sub>sat</sub>	NWTIS 3K <sub>sat</sub>
						T <sub>dan</sub>	NWTIS 3K <sub>dan</sub>
WEB dizajn i programiranje	84	120	122466	2	55/29	T <sub>sat</sub>	WDIP 2K <sub>sat</sub>
						T <sub>dan</sub>	WDIP 2K <sub>dan</sub>
				3	25/30/29	T <sub>sat</sub>	WDIP 3K <sub>sat</sub>
						T <sub>dan</sub>	WDIP 3K <sub>dan</sub>

### 3.2. Vremenski nizovi

Pod vremenskim nizom misli se na uređeni niz opažanja, a uređeni su najčešće u odnosu na vrijeme, tj. na jednake vremenske intervale. Opažanja nisu međusobno nezavisna već i zbog toga što je bitan njihov vremenski poredak (Kovačić, 1995, str. 2). Analiza vremenskih nizova česta je u geofizici, meteorologiji, medicini, ekonomiji te u analizi pokazatelja na burzama.

Kada se neprekinuta pojava prati u određenim vremenskim intervalima, diskretnu vrijednost možemo dobiti vremenskim agregiranjem ili kumuliranjem kontinuiranog vremenskog niza u određenom intervalu vremena. Analizi vremenskih nizova može se pristupiti u vremenskoj domeni i u frekvencijskoj domeni. Pri analizi u vremenskoj domeni, vremenski niz se promatra kao funkcija vremena i koristi se autokorelacijska funkcija koja opisuje uzajamnu ovisnost dva opažanja iz vremenskog niza. Pri analizi u frekvencijskoj domeni, vremenski nizovi se prikazuju kao funkcija frekvencije i koristi se spektralna funkcija gustoće. Spektralna funkcija gustoće ukazuje na doprinose periodičnih komponenti na različitim frekvencijama ukupnoj varijanci vremenskog niza. Analiza iregularnih ciklusa tj. ciklusa koji se ne pojavljuju na jednak način i

jednakim intenzitetom u jednakim vremenskim periodima, temelji se na spektru. Fourierova analiza je metodološka osnova analize vremenskog niza u frekencijskoj domeni.

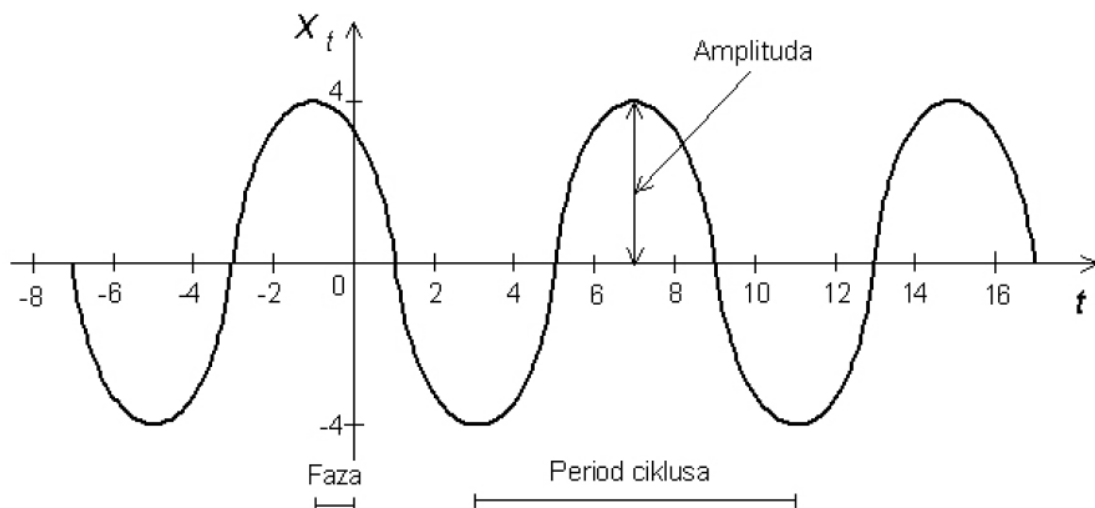
Jednostavan ciklički model s poznatim periodom može se prikazati:

$$X_t = A \cos(\omega t + \theta), \quad t = 0, \pm 1, \pm 2, \dots$$

Gdje je  $\omega$  kutna frekvencija ili frekvencija kosinusnih varijacija,  $A$  je amplituda tih varijacija, a  $\theta$  je faza tih varijacija, odnosno njihov fazni pomak. Kut  $\omega t$  je neimenovani broj, a  $\omega$  je frekvencija izražena u radijanima po vremenskoj jedinici.

Period ciklusa  $2\pi/\omega$  iskazan je u jedinici vremena i zove se dužina ciklusa.

Slika 2 prikazuje jednostavni ciklički model.



*Slika 2: Jednostavni ciklički model*

*Izvor: Kovačić, 1995., p 254*

Pojava vrha na spektralnoj funkciji gustoće pokazatelj je prisutnosti ciklične komponente u nizu. Gladak oblik spektra s visokom vrijednošću ordinate oko nulte frekvencije može biti pokazatelj postojanja trenda (Kovačić, 1995, str. 289). Zanimanje za vremenske nizove povećava se s većom učestalošću longitudinalnih baza podataka. Upravo zbog specifičnosti vremenskih nizova, dubinska analiza podataka iz takvih skupova drži se jednim od većih izazova i značajnim područjem istraživanja (Fu, 2011; Yang i Wu, 2006). Najznačajnije karakteristike takvih baza podataka su ogromna količina podataka, prisutnost šuma i nelinearne veze među elementima podataka (Wilson, 2017). Problem prevelike količine podataka nastoji se riješiti sažimanjem tako

da se sa što manjim brojem podataka pokrije većina varijacija. Diskretna Fourierova transformacija (engl. *discrete Fourier transform*, DFT), odnosno brza Fourierova transformacija (engl. *fast Fourier transform*, FFT) kao unaprijeđeni algoritam za njeno računanje (Chen i Chen, 2014) jedna je od popularnih tehnika transformacije vremenskih nizova u frekvencijsku domenu i obrnuto te način sažimanja podataka u vremenskim nizovima. Spektralna analiza ne drži se konvencionalnom metodom analize vremenskih nizova no ona pruža uvid u složene ili autokorelacijske strukture te se u slučajevima gdje složeni uzorci u periodogramu predstavljaju cikličke trendove i autokorelacijske strukture drži boljim odabirom od ARMA (engl. *autoregressive moving average*) modela (Bagnall i Janacek, 2005).

### 3.3. Vrijeme kumulacije

Općenito, vremenski niz je skup opažanja  $X_i$ , koja su zabilježena u vremenu  $t$ . U ovom slučaju vremenski niz je diskretan  $t_i$  gdje je  $i=0,1,2,3, \dots, N$ , a brojevi predstavljaju sate ili dane, ovisno o tome u kojim je jednakim i jednako udaljenim periodima  $T$  signal mjeran.

Praktično, to znači da je  $T = t_i - t_{i-1}$  u stvari period vremena i zovemo ga vrijeme kumulacije. Ovdje se koriste 2 vremena kumulacije  $T$ ,  $T_{sat} = 1$  sat, i  $T_{dan} = 1$  dan. To znači da je bilježen broj studentskih klikova u jednom satu, odnosno u jednom danu.

Kao rezultat tih mjerenja generirani su vremenski nizovi za svakog studenta opisujući broj i vremenski slijed studentskih klikova na određeni LMS predmet, a za vrijeme trajanja tog predmeta.

### 3.4. Fourierova transformacija

Iz generiranih vremenskih nizova izračunata je DFT. Normalizirana DFT vremenskog niza  $x(n)$ ,  $n=0,1 \dots N-1$  je vektor kompleksnih brojeva  $X(f)$ :

$$X(f_{k/N}) = \frac{1}{\sqrt{N}} \sum_n^{N-1} x(n) e^{-i2\pi kn/N}, k = 0,1 \dots N-1$$

Vremenski nizovi u ovom radu su stvarni signali, a Fourierovi koeficijenti su simetrični oko sredine. Fourierova transformacija predstavlja linearnu kombinaciju kompleksnih sinusoida

$$s_f(n) = \frac{e^{i2\pi kn/N}}{\sqrt{N}}.$$

Prema tome, Fourierovi koeficijenti predstavljaju amplitude svake od tih sinusoida, nakon što je signal  $x$  projiciran na njih. Zatim su korištene dominantne frekvencije kako bi se pojednostavile vremenski nizovi i pri tom očuvale najvažnije periodičnosti. Pod dominantnim frekvencijama se misli na one koje nose najviše energije signala. Način da se identificiraju te frekvencije je izračunavanje periodograma kvadriranjem veličine Fourierovih koeficijenata:

$$P\left(f_{\frac{k}{N}}\right) = \|X(f_{k/N})\|^2, \quad k = 0, 1 \dots \left[\frac{N-1}{2}\right]$$

Dominantne frekvencije su vrhovi u periodogramu i odgovaraju koeficijentima s najvećom veličinom. Kada se u daljnjem tekstu spominju najbolji ili najveći Fourierovi koeficijenti, misli se na one koji imaju najveći vrh u periodogramu.

### **3.5. Iz vremenske u frekvencijsku domenu**

Predlaže se prebacivanje podataka iz vremenske u frekvencijsku domenu kako bi se primjenom tehnika za uklanjanje šuma, smanjila količina podataka potrebna za uspješno modeliranje te eventualno otkrile određene periodičnosti karakteristične za pojedine studentske grupe. To su periodičnosti koje u vremenskoj domeni nisu vidljive, a za koje se pretpostavlja da su ključ mogućeg modeliranja.

EDM je kao novo istraživačko područje, otvoreno primjeni metoda koje su do sada bile najčešće korištene u prirodnim znanostima. Primjer toga je spektralno učenje kao novi pristup otkrivanju znanja (Falakmasir i sur., 2013), gdje je spektralni algoritam primijenjen kako bi se poboljšalo vrijeme podešavanja parametra otkrivanja znanja, a da se pri tom zadrži ista točnost predikcije pri korištenju skrivenog Markovljevog modela za određivanje studentskog znanja u adaptivnim edukacijskim sistemima i tutorima. Warnakulasooriya i Galen (2012) su uveli pojam fraktalne dimenzije jer su zaključili da studentski odgovori u tutorskom sistemu, kada su promatrani kao vremenski nizovi imaju karakteristike nasumičnog ili Brownovog gibanja. Ti su primjeri dali poticaj za primjenu FT na vremenske nizove studentskih logova u LMS.

### **3.6. Sažimanje podataka i uklanjanje šuma**

Kao što se u tehnologiji mogu poboljšati snimke zvuka ili fotografije uklanjanjem Fourierovih koeficijenata iz frekvencijske domene koji se mogu povezati sa šumom (Reis i sur., 2009), ovdje



se predlaže smanjenje broja FT koeficijenata iz skupa podataka na način da se pri tom sačuva ili čak poboljša informacija o uspjehu studenta sadržana u vremenskim nizovima (Vlachos i sur., 2004).

Važno je pitanje koji su to FT koeficijenti koji se važni za predviđanje, odnosno otkrivanje sličnosti. Neki autori koriste prvih  $k$  FT koeficijenata (Agrawal i sur., 1993) no za vremenske nizove s jakom periodičnošću preporuča se korištenje  $k$  najboljih FT koeficijenata s obzirom na to da se najveća snaga ne nalazi samo u niskim frekvencijama nego je prisutna u cijelom spektru (Vlachos i sur., 2004). Rad Agrawal i sur. (1993) se bavi brzim otkrivanjem sličnosti kod vremenskih nizova na izmišljenom skupu podataka koji nema veze s edukacijskim okruženjem, a rad autora Vlachos i sur. (2004) se bavi otkrivanjem sličnosti i periodičnosti u upitima na online tražilici.

### **3.7. Klasifikacija**

Imamo li skup podataka (instanci) i dodjeljujemo li svaki novi slučaj nekoj od unaprijed definiranih klasa, a temeljem promatranih atributa, govorimo o klasifikaciji. Kod predviđanja studentske uspješnosti koriste se različite tehnike ovisno o tome je li zavisna varijabla kontinuirana vrijednost ili je kategorijska vrijednost. Ako je zavisna varijabla kategorijska vrijednost, radi se o klasifikaciji, a ako je zavisna varijabla kontinuirana vrijednost radi se o regresiji (Hämäläinen i Vinni, 2011). U ovom radu radi se o kategorijskim klasama (2 i 3 klase), pa govorimo o klasifikaciji.

U situaciji nadziranog učenja tj. kada točno znamo koliko je klasa i utvrđujemo pravila temeljem kojih možemo klasificirati nove instance u postojeće klase, radi se o klasifikaciji (Duda i sur., 2000). Kod nenadziranog učenja, tj. kada slučajeve grupiramo u skupine sličnih slučajeva, govorimo o klasteriranju (Jain i sur., 1999). S obzirom na to da se u ovom radu opisuje proces nadziranog učenja, i po ovom kriteriju radi se o klasifikaciji.

### **3.8. Analiza glavnih komponenti**

Cilj preliminarnog istraživanja bio je provjeriti intuitivnu pretpostavku da se informacija o uspješnosti studenta može dobiti iz vremenskog uzorka studentovih logova na LMS. Analiza glavnih komponenti (engl. *Principal Component Analysis*, PCA) je primijenjena na podatke o

ocjenama na ispitu iz jednog hibridnog predmeta (Napredne WEB tehnologije i servisi) i vremenskom slijedu studentskih pristupa na taj predmet na LMS-u, tijekom trajanja tog predmeta u akademskoj godini 2012./2013. Analizirani vremenski niz pokazao je da se vremenski niz logova na LMS razdvaja prema kategorijama studentske uspješnosti.

PCA provedena je pomoću Matlaba uz dodatak PLS Toolbox (7.0.3. (2013) Eigenvector Research, Inc., Wenatchee, WA, USA).

## 4. MODELIRANJE

### 4.1. Prethodna obrada podataka

Tipovi studenata koji studiraju na Medicinskom fakultetu i Fakultetu organizacije i informatike se razlikuju, a razlikuju se i po frekvenciji pristupa LMS-u. Jedino što je zajedničko je postojanje i korištenje LMS-a u klasičnoj nastavi. Od studenata medicine se pri upisu traže dobre ocjene iz fizike, kemije i biologije, a studenti informatike su više tehnološki orijentirani i kod kojih se traže bolje ocjene iz matematike. Očekivano, studenti Fakulteta organizacije i informatike imaju puno veću frekvenciju pristupa LMS-u od studenata medicine.

Predmet Fizika i biofizika (FIB) predaje se na prvoj godini studija Medicine. Upisuje ga relativno velik broj studenata (308 u 2012., a 310 u 2013. godini), a studentima je putem LMS-a dostupan velik broj raznovrsnih materijala: opće informacije, predavanja u pdf formatu, video predavanja, animacije, simulacije, skripta (engl. *textbook*), formativni testovi za samoprovjeru, ishodi učenja, ankete (ugrađene Moodle ankete). Nastava je organizirana u obliku predavanja, seminara i vježbi koji se održavaju u učionicama i laboratorijima te se isprepliće s e-nastavom u formi dodatnih materijala za učenje, online kvizova i zadataka. Predmet traje približno četiri mjeseca. Drugi predmet obuhvaćen ovim istraživanjem predaje se na četvrtoj godini studija medicine (Fizika medicinske dijagnostike, FMD) i također ga upisuje velik broj studenata (255 u 2012., a 269 u 2013. godini), ali traje samo jedan tjedan i nema tako velik broj nastavnih materijala na LMS-u. Radi se također o hibridnom okruženju za učenje. Sve što je dostupno na LMS-u studenti koriste samo ako to žele.

Na Fakultetu organizacije i informatike obrađeni su podaci o predmetu preddiplomskog studija, WEB dizajn i programiranje (WDIP) i o predmetu diplomskog studija Napredne WEB tehnologije i servisi (NWTIS). Predmete upisuje manji broj studenata nego predmete na Medicinskom fakultetu. Predmet WEB dizajn i programiranje odslušalo je 80 studenata u 2012., a 84 studenata u 2013. godini. Predmet Napredne WEB tehnologije i servisi odslušalo je 64 studenta u 2012., a 51 student u 2013. godini. Za predmet WEB dizajn i programiranje studentima je na LMS-u dostupan velik broj raznovrsnih materijala: opće informacije, predavanja u pdf formatu, programski primjeri s predavanja, dodatni izvori u pdf formatu, informacije o prisustvovanju na predavanjima, seminarima i laboratorijskim vježbama, formativni testovi za samoprovjeru, ishodi učenja, ankete (ugrađene Moodle ankete), wiki za seminare, forumi, zadaće, projektni zadaci, poveznice na korisne web sadržaje. Nastava je

organizirana u obliku predavanja, seminara i vježbi koji se održavaju u učionicama i laboratorijima te se isprepliće s e-nastavom u formi dodatnih materijala za učenje, online kvizovima, forumima (za zadaće i projekt) i zadacima. Za predmet Napredne WEB tehnologije i servisi studentima je na LMS-u dostupan također velik broj raznovrsnih materijala: opće informacije, predavanja u pdf formatu, programski primjeri s predavanja, programski primjeri s vježbi, video materijali s vježbi, dodatni izvori u pdf formatu, informacije o prisustvovanju na predavanjima i laboratorijskim vježbama, formativni testovi za samoprovjeru, ishodi učenja, ankete (ugrađene Moodle ankete), forumi, zadaće, projektni zadaci, poveznice na korisne web sadržaje. Nastava je organizirana u obliku predavanja i vježbi koji se održavaju u učionicama i laboratorijima te se isprepliće s e-nastavom u formi dodatnih materijala za učenje, online kvizovima, forumima (za zadaće i projekt) i zadacima. Usprkos manjem broju upisanih studenata, ukupan broj logova na LMS-u po predmetu je usporediv s ukupnim brojem logova na predmete s Medicinskog fakulteta što se može objasniti činjenicom da su studenti Fakulteta organizacije i informatike više usmjereni na korištenje informacijsko-komunikacijskih tehnologija te da je nastava koncipirana tako da su neki sadržaji na LMS-u obavezni za sve studente (zadaće i projektni zadaci). U tablicama 3 i 4 prikazani su usporedni podaci o elementima i aktivnostima na promatranim predmetima, njihovom broju i broju dnevničkih zapisa.

Iz tablica 3 i 4 vidljivo da se studenti služe LMS-om uglavnom kako bi nešto pogledali. Na svim promatranim predmetima i akademskim godinama, u 98 do 100% slučajeva zabilježena je akcija pogledaj (engl. *view*). Najčešće se pristupa sljedećim elementima: opće informacije, predavanja u pdf formatu, programski primjeri s predavanja, programski primjeri s vježbi, video materijali s vježbi, dodatni izvori u pdf formatu, video predavanja, animacije, simulacije, skripta (73-82% ukupnih pristupa nekom predmetu). Na predmetima Fizika i biofizika te Fizika medicinske dijagnostike je postotak pristupa navedenim elementima nešto viši (77-82%) nego postotak pristupa spomenutim elementima na predmetima Napredne WEB tehnologije i servisi te WEB dizajn i programiranje (73-75%). Pristup forumima varira od 3-12% ovisno o predmetu i promatranoj akademskoj godini. Predmeti Napredne WEB tehnologije i servisi te WEB dizajn i programiranje imaju veći postotak pristupa materijalima koji se odnose na zadaće i projektne zadatke (10-17%), a predmeti Fizika i biofizika te Fizika medicinske dijagnostike imaju veći broj pristupa kvizovima za samoprovjeru (FIB-6-9%, FMD-visokih 14-18%).

Tablica 3: Usporedni podaci o elementima i aktivnostima na promatranim predmetima, njihovom broju i broju dnevničkih zapisa za akademsku godinu 2012./2013.

Element	Aktivnost	Fizika i biofizika		Fizika medicinske dijagnostike		Napredne WEB tehnologije i servisi		WEB dizajn i programiranje	
		Br. akt./element	Broj logova	Br. akt./element	Broj logova	Br. akt./element	Broj logova	Br. akt./element	Broj logova
assign	submit					2	360	2	448
assign	view					4	5826	4	7351
blog	view					1	1	1	7
choice	choose					2	171	2	242
choice	view					1	1135	1	2192
course	report log	6	180			1	55	2	3
course	view	1	59806	1	6825	2	53242	2	68164
discussion	mark read	1	19	1	3				
feedback	startcomplete					1	343	1	436
feedback	submit					1	260	1	408
feedback	view					2	953	2	1771
forum	add discussion	2	61	1	24	2	68	2	8
forum	delete discussion	2	23	1	2			1	1
forum	mark read	1	12	1	1				
forum	search	1	16			1	5	1	8
forum	update post	2	19	1	2	2	24	1	5
forum	user report	1	6					1	1
forum	view discussion	3	15590	3	1942	3	13181	3	5033
page	view	3	7885	2	729	1	8157	1	2916
quiz	attempt	2	846	2	347	1	502	1	552
quiz	close attempt	1	815	1	325	1	482	1	546
quiz	continue attempt	1	1217	1	490	1	2467	1	2772
quiz	view	8	6619	8	2462	2	1683	2	1901
resource	view	4	11593	3	6997	1	9617	1	11054
survey	view	6	1758						
wiki	edit							5	1047
wiki	view							1	3791
workshop	add assessment					2	429	2	372
workshop	add submission					3	86	2	100
workshop	view					2	2941	2	3492
<b>Ukupno</b>			<b>106465</b>		<b>20149</b>		<b>101988</b>		<b>114621</b>

Tablica 4: Usporedni podaci o elementima i aktivnostima na promatranim predmetima, njihovom broju i broju dnevničkih zapisa za akademsku godinu 2013./2014.

Element	Aktivnost	Fizika i biofizika		Fizika medicinske dijagnostike		Napredne WEB tehnologije i servisi		WEB dizajn i programiranje	
		Br. akt./element	Broj logova	Br. akt./element	Broj logova	Br. akt./element	Broj logova	Br. akt./element	Broj logova
assign	submit					1	73	1	328
assign	view					3	5724	4	5343
blog	view					1	1	1	4
choice	choose							1	36
choice	view							1	1163
course	report log	1	6			1	37	2	10
course	view	1	58402	2	8184	2	42266	2	72500
discussion	mark read	1	13					1	2
feedback	submit					1	228	1	600
feedback	view					1	838	1	2317
forum	add discussion	1	24	1	17	2	46	2	35
forum	delete discussion	1	3	1	1	1	1		
forum	mark read			1	1			1	1
forum	search	1	19			1	10	1	10
forum	update post	2	7	1	5	3	33	3	90
forum	view discussion	3	12332	3	2012	3	8958	3	11054
page	view	2	5439	1	127	1	5883	1	2274
quiz	attempt	1	607	1	304	1	191	1	330
quiz	close attempt	1	591	1	292	1	191	1	330
quiz	continue attempt	1	766	1	412	1	1348	1	2302
quiz	view	6	3584	6	1813	2	697	2	1264
resource	view	2	16505	2	7988	1	7471	1	17179
survey	view	5	1025						
wiki	edit							7	907
wiki	view							1	4387
workshop	add assessment					3	332		
workshop	add submission					2	23		
workshop	view					2	1565		
<b>Ukupno</b>			<b>99323</b>		<b>21156</b>		<b>75916</b>		<b>122466</b>

Izbor predmeta i studenata koji se međusobno razlikuju omogućit će provjeru je li predloženi način modeliranja primjenjiv u različitim okruženjima, odnosno je li dovoljno robusan. Podaci o svim predmetima i inicijalnim skupovima podataka nalaze se tablici 1 (za 2012./2013.) i tablici 2

(za 2013./2014.) u dijelu 3.1.

U dnevničkim zapisima (log) LMS-a čuvaju se izvještaji o aktivnostima sudionika. Studentski log je izvještaj o aktivnosti na LMS-u: kojim stranicama je i kada student pristupao, s koje IP adrese i koje su bile aktivnosti (pogledaj, dodaj, osvježi i izbriši). Svaka aktivnost je zabilježena kao jedan klik. Riječi klik, log, pristup i dnevnički zapis ovdje se koriste naizmjenično i s istim značenjem. Za svakog studenta kreira se vremenski niz koji sadrži broj i vremenski slijed njegovih pristupa određenom predmetu na LMS-u za vrijeme trajanja tog predmeta.

Model je konstruiran tako da je zavisna varijabla uspjeh studenta na završnom ispitu iz određenog predmeta. Svi opisani predmeti ocjenjuju se ocjenama od 1 (nedovoljan) do 5 (izvrstan). Jedino se predmet Fizika medicinske dijagnostike ocjenjuje s *prošao/pao*. Numerički podaci za zavisnu varijablu podijeljeni su u kategorijske klase u svrhu povećanja interpretabilnosti i razumijevanja (Romero, Ventura i García, 2008) i to na 2 načina: u dvije i u tri kategorije. Kategorije su formirane na sljedeći način: dvije kategorije (Dobri (*D*) – ocjene 3,4,5 i Loši (*L*) – ocjene 1 i 2), a tri kategorije (Vrlo dobri (*VD*) – ocjene 4 i 5, Prosječni (*P*) – ocjena 3, Loši (*L*) – ocjene 1 i 2). Kategorije su formirane na opisan način kako bi distribucija ispitanika po kategorijama bila što ujednačenija. S druge strane, takvo formiranje kategorija može predstavljati izvjesno ograničenje o kojem treba voditi računa kod diskusije rezultata. Smještanje studenata s ocjenama 1 i 2 istu kategoriju ima svoje opravdanje, jer i studenti s ocjenom dovoljan predstavljaju rizične studente i potrebna im je pomoć (Romero i sur., 2013). Broj studenata unutar pojedinih kategorija prikazan je u tablicama 1 i 2 i ujednačen je.

Nezavisne varijable (vremenski nizovi pristupa LMS-u) su iz vremenske domene transformirane u frekvencijsku domenu. Vrijeme kumulacije pristupa (period u kojem se signal mjeri) je 1 sat i 1 dan.

Podaci o pristupima prikupljenim u LMS-ovoj tablici nisu odmah prikladni za dubinsku analizu već je potrebno podatke prethodno obraditi. Ukratko, u fazi prethodne obrade podataka prilagođavaju se podaci iz tablice LMS-ove baze pristupa studenata pojedinim resursima promatranog predmeta tako da se dobiju vremenski nizovi pristupa (logova) po pojedinom studentu za određeni predmet. Vremenski nizovi se zatim prebacuju u frekvencijsku domenu primjenom DFT. Iz dobivenih periodograma odabire se određeni broj (npr. 20) najintenzivnijih (najvećih) koeficijenata po studentu pomoću kojih se izradi zajednička tablica najintenzivnijih Fourierovih koeficijenata koja se koristi za modeliranje. Navedeni postupak bi se ukratko mogao

nazvati uklanjanjem šuma i otkrivanje karakteristične periodičnosti koja nije bila vidljiva u vremenskoj domeni.

Podrobnije, podaci u LMS-ovoj tablici sadrže podatke o točnom vremenu svakog klika svih studenata upisanih na promatrani kolegij. Međutim, da bi se ti podaci mogli pretvoriti u vremenski slijed potrebna je dodatna obrada.

Prvi korak u dodatnoj obradi je određivanje vremena kumulacije  $T$ . Kao što je opisano u dijelu 3.3.,  $T$  je period u kojem se zbrajaju svi logovi određenog studenta. Ovdje su korištena dva vremena kumulacije,  $T_{sat}=1$  sat i  $T_{dan}=1$  dan. Potom se originalni podaci o pristupima dobiveni iz LMS-ove baze prebacuju u lokalnu MySQL bazu (MySQL 5.6.17, Oracle Corporation, Redwood City, California) te se stvaraju tri nove tablice po predmetu za određenu godinu. Dvije tablice pristupa s različitim vremenima kumulacije i jedna s uspjehom studenata. Tablica za npr. predmet Fizika i biofizika prenesena je iz LMS-a (*FIBpr*), a tablice *FIBprpodanu* i *FIBprposatu* su preračunate tablice za dva odabrana vremena kumulacije. Tablica *uspjehFIB* sadrži podatke o studentskom uspjehu na predmetu Fizika i biofizika u određenoj godini.

Prema odabranom vremenu kumulacije i trenutku početka promatranog predmeta iz tablice kakva je prikazana na slici 1 pune se nove tablice (*FIBprposatu* i *FIBprpodanu*) u kojima je vrijeme pristupa pretvoreno u sate (ili u drugom slučaju u dane) protekle od početka predmeta (Prilozi, kod 1).

Nova tablica ima tri kolone: indeks, sat (dan) kad je kliknuto i identifikacijski broj studenta. Novonastalu tablicu treba grupirati po satima (danima) i studentima (Prilozi, kod 2). Međutim, to još uvijek nisu vremenski nizovi pogodni za primjenu FFT.

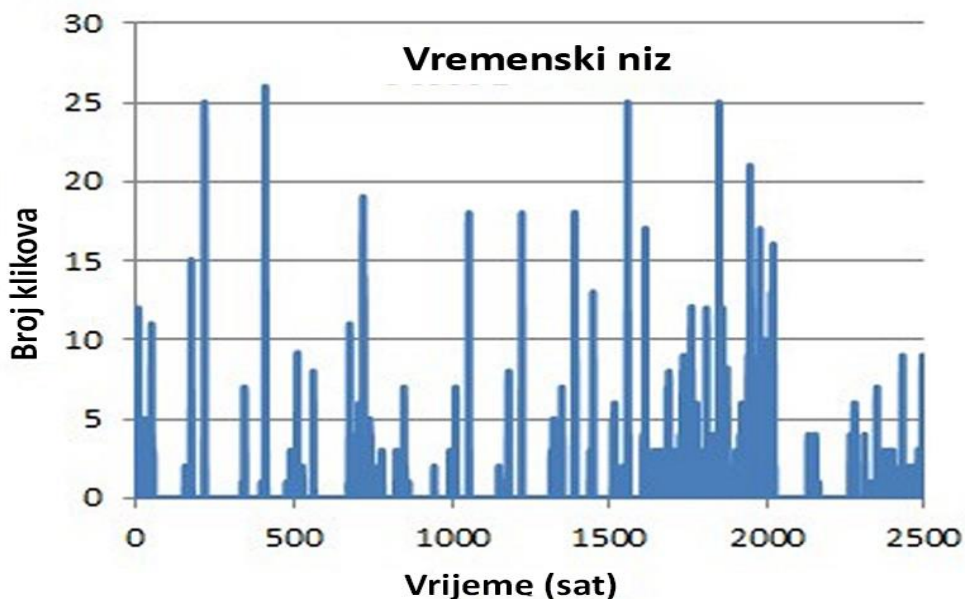
Iz MySQL-a prebacujemo tablicu o uspjehu studenata i tablicu s pristupima po studentu po danu/satu, u Excel kao međukorak za unos u Matlab u kojem se radi FT. Međukorak je napravljen iz tehničkih razloga, unos podataka u Matlab preko Excela bio je najpouzdaniji.

Za transformiranje podataka o logovima u vremenske nizove, a zatim u periodograme korištene su kratke skripte napisane u Matlabu (Matlab 7.11.0 (2010b), The MathWorks, Inc, Natick, Massachusetts, USA). Nakon prebacivanja iz Excela u Matlabu su nastale dvije matrice za svaki predmet. Jedna s vremenima pristupa za sve klikove svakog studenta, a druga sa studentskim rezultatima.



Iz tih matrica formira se tablica (matrica) s vremenskim nizom za svakog studenta. Svaka kolona predstavlja vremenski niz klikova jednog studenta. Studenti ne pristupaju LMS-u ravnomjerno. Na primjeru vremenskog niza s vremenom kumulacije od jednog sata to izgleda ovako: npr. student klikne na neki sadržaj u prvom, devetom, tridesetšestom satu itd. Vremenski niz koji tako nastaje je neravnomjerno uzorkovan. Za FFT je, međutim, nužno ravnomjerno uzorkovanje. U vremenskom nizu potreban je podatak za svaki sat. Zbog toga je generirana matrica  $m \times n$  ( $m$  je broj studenata, a  $n$  dužina vremenskog niza) koja je ispunjena nulama. Potom su nule zamijenjene točnim podatkom o broju klikova u određenom satu za određenog studenta. Tek je takav, ravnomjerno uzorkovan vremenski niz, moguće iskoristiti za FFT (Prilozi, kod 3 u Matlabu).

Na slici 3 prikazan je vremenski niz logova za jednog, nasumično odabranog, studenta. Kada je kao  $T$  odabran sat, broj točaka u vremenskom nizu za jednosemestralni predmet je približno 2800, a kada je kao  $T$  odabran dan, broj točaka je približno 120. Nizovi poput gore opisanih, kreirani su za svakog studenta koji je upisan u promatrane predmete.



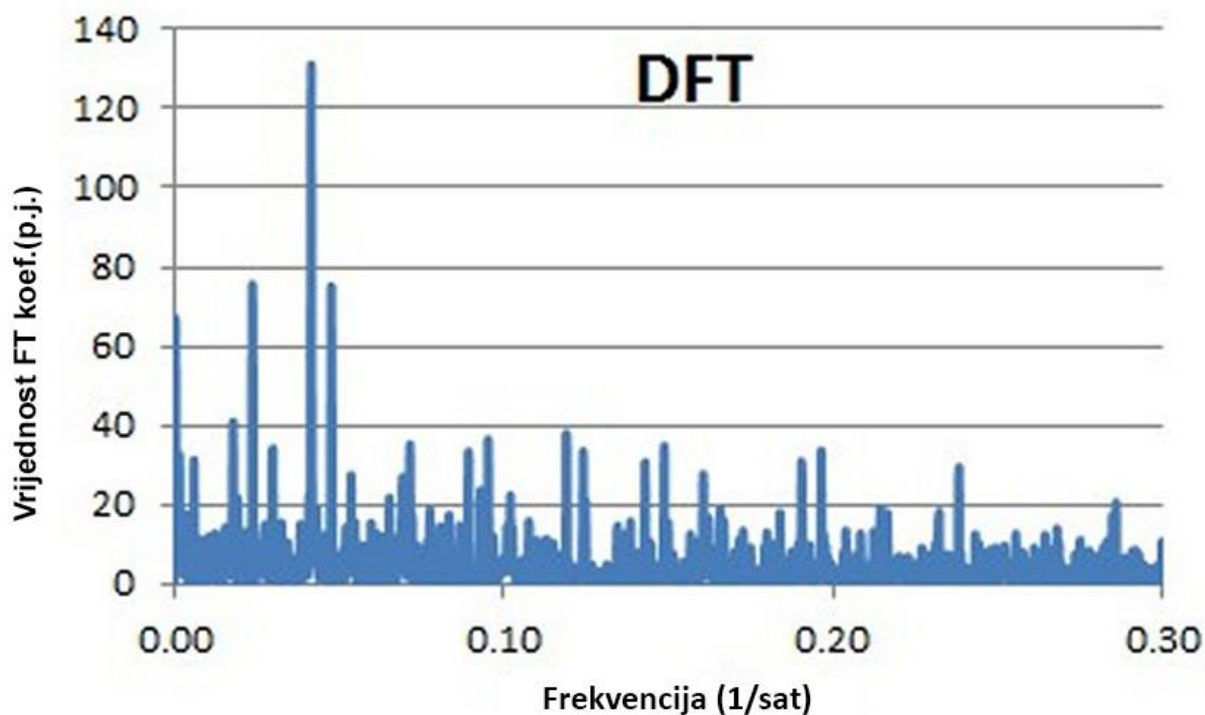
*Slika 3: Vremenski niz logova za određenog studenta*

*Izvor: Gamulin, Gamulin i Kermek (2016)*

Sljedeći korak u obradi podataka je računanje DFT za sve generirane vremenske nizove i izrada pripadajućih periodograma. Broj Fourierovih koeficijenata u periodogramu jednak je broju

točaka u prvotnom vremenskom nizu. Slijedi računanje FT koeficijenata u Matlabu, izdvajanje 20 najvećih koeficijenata po studentu i izrada zajedničke tablice najvećih koeficijenata za cijeli predmet (Prilozi, kod 4).

Primjerak periodograma za jednog studenta prikazan je na slici 4.



Slika 4: Primjerak periodograma, odnosno Fourierovih koeficijenata za jednog studenta

Izvor: Gamulin, Gamulin i Kermek (2016)

Iz svakog periodograma odabere se 20 najintenzivnijih koeficijenata za svakog studenta. Kreirani skupovi od po 20 najintenzivnijih koeficijenata i pripadajućih frekvencija za svakog studenta ubacuju se u novu tablicu koja sadrži po 20 najintenzivnijih koeficijenata i pripadajućih frekvencija za sve studente upisane u predmet. U tablici su u prvoj koloni navedene sve moguće frekvencije, a u kolonama za svakog pojedinog studenta su navedeni pripadajući koeficijenti (iz skupa 20 najintenzivnijih za svakog studenta). Ako za neku frekvenciju nema pripadajućeg koeficijenta, polje se popunjava nulom. U zadnjoj koloni prikazan je broj koeficijenata (koji nisu nula) u određenom retku.

Tablica se zatim sortira po vrijednostima iz zadnje kolone od najveće prema najmanjoj vrijednosti te se izradi zajednička lista najintenzivnijih koeficijenata koji se pojavljuju u skupovima podataka studenata. U konačnoj listi prikazani su najzastupljeniji i najintenzivniji koeficijenti za promatrani predmet. Konačne tablice za promatrane predmete imaju od 73 do 403 koeficijenta.

Na kraju su podaci koji su u Matlabu priređeni za obradu, učitani u RapidMiner (RapidMiner 7.1) pomoću Excela (Prilozi, kod 5).

Napravljena je matrica uspjeha studenata i najintenzivnijih Fourierovih koeficijenata. Uspjehu studenata se dodijeli uloga zavisne/ ciljne varijable i tada su podaci spremni za modeliranje.

Na ovaj način formirani su iz 13 inicijalnih skupova varijabli za svaku od promatrane dvije godine (tablica 1 i tablica 2) skupovi podataka za modeliranje koji odgovaraju različitim dužinama vremenskih nizova. Za vrijeme kumulacije od jednog sata formirano je osam skupova s kraćim vremenskim nizovima (1000, 1400, 1800, 2000, 2150, 2500, 2800 sati i skup s maksimalno mogućom dužinom u satima). Za vrijeme kumulacije od jednog dana formirano je šest skupova s kraćim vremenskim nizovima (60, 80, 100, 120, 150 dana i s maksimalno mogućom dužinom u danima)

Stvarni skupovi podataka za modeliranje generirani su iz gore opisanih skupova podataka tako da je postupno smanjivan broj Fourierovih koeficijenata počevši od maksimalnog broja koeficijenata, pa do četiri koeficijenta (dio 4.3.). Najprije su iz skupa podataka odstranjeni najmanje zastupljeni koeficijenti, a zatim postupno sve zastupljeniji koeficijenti dok na kraju nisu ostala četiri najzastupljenija koeficijenta. Tako je generirano i po 399 skupova podataka za modeliranje iz svakog inicijalnog skupa podataka. Budući da su vremenski nizovi skraćivani, formirano je 180 skupova podataka i za svaki od tih skupova provedeno je iterativno smanjivanje broja upotrijebljenih koeficijenata (broj iteracija je puni broj koeficijenata umanjen za 4). Korištena su 4 algoritma, s time da je SVM (metoda potpornih vektora, engl. *support vector machines*) korišten samo za binarnu klasifikaciju. Ukupno je izračunata točnost predikcije za 100919 modela.

## 4.2. Algoritmi i softver

Prema dostupnoj literaturi, ovi se algoritmi uobičajeno koriste za predikcijske modele (Hastie i sur., 2008): metoda potpornih vektora (engl. *support vector machines*, SVM), metoda  $k$  najbližih susjeda (engl. *k-nearest neighbors*, kNN), umjetna neuronska mreža (engl. *artificial neural network*, ANN) te naivni Bayesov klasifikator (engl. *naïve Bayes*, NB). Algoritmi su korišteni sa zadanim postavkama za alat RapidMiner (RapidMiner 7.1., RapidMiner GmbH, Dortmund, Germany). To znači da je metoda potpornih vektora koristila radijalnu jezgenu funkciju, parametar troška, a kernel gama je 1. Pri korištenju metode  $k$  najbližih susjeda, broj najbližih susjeda  $k$  je 30, a mjeri se euklidska udaljenost. Umjetne neuronske mreže širenja unatrag konfigurirane su u tri sloja s brojem čvorova u skrivenom sloju koji odgovara zbroju broja atributa i broja klasa, podijeljenim s 2, a zatim dodan 1. Aktivacijska funkcija je sigmoidnog oblika. Naivni Bayesov klasifikator je korišten uz Laplaceovu korekciju. U svim modelima korištena je *split* validacija sa slučajnim (engl. *shuffled sampling*) uzorkovanjem. 70% podataka je korišteno za izgradnju modela, a 30% za validaciju. Odabir atributa je optimiziran pomoću genetskog algoritma jer je pokazano da je točnost predikcije veća uz primjenu takve optimizacije (Oreški i sur., 2012; Gamulin i sur., 2014). Pri korištenju genetskog algoritma kao selekcijska shema korišten je rulet (engl. *roulette wheel*), a tip križanja (engl. *crossover type*) je jednolik (engl. *uniform*): Težine su normalizirane, vjerojatnost uključivanja nekog atributa (*p initialize*) je 0,5, vjerojatnost promjene nekog atributa (engl. *p mutation*) je -1,0 (odabrana vrijednost parametra -1,0 znači da je vjerojatnost promjene nekog atributa  $1/n$ , gdje je  $n$  ukupan broj atributa), vjerojatnost odabira za križanje (engl. *p crossover*) je 0,5. Parametar minimalni broj atributa je 1, parametar veličina populacije je 5, maksimalni broj generacija je 30, a maksimalna prikladnost (engl. *maximal fitness*) je postavljena na beskonačno.

U radu je korišteno nekoliko algoritama često korištenih u literaturi kako bi se pokazalo da svi ti algoritmi daju točnost veću od slučajnog pogađanja i usporedivu s rezultatima drugih modela.

Primjena 4 algoritma bila je vremenski vrlo zahtjevna jer je proveden izuzetno velik broj izračuna. Pojedini izračuni potrajali su i po nekoliko sati. To je razlog za primjenu jednostruke validacije na podskupu od 30% podataka.

#### Algoritmi korišteni u radu:

- metoda potpornih vektora (SVM) – Kotsiantis i sur. 2004; Fawcett, 2006; Lara i sur., 2014; Lykourantzou i sur., 2009; Thai-Nghe i sur., 2009; Kotsiantis i sur., 2010, Gupta i sur., 2016.
- metoda k najbližih susjeda (kNN) – Kotsiantis i sur., 2004; Kotsiantis i sur., 2010.
- umjetna neuronska mreža (ANN) – Jovanović i sur., 2012; Kotsiantis i sur., 2004; Delgado i sur., 2006; Paliwal i Kumar, 2009; Lara i sur., 2014; Lykourantzou i sur., 2009; Kotsiantis i sur., 2010, Gupta i sur., 2016.
- naivni Bayesov klasifikator (NB) – Jovanović i sur., 2012; Kotsiantis i sur., 2004; Thai-Nghe i sur., 2009; Kotsiantis i sur., 2010.

#### Softver korišten u radu:

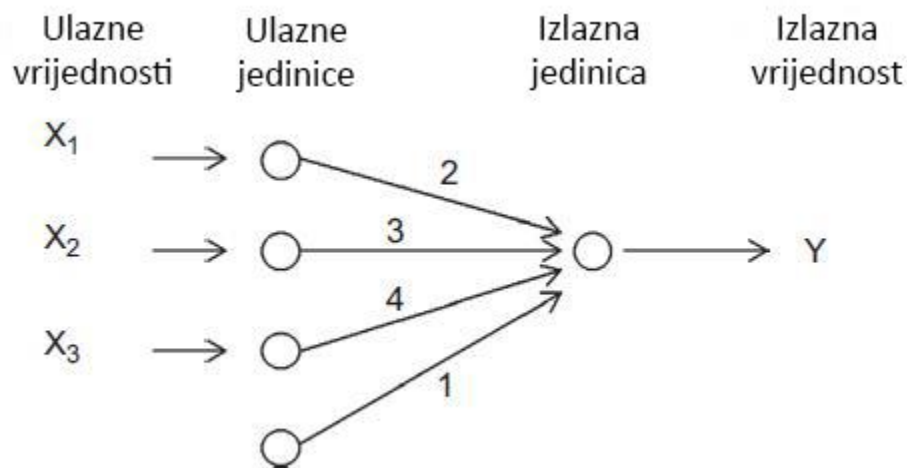
- MySQL 5.6.17 (Oracle Corporation, Redwood City, California, USA) za dio prethodne obrade podataka;
- Matlab 7.11.0 (2010b) (The MathWorks, Inc, Natick, Massachusetts, USA) za dio prethodne obrade podataka;
- RapidMiner 7.1. (RapidMiner GmbH, Dortmund, Njemačka) za modeliranje;
- Matlab uz dodatak PLS Toolbox (7.0.3. (2013) Eigenvector Research, Inc., Wenatchee, WA, USA) za PCA za preliminarno istraživanje tj. je li moguće odvojiti različite kategorije studenata;
- Excel, Office 2010 (Microsoft, Redmond, Washington, USA) je korišten za učitavanje podataka u Matlab u i RapidMiner jer je iskustvo pokazalo da se podaci tako učitavaju s najmanje problema;
- TeamViewer, Version 11 (Teamviewer GmbH) je korišten za upravljanje udaljenim računalima.

### 4.2.1. Umjetna neuronska mreža

Umjetna neuronska mreža je računski i matematički model nadahnut biološkim živčanim sistemom (Kotu i Deshpande, 2014, str 124). Veza između ulaznih i izlaznih varijabli se modelira razvijajući matematičko objašnjenje vrlo slično biološkom funkcioniranju živčane stanice. Za jednostavan linearni matematički model

$$Y=I+2X_1+3X_2+4X_3$$

gdje je  $Y$  ciljna varijabla, a  $X_1$ ,  $X_2$  i  $X_3$  atributi, 2, 3 i 4 su koeficijenti atributa (težine)  $X_1$ ,  $X_2$  i  $X_3$ , topologija modela izgleda kao na slici 5.

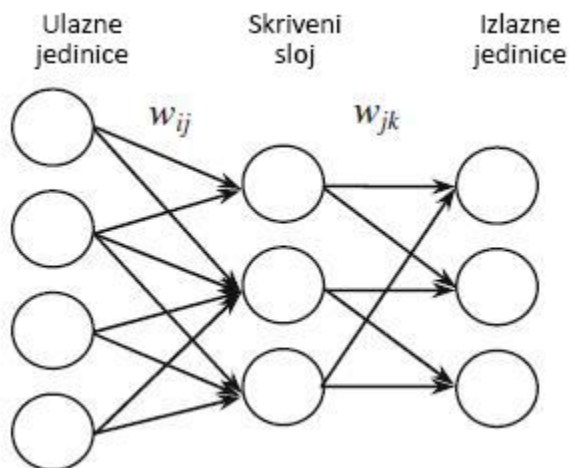


Slika 5: Topologija modela umjetne neuronske mreže

Izvor: Kotu i Deshpande, 2014, str. 124

Umjetna neuronska mreža može imati više slojeva, a prvi sloj najbliži ulazu je ulazni sloj i on prima informaciju koja treba biti procesuirana, a sloj najbliži izlazu je izlazna jedinica i tamo se nalaze rezultati obrade. Između se nalaze skrivene jedinice (čvorovi ili neuroni). Kada se signal kreće u samo jednom smjeru, bez petlji, to je umjetna neuronska mreža s propagacijom unaprijed. Modeliranje složenih, nelinearnih veza moguće je uz skrivene slojeve u topologiji (slika 6). Aktivacijska funkcija u izlaznom sloju se sastoji od agregacijske funkcije koja obično sumarizira i transferne funkcije koja služi za linearnu i nelinearnu transformaciju. Zbog brojnih mogućnosti odabira topologije, skrivenih slojeva i transferne funkcije, pronalaženje optimalnog rješenja za umjetnu neuronsku mrežu prilično je zahtjevno.

Propagacija greške unatrag je najpoznatiji algoritam kojim umjetna neuronska mreža uči o vezi između ulaznih atributa i izlazne zavisne varijable. Procjenjuje se pogreška između procijenjenog i stvarnog izlaza. Ta se greška koristi da se prilagode težine veza kako bi greška u sljedećim izračunima bila manja.



Slika 6: Umjetna neuronska mreža s propagacijom unaprijed i skrivenim slojem

Izvor: Kotsiantis, Zaharakis i Pintelas, 2006

Pravilo za korigiranje težina je

$$\Delta W_{ji} = \eta \delta_j O_i$$

$\eta$  je pozitivan broj koji se zove stopa učenja, a određuje veličinu koraka u postupnoj potrazi. Velik broj omogućuje bržu propagaciju unatrag, ali i povećava vjerojatnost da se cilj nikad ne dostigne.

$O_i$  je izlazna vrijednost koju je izračunala izlazna jedinica  $i$

$\delta_j = O_j(1 - O_j)(T_j - O_j)$  za izlazne jedinice, gdje je  $T_j$  željena izlazna vrijednost za izlaznu jedinicu  $j$  i

$\delta_j = O_j(1 - O_j) \sum_k \delta_k w_{kj}$  za skrivene jedinice

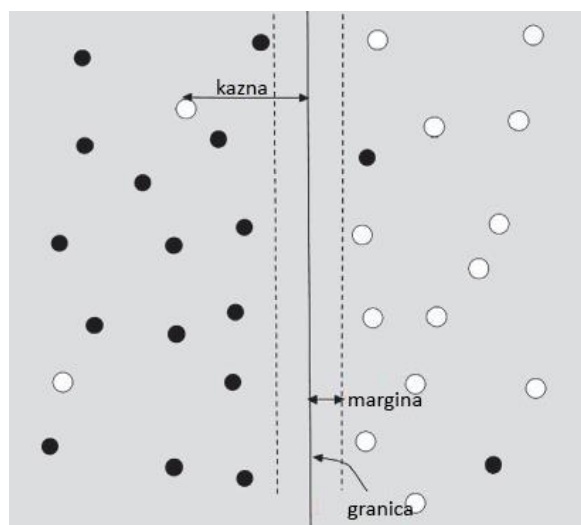
(Kotsiantis i sur., 2006)

Općenito se drži da je bolje imati previše skrivenih jedinica nego premalo. S premalo skrivenih jedinica može se dogoditi da model nema dovoljno fleksibilnosti da uhvati nelinearnost u

podacima, a s previše skrivenih jedinica, težine mogu kolabirati u nulu. Uobičajeno je da broj skrivenih jedinica bude između 5 i 100, a broj se povećava s brojem ulaza i brojem slučajeva za učenje. Ako se ne specificira broj skrivenih slojeva, koristi se predefiniрани skriveni sloj, s brojem neurona koji se računa tako da se zbroj broja atributa i broja klasa podijeli s 2 i zatim se doda 1. Aktivacijska funkcija je sigmoidnog oblika. Broj ciklusa učenja je 5000 po zadanim postavkama. Stopa učenja (engl. *learning rate*) je vrijednost od 0 do 1, a određuje koliko osjetljiva mora biti promjena u težini pri ispravku pogreške. Ako je vrijednost bliže nuli znači da će ispravljena vrijednost biti temeljena više na prethodnoj težini, a manje na ispravci pogreške. I obrnuto, vrijednost bliža 1 bit će uglavnom temeljena na ispravci pogreške. U radu je korištena stopa učenja u vrijednosti od 0,3 po zadanim postavkama. Momentum je vrijednost koja smanjuje rizik upadanja u lokalne minimume i nastoji pronaći globalno optimalne rezultate. U radu je korišten momentum u vrijednosti od 0,2 po zadanim postavkama.

#### 4.2.2. Metoda potpornih vektora

Baš kao i mnogo drugih tehnika nadziranog strojnog učenja, algoritmi potpornih vektora su relativno nov pojam, a jedan od prvih opisa je dao Cortes (1995), koji se bavi algoritmima za optičko prepoznavanje slova. Algoritam prilagođava granicu na skup točaka koje pripadaju istoj klasi. Najbolja je granica s najmanje loših klasifikacija i najvećom marginom tj. prosječnom geometrijskom udaljenošću između dvije klase (slika 7).

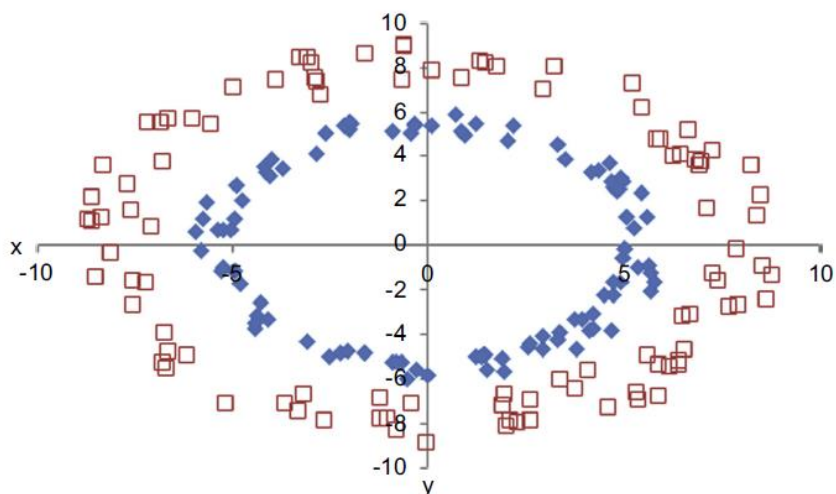


Slika 7: Temeljni pojmovi za SVM algoritam – granica, margina i kazna

Izvor: Kotu i Deshpande, 2014, str. 136

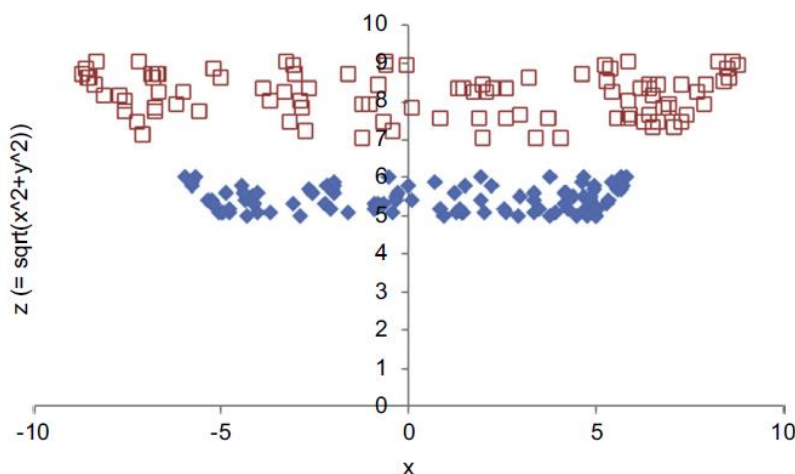


Rijetko je moguće pronaći podatke koji se mogu linearno razdvojiti. Kad se to i dogodi, mogu postojati točke koje se nalaze unutar margine. Tada se naplaćuje kazna za svaku točku unutar margine (parametar C), a kao najbolja granica odabire se ona koja ima minimalni trošak kazne. Kada se između dvije klase ne može povući ravna linija, koristi se jezgrena funkcija kojom se prostor podataka preslikava u prostor više dimenzije tako da klase koje u originalnom prostoru nisu linearno separabilne u prostoru s više dimenzija postanu linearno separabilne. (slike 8 i 9).



Slika 8: Klase koje se ne mogu linearno razdvojiti

Izvor: Kotu i Deshpande, 2014, str. 137



Slika 9: Transformacija u klase koje se mogu linearno razdvojiti

Izvor: Kotu i Deshpande, 2014, str. 137

Pri korištenju metode potpornih vektora upotrijebljena je radijalna jezgrena funkcija, kao jedna od najčešće korištenih u tu svrhu. Konstanta kompleksnosti  $C$  je 0 jer je preporuka da se krene s početnim postavkama RapidMinera. I ostali parametri su korišteni u predefiniranim vrijednostima, pa je tako korištena vrijednost jezgrena gama (engl. *kernel gamma*) je 1, jezgrena memorija (engl. *kernel cache*) je 200, epsilon konvergencije je 0,001, maksimalni broj iteracija je 100000,  $L_{\text{pos}}$  je 1 i  $L_{\text{neg}}$  je 1.

Metoda potpornih vektora je algoritam koji pronalazi višedimenzionalne granice koje razdvajaju točke koje pripadaju različitim klasama. Model je vektorska jednadžba, a svi atributi bi trebali biti numerički. Ciljna varijabla može biti kategorijska ili numerička. Algoritam nije sklon prenaučivosti, a male razlike u ulaznim podacima ne rezultiraju drugačijim rezultatima. Drži se dobrim algoritmom za nelinearne veze. Algoritam može biti spor u fazi učenja zbog optimiziranja kombinacija parametara. Obično se koristi za prepoznavanje teksta, otkrivanje krivotvorina, i modeliranje crnih labudova (engl. *black-swan*), tj. događaja koji iznenađuju promatrača, a imaju velik utjecaj (Kotu i Deshpande, 2014, str. 137).

Kada imamo podatke koji se mogu linearno razdvojiti, točke koje se nalaze na marginama ravnine koja optimalno razdvaja podatke, nazivamo točkama potpornih vektora, a rješenje je linearna kombinacija tih točaka. Ostali podaci se zanemare. Na složenost SVM modela ne utječe broj atributa, a broj potpornih vektora koje SVM algoritam odabire obično je mali. Zbog toga su SVM prikladni za situacije u kojim je broj atributa velik u odnosu na broj slučajeva (Kotsiantis i sur., 2006) Dobra je strana što SVM uvijek dostiže globalni minimum, za razliku od ANN. Međutim, SVM metode su binarne i ne mogu rješavati probleme s većim brojem klasa (Kotsiantis i sur., 2006).

### 4.2.3. Metoda $k$ najbližih susjeda

Osnovna ideja ovog algoritma opisana je starom izrekom: „Svaka ptica svome jatu leti“, odnosno slične skupine podataka se grupiraju u susjedstvu u  $n$ -dimenzionalnom prostoru i imaju iste ciljne klase. Cijeli skup podataka za učenje se pamti i svaki novi ulazni podatak se uspoređuje s cijelim skupom podataka za učenje kako bi se pronašla najbliža vrijednost. Ciljna vrijednost te najbliže vrijednosti za učenje je ujedno i ciljna vrijednost za novu ulaznu vrijednost za testiranje. Radi se o neparametrijskoj metodi koja ne traži formulu odnosa između atributa i ciljne varijable niti određenu distribuciju skupa podataka (Altman, 1992).

Cilj je kNN algoritma odrediti koja je točka za učenje najbliža određenoj točki za testiranje, koristeći se pri tom mjerom blizine. Kako bi se kvantificirala sličnost između dva sloga podataka koriste se tehnike kao što su udaljenost, korelacija, Jaccardova sličnost i kosinusna sličnost (Tan i sur., 2005). Najuobičajenija mjera udaljenosti je euklidska udaljenost (Larose i Larose, 2015, str. 302). Za  $n$ -dimenzionalni prostor, gdje je  $X$  jednako  $(x_1, x_2, \dots, x_n)$ , a  $Y$  je  $(y_1, y_2, \dots, y_n)$ , udaljenost  $d$  je

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

Mjeri se euklidska udaljenost između dvije točke, a  $k$  je broj najbližih susjeda, početne postavke u RapidMineru su  $k=30$ .

Metoda  $k$  najbližih susjeda je poznata kao „lijeni učenik“ kod kojeg nema generalizacije modela. Svi novi podaci se uspoređuju sa sličnim poznatim podacima u skupu za učenje. Cijeli skup podataka za učenje je model. Nema nekih ograničenja kakvi bi trebali biti ulazni podaci, no rezultati su bolji s numeričkim podacima. Podaci trebaju biti normalizirani jer na rezultat jako utječu veličine i mjerne jedinice atributa. Ciljna varijabla je kategorijska (Kotu i Deshpande, 2014). Za izgradnju modela ne treba puno vremena, ali je potrebno relativno više vremena za klasifikaciju (Kotsiantis i sur., 2006). Dobro se nosi s podacima koji nedostaju. Pogodan je za nelinearne veze (Kotu i Deshpande, 2014). Odabir vrijednosti  $k$  je proizvoljan. Za točnije rezultate koristi se nekoliko težinskih shema koje mijenjaju mjere udaljenosti i glasaju (engl. *voting*) za utjecaj svakog slučaja (Kotsiantis i sur., 2006). Ne postoji opis modela. Koristi se za procesiranje slika i u slučajevima gdje je prihvatljivo duže vrijeme za odgovor (Kotu i Deshpande, 2014).

#### 4.2.4. Naivni Bayesov klasifikator

Naivni Bayesov klasifikator ima korijene u statistici i teoriji vjerojatnosti. Veza između atributa i ciljne varijable je probabilistička, a algoritam polazi od jake, ali ponekad naivne pretpostavke o međusobnoj nezavisnosti atributa. Odatle i ime algoritma. Iako pretpostavka o nekoreliranosti atributa nije uvijek točna, jednostavnost i robusnost algoritma uvelike nadilazi ograničenja zbog pretpostavke o nekoreliranosti. Prema Rish (2001) algoritam može dobro raditi s blago

koreliranim atributima. Bayesov teorem je objavljen 1763. godine i matematički opisuje uvjetnu vjerojatnost.

Recimo da je  $X$  skup atributa  $\{X_1, X_2, X_3, \dots, X_n\}$ , a  $Y$  je ishod. Vjerojatnost ishoda  $P(Y)$  se zove *a priori* vjerojatnost i može se izračunati iz skupa podataka.  $P(Y/X)$  je uvjetna vjerojatnost ili *a posteriori* vjerojatnost, a daje vjerojatnost ishoda ako nam je poznata vrijednost  $X$ . Bayesov teorem tvrdi

$$P(Y/X) = \frac{P(Y) * P(X|Y)}{P(X)}$$

gdje je  $P(X/Y)$  *a posteriori* vjerojatnost atributa  $X$  klase, tj. vjerojatnost da će atribut imati određenu vrijednost ako je poznata vrijednost ishoda  $Y$ .  $P(X)$  je *a priori* vjerojatnost atributa.

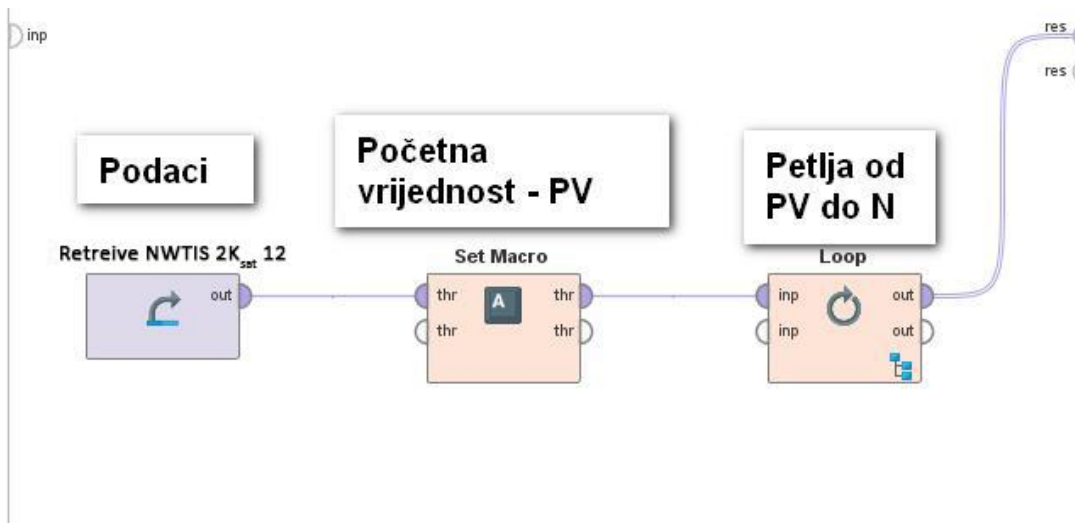
Laplaceov ispravak dodaje kontroliranu pogrešku na  $P(X/Y)$  i tako rješava problem nedostajućih vrijednosti. Problem nedostajućih vrijednosti je naročito naglašen jer se *a posteriori* vjerojatnost dobiva množenjem, pa vjerojatnost nula narušava izračun (Kotsiantis i sur., 2006).

Naivni Bayesov klasifikator je algoritam koji predviđa ciljnu varijablu temeljem Bayesovog teorema i računanja *a posteriori* vjerojatnosti i *a priori* vjerojatnosti klase. Model je tablica vjerojatnosti i *a posteriori* vjerojatnosti za svaki atribut. Nema ograničenja kakvi trebaju biti atributi, ali radi bolje s kategorijskim atributima. Vrijeme potrebno za modeliranje je relativno kratko. Algoritam ima jake statističke temelje. Skup podataka za učenje mora biti reprezentativan u odnosu na populaciju i mora sadržavati sve kombinacije ulaznih i izlaznih podataka. Atributi trebaju biti nezavisni. Primjenjuje se u otkrivanju spama i dubinskoj analizi teksta (Kotu i Deshpande, 2014).

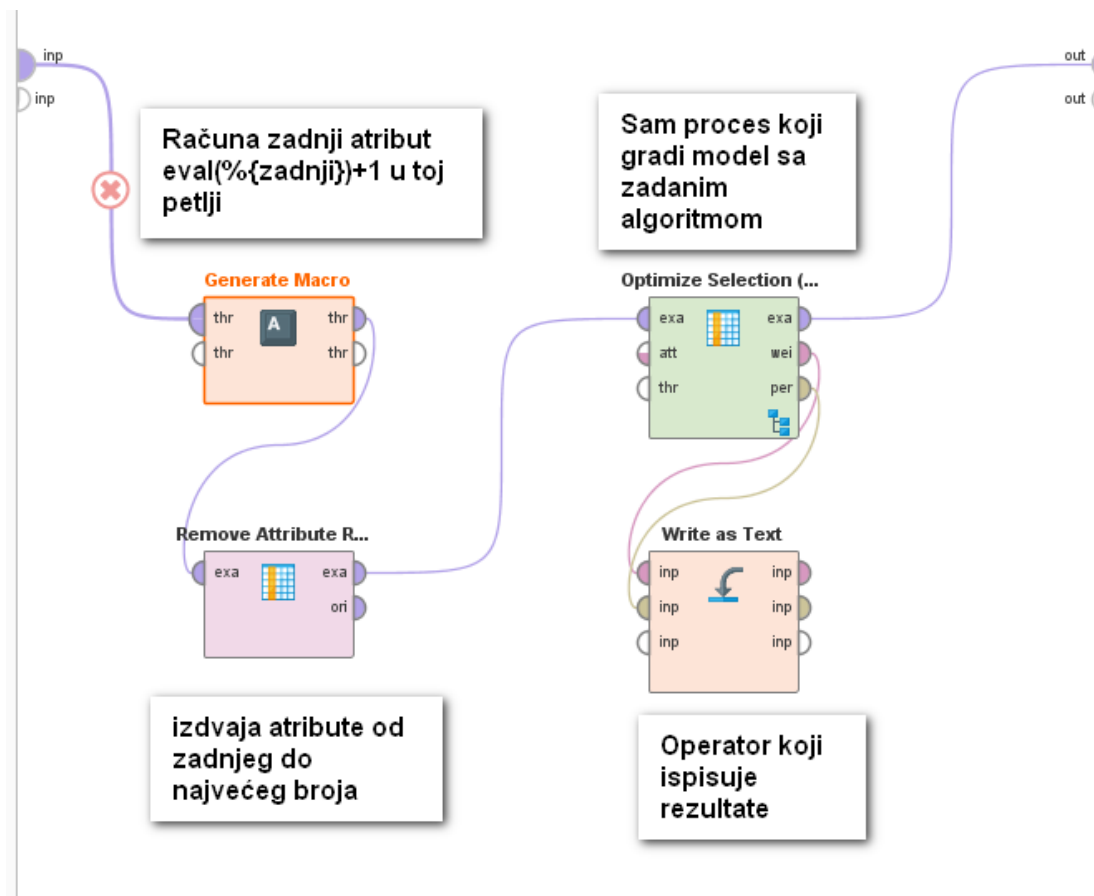
### 4.3. Model

Kao zavisna varijabla u ovdje opisanim modelima uzet je uspjeh studenta na završnom ispitu iz određenog predmeta. Numerički podaci za zavisnu varijablu podijeljeni su u kategorijske klase: u dvije i u tri kategorije. Nezavisne varijable (vremenski nizovi pristupa LMS-u) su iz vremenske domene transformirane u frekvencijsku domenu i pripremljene za modeliranje na način opisan u dijelu 4.1.

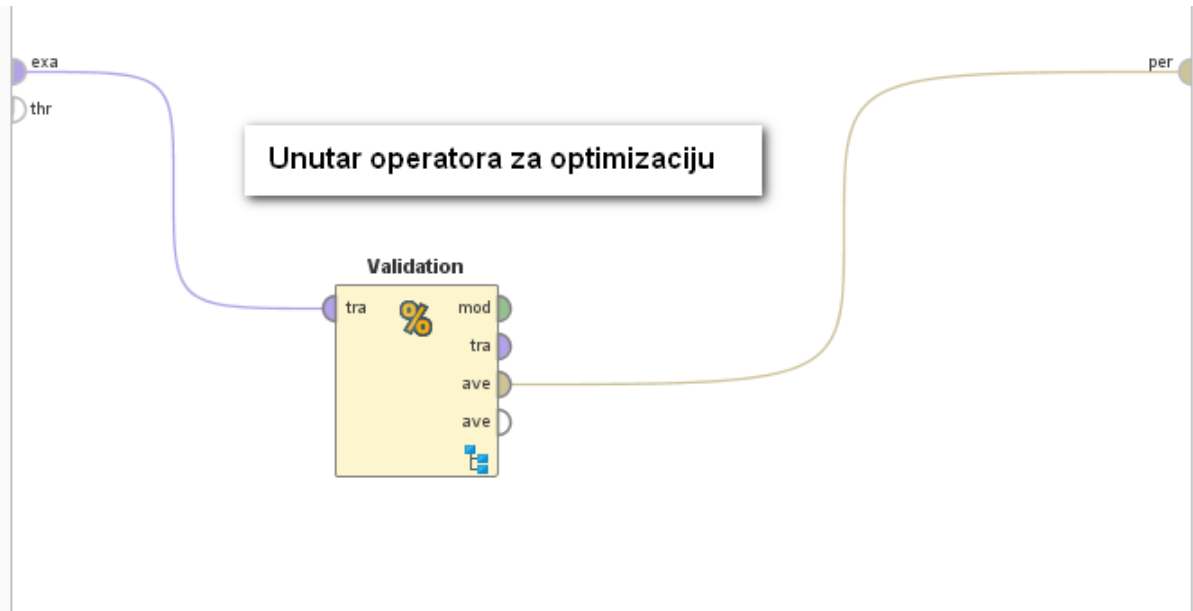
Na slici 10 prikazan je osnovni proces u RapidMineru.



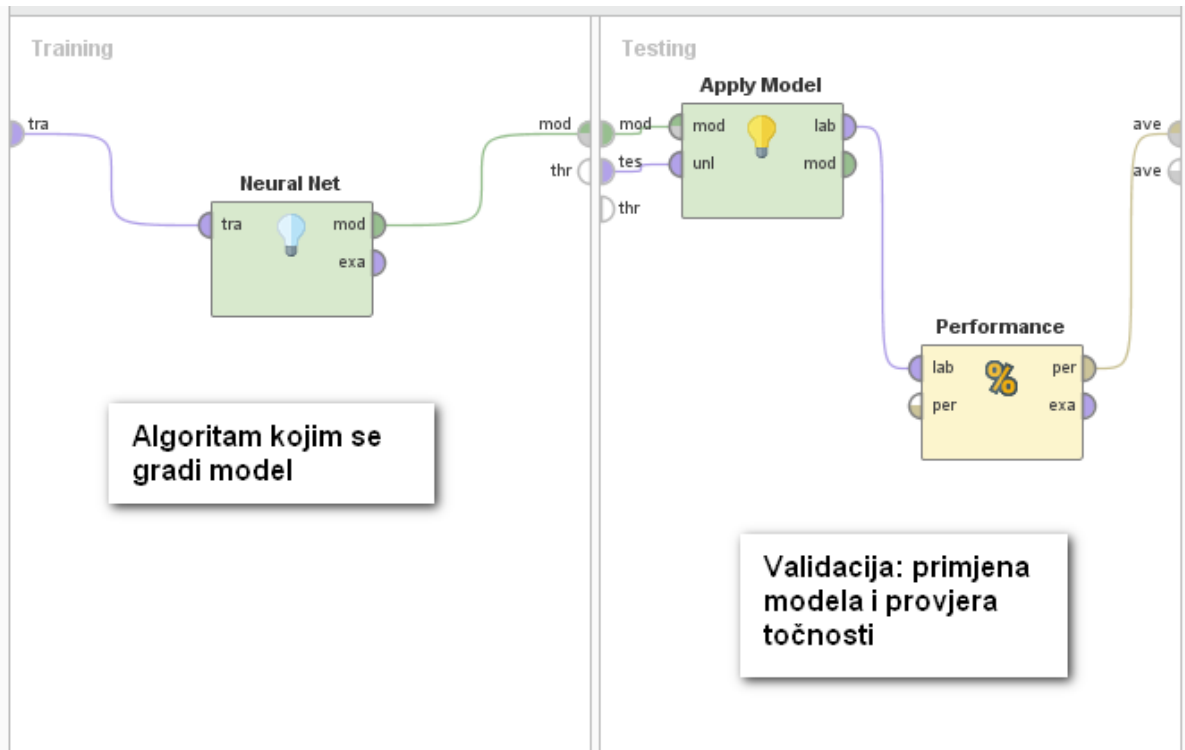
Slika 10: Osnovni proces



Slika 11: Petlja za računanje modela s postupnim povećavanjem broja atributa koje uzima u obzir



Slika 12: Postupak modeliranja i validacije unutar operatora za optimizaciju



Slika 13: Validacija

Iz 70% podataka o FT koeficijentima za promatrani predmet i odabrano vrijeme (npr. prvih 1000, 1400, 1800 itd. sati) te podataka o ishodu ispita (2 i 3 kategorije) provodi se pripisavanje modela (skup podataka za učenje), a na preostalim 30% podataka (skup podataka za testiranje) se za svaki od dobivenih modela izračunava točnost predikcije modela. Postupak se iterativno ponavlja tako da je broj odabranih koeficijenata od 4 pa sve do maksimalnog broja koeficijenata u tom skupu podataka (slika 11). Kod bržih algoritama izračunavane su točnosti predikcije za modele sa svim mogućim brojevima koeficijenata, a kod sporijih algoritama (ANN) broj koeficijenata po modelu se dijelom reducira tako da se uzimao svaki drugi koeficijent do 100, a iznad toga svaki dvadeseti. U svakoj od iteracija primijenjen je i genetski algoritam za odabir atributa.

Dobivene tekstualne datoteke bile su velike. Ovisno o broju upotrijebljenih FT koeficijenata, veličina datoteke kretala se od nekoliko tisuća pa do 70.000 linija. Ručno očitavanje podataka iz takvih datoteka bi bilo jako vremenski zahtjevno i podložno pogreškama, pa se pristupilo izradi programa za očitavanje i sortiranje rezultata. Program je napravljen pomoću integriranog razvojnog okruženja Lazarus (2017) u Free Pascalu (2017). Program omogućava učitavanje više tekstualnih datoteka dobivenih modeliranjem u RapidMineru te odabir podataka potrebnih za ovo istraživanje.

Program iz tekstualnog rezultata bira model koji ima najveću točnost, a u slučaju da više modela daje jednaku maksimalnu točnost predikcije, bira se onaj koji ima najmanji broj atributa tj. FT koeficijenata (Prilozi, kod 6). Opisani postupak izbora optimalnog modela predstavlja tehniku sažimanja.

Kako bi se omogućilo paralelno procesiranje, repozitorij podataka i procesa smješten je u oblaku (Google Drive). Tako su se procesi mogli pokretati na više različitih računala istovremeno, neovisno o njihovoj lokaciji. Rezultati su se također spremali u oblaku tako da je svako računalo vidjelo sve rezultate, pa se znalo koji su podaci obrađeni, a koji nisu. Istovremeno se stvarala i sigurnosna kopija što je smanjivalo mogućnost gubitka podataka ako bi se koje računalo srušilo. Podaci su se obrađivali na 4 računala spojena na Internet, a sva su računala pomoću TeamViewer-a mogla biti kontrolirana s jednog mjesta. Paralelno procesiranje je bilo nužno zbog velikog broja promatranih skupova podataka i iznimno velikog broja izračunatih modela. Korištena su računala sljedećih karakteristika:

Stolno računalo 1, procesor i7, 16 Gb RAM, 250 Gb SSD (engl. *solid state drive*), 2 Tb HDD (engl. *hard disk drive*), Windows 10

Stolno računalo 2, procesor i5, 8 Gb RAM, 250 Gb SSD, 1 Tb HDD, Windows 10

Stolno računalo 3, procesor i5, 8 Gb RAM, 128 Gb SSD, 1 Tb HDD, Windows 10

Prijenosno računalo 1, procesor i5, 4 Gb RAM, 1 Tb HDD, Windows 7

## 4.4. Evaluacija

### 4.4.1. Točnost

Točnost je jedna od mjera vrednovanja sposobnosti algoritma da točno razvrsta što veći broj uzoraka iz skupa podataka. Točnost je kriterij kojim se mjeri udio studenata čiji je status točno predviđen određenom tehnikom, odnosno točnost mjeri postotak točno predviđenih ishoda u odnosu na sve ishode.

Pomoću matrice zabune (engl. *confusion matrix*) moguće je analizirati koliko se rezultati dobiveni klasifikacijom razlikuju od stvarnih vrijednosti (Powers, 2011).

Matrica zabune za 2 klase prikazana je na tablici 5.

Tablica 5: Matrica zabune za 2 klase

		Predviđeni ishodi	
		Pozitivni	Negativni
Poznati ishodi	Pozitivni	TP	FN
	Negativni	FP	TN

U tablici 3 su navedene sljedeće vrijednosti:

TN=broj ispravno predviđenih negativnih ishoda

FP=broj pogrešno predviđenih pozitivnih ishoda

FN=broj pogrešno predviđenih negativnih ishoda

TP=broj ispravno predviđenih pozitivnih ishoda

Klasifikator ima bolju točnost što su vrijednosti na glavnoj dijagonali matrice zabune veće, a vrijednosti izvan dijagonale bi trebale težiti nuli.

Iz matrice zabune moguće je dobiti sljedeće mjere za vrednovanje klasifikatora (Lara *et al.*, 2014.):



- točnost (engl. *accuracy*,  $Ac$ ) je omjer broja ispravno predviđenih slučajeva i ukupnog broja slučajeva, a računa se

$$Ac = \frac{TP + TN}{TP + TN + FP + FN}$$

- preciznost (engl. *precision*,  $Pr$ ) je omjer ispravno predviđenih pozitivnih ishoda i ukupnog broja pozitivnih ishoda, a računa se

$$Pr = \frac{TP}{TP + FP}$$

- opoziv (engl. *recall*,  $Re$ ) je omjer ispravno predviđenih pozitivnih ishoda i zbroja ispravno predviđenih pozitivnih ishoda i pogrešno predviđenih negativnih ishoda, a računa se

$$Re = \frac{TP}{TP + FN}$$

- specifičnost (engl. *specificity*,  $Sp$ ) je omjer ispravno predviđenih negativnih ishoda i zbroja ispravno predviđenih negativnih ishoda i pogrešno predviđenih pozitivnih ishoda, a računa se

$$Sp = \frac{TN}{TN + FP}$$

Često se koristi i  $F$  – mjera, koja predstavlja harmonijsku sredinu preciznosti i opoziva. Točnost kao mjera vrednovanja klasifikatora se najčešće koristi no treba imati na umu da je, posebice kada ishodi nisu ravnomjerno raspoređeni po klasama, potrebno u obzir uzeti i druge mjere vrednovanja klasifikatora. Uz balansiranu zastupljenost klasa važno je i kolika je cijena pogreške. Je li lošiji FP ili FN? U situaciji predviđanja uspješnosti studenata je li lošije predvidjeti da će proći iako će pasti nego predvidjeti da pasti iako će proći. Međutim, s obzirom na to da se u ovom radu provodi vrlo velik broj izračuna, a ishodi su jednoliko raspoređeni po klasama, kao mjera vrednovanja uzeta je samo točnost.

Iz relevantne literature vidljivo je da točnost predviđanja studentskog modela ovisi i o broju klasa ciljne varijable. U pravilu, što je veći broj klasa ciljne varijable, točnost je manja (Romero, Ventura, Espejo i Hervas, 2008).

#### 4.4.2. Vrednovanje

Za procjenu pogreške klasifikatora u radu je korištena jednostruka validacija, koja se još zove i metoda izdvajanja (engl. *holdout method*), a predstavlja najjednostavniju varijantu unakrsnog vrednovanja (Kohavi, 1995). Primjeri se razdvajaju na skup za učenje (engl. *training set*) i skup za provjeru (engl. *test set*) (Larose i Larose, 2015, str. 161) u omjeru 70% slučajeva za učenje i 30% slučajeva za testiranje (slika 12 i slika 13). Odabir slučajeva za podskupove je slučajan (engl. *shuffled sampling*). Zbog ograničenih resursa i velikog broja modela u ovom istraživanju je napravljen kompromis te je provedena jednostruka validacija. U budućim istraživanjima, koja neće zahvaćati tako široko područje, bit će potrebno provesti procjenu greške generalizacije s više preklopa.

#### 4.4.3. Test statističke značajnosti

Test statističke značajnosti provodi se kako bismo dobili odgovor na pitanje jesu li uočene razlike u točnosti klasifikacije samo slučajne ili su i statistički značajne. Friedmanov test je neparametrijski ekvivalent testa ANOVA za ponovljena mjerenja. Oba testa preporučuju se za usporedbu nekoliko (više od dva) osnovnih skupova s time što Friedmanov test ne pretpostavlja da uzorci imaju normalnu distribuciju. Pretpostavke za primjenu Friedmanovog testa su (Daniel i Cross, 2013, p.713, Laerd Statistics, 2017) da se jedna grupa mjeri u tri ili više različitih prilika, da su sva opažanja nezavisna u odnosu na opažanja koja se pojavljuju u svim drugim grupama, da zavisna varijabla treba biti ordinalna ili kontinuirana (kvantitativna) i da uzorci ne moraju imati normalnu distribuciju. Friedmanov test će biti korišten za usporedbu točnosti predikcije po algoritmima (dio 5.2.1.) Usporedit će se medijani točnosti dobiveni primjenom 4 spomenuta algoritma za različite skupove podataka, pri tome vodeći računa da opažanja koja se pojavljuju u jednoj grupi budu nezavisna u odnosu na opažanja koja se pojavljuju u svim drugim grupama. Prema nultoj hipotezi primijenjenog Friedmanovog testa medijani točnosti svih primijenjenih algoritama bi trebali biti jednaki.

Zatim će u dijelu 5.2.2. biti proveden Friedmanov test po grupama predmeta koji su usporedivi s obzirom na dužine vremenskih nizova, broj klasa i vrijeme kumulacije, kako bi se provjerilo ima li unutar grupe statistički značajne razlike između medijana točnosti predikcije za različite dužine

vremenskog niza, pri tome vodeći računa da opažanja koja se pojavljuju u jednoj grupi budu nezavisna u odnosu na opažanja koja se pojavljuju u svim drugim grupama. Prema nultoj hipotezi primijenjenog Friedmanovog testa medijani točnosti za sve promatrane dužine vremenskih nizova su jednaki.

Ako je nulta hipoteza odbačena, provodi se *post-hoc* Nemenyi test (Nemenyi, 1963). Uspješnost određena dva klasifikatora se statistički značajno razlikuje ako se odgovarajući prosječni medijani razlikuju barem za kritičnu razliku.

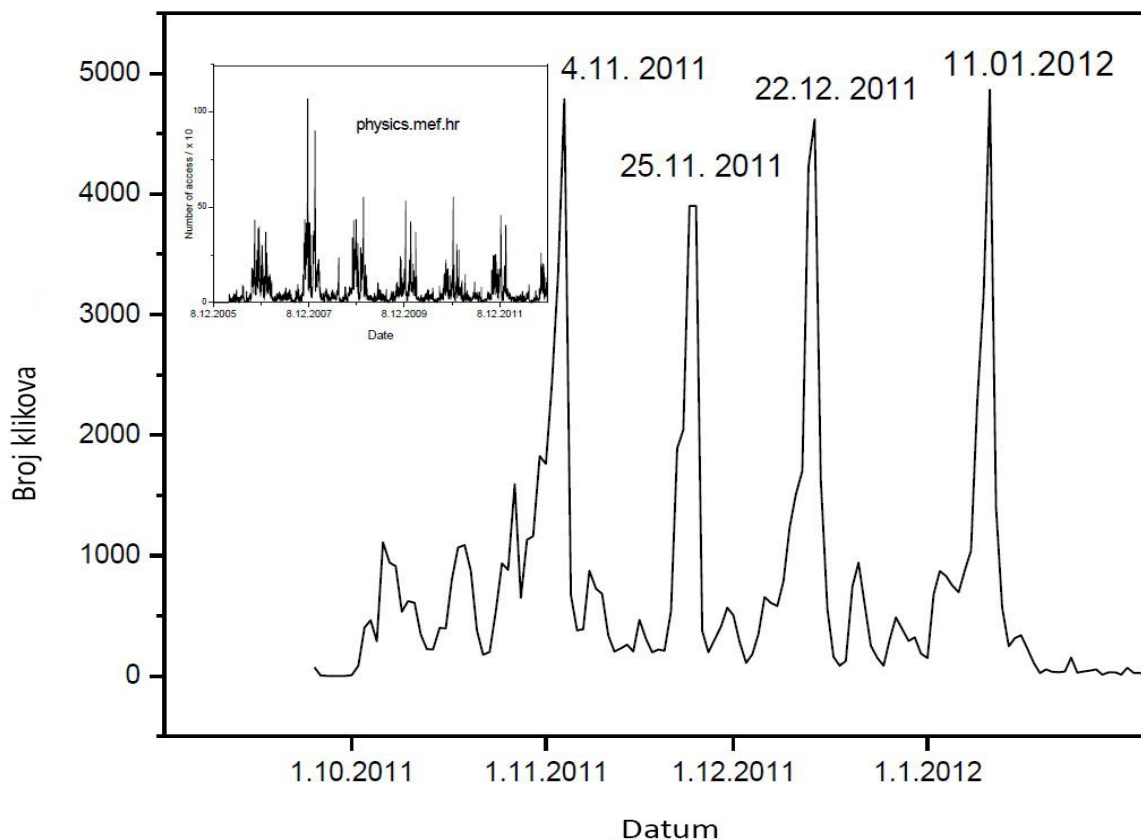
Wilcoxonov test usklađenih parova (Wilcoxon, 1945) bit će korišten u dijelu 5.2.3. i za hipotezu 2. Wilcoxonov test je neparametrijska alternativa t-testu za zavisne uzorke. Pretpostavke za primjenu Wilcoxonovog testa su (Daniel i Cross, 2013, p.682, Statistics Solutions, 2017) da dva uzorka moraju sadržavati zavisna opažanja, da su parovi opažanja slučajni i nezavisni, da je varijabla kontinuirana, a mjerna skala barem intervalna. U dijelu 5.2.3. usporedit će se točnosti predikcije po vremenu kumulacije (1 sat i 1 dan). Bit će korišten Wilcoxonov test usklađenih parova kako bi se ispitalo postoji li statistički značajna razlika između točnosti izračunate na skupovima podataka s vremenom kumulacije jedan sat, diskretiziranim u 2 kategorije i diskretiziranim u 3 kategorije te na skupovima podataka s vremenom kumulacije jedan dan, diskretiziranim u 2 kategorije i diskretiziranim u 3 kategorije. Pri tome će se voditi računa da opažanja koja se pojavljuju u jednoj grupi budu nezavisna u odnosu na opažanja koja se pojavljuju u svim drugim grupama. Testirat će se točnosti predikcije dobivene upotrebom odgovarajućih algoritama. Prema nultoj hipotezi Wilcoxonovog testa usklađenih parova nema statistički značajne razlike u točnosti predikcije temeljene na skupovima podataka s vremenom kumulacije 1 sat i skupovima podataka s vremenom kumulacije 1 dan.

Za hipotezu 2, koja glasi: *Primjenom tehnike sažimanja vremenskih nizova postiže se bolja točnost predikcije u odnosu na modele temeljene na punom skupu podataka*, Wilcoxonov test usklađenih parova bit će korišten za usporedbu točnosti na odgovarajućim parovima: na modelima s punim brojem koeficijenata i modelima sa smanjenim brojem koeficijenata, tj. prije i poslije transformacije predložene u ovom radu, pri tome vodeći računa da opažanja koja se pojavljuju u jednoj grupi budu nezavisna u odnosu na opažanja koja se pojavljuju u svim drugim grupama.

## 5. REZULTATI

### 5.1. Prethodni rezultati

Proučavajući u prethodnim istraživanjima utjecaj rezultata primjene web-formativnog ispitivanja na rezultate ispita te modeliranjem studentskog uspjeha u hibridnom okruženju za učenje temeljem atributa kao što su rezultati web-formativnih kvizova na seminarima, rezultati web-formativnih kvizova na vježbama, vrijeme utrošeno na rješavanje web-formativnih kvizova, rezultati pismenih kolokvija, rezultati testova za samoprocjenu (LMS), vrijeme utrošeno za rješavanje testova za samoprocjenu, broj logova na promatrani predmet na LMS-u, uočena je periodičnost pristupa studenata na LMS koja korespondira s datumima kolokvija i ispita (slika 14) za predmet Fizika i biofizika u akademskoj godini 2011./2012. Ista periodičnost uočena je i u periodu 2006.-2011. kada su pomoćni nastavni materijali bili smješteni na stranici physics.mef.hr. (umetak slike 14).

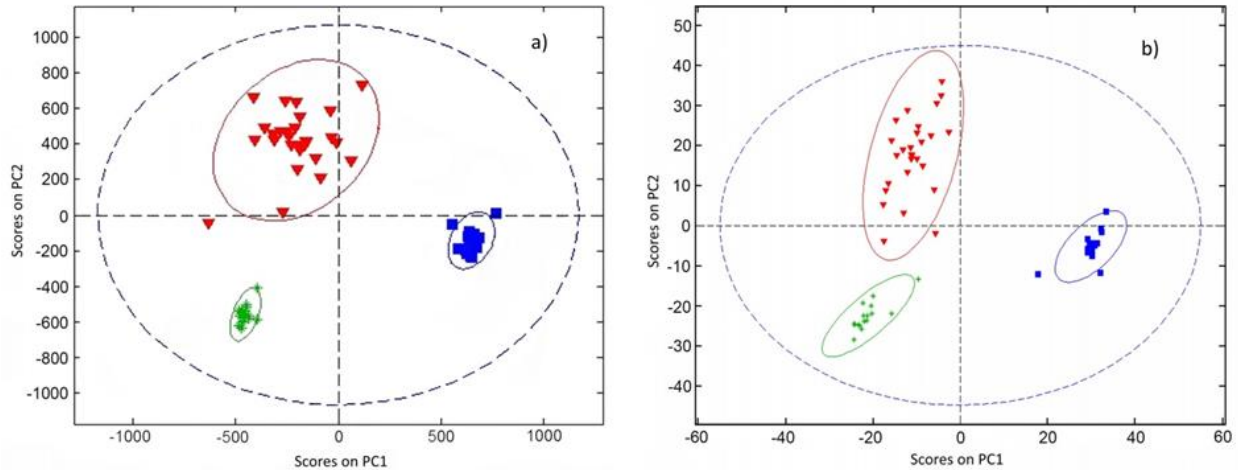


Slika 14: Broj klikova na LMS i datumi kolokvija i ispita

Izvor: Gamulin, J., O. Gamulin, D. Kermek (2013)

Zahvaljujući uočenoj periodičnosti pretpostavljeno je da u vremenskim nizovima klikova postoji informacija o obrascima studentskog ponašanja, pa tako i o uspješnosti studenata. Za preliminarno testiranje te pretpostavke iskorištena je PCA (Abdi i Williams, 2010). Vremenski nizovi klikova imaju puno sličnosti sa spektrima fizičkih ili kemijskih sistema pa je analiza provedena PCA alatom iz Matlabovog kemometrijskog skupa alata PLS\_Toolbox (7.0.3. (2013) Eigenvector Research, Inc., Wenatchee, WA, USA). Na podatke o vremenskom nizu pristupa studenata LMS-u za jedan od analiziranih predmeta (Napredne WEB tehnologije i servisi u ak. g. 2012./2013.) primijenjena je PCA. Postupak je ponovljen i na podacima o frekvencijskim spektrima dobivenim iz navedenih vremenskih nizova. Cilj je bio uočiti da li će se primjenom PCA vremenski nizovi klikova grupirati na prikazu faktorskih bodova u skupine koje odgovaraju definiranim kategorijama ishoda na ispitu (opisanim u dijelu 4.1.). PCA *score* graf (prikaz faktorskih bodova) vremenskog niza za predmet Napredne WEB tehnologije i servisi, prikazan je na slici 15(a), a može se uočiti razdvajanje vremenskog niza prema kategorijama uspješnosti studenata. Transformacijom iz vremenske u frekvencijsku domenu upotrebom DFT, mogu se dobiti dodatne informacije, poput periodičnosti studentskih logova (Bagnall i Janacek, 2005). Pretpostavka je da će informacija pohranjena u vremenskom nizu biti očuvana te da će biti otkrivena, do tada skrivena, informacija o uzorku (engl. *pattern*) studentskih pristupa. Kako bi se provjerila pretpostavka da će informacija biti očuvana i u frekvencijskoj domeni, provedena je PCA frekvencijskog (spektra) periodograma izračunatog pomoću DFT. Prikaz faktorskih bodova PC1 i PC2 komponente prikazan je na slici 15(b).

PCA prikaz faktorskih bodova na vremenskom nizu logova (slika 15(a)) i na frekvencijskom spektru (slika 15(b)) su slični što pokazuje da je informacija o studentskom uspjehu očuvana. Na oba grafa postotak objašnjene varijance je sličan, oko 70% (slika 15 (a) PC1 - 45,7%, PC2 - 24,2%; slika 15 (b) PC1 - 48,3%, PC2 - 22,8%) Sistem je sam odredio broj glavnih komponenti koje je zadržao. Automatski računa 20 glavnih komponenti, ali sugerira zadržavanje manjeg broja na osnovi postotka objašnjene varijance, u slučaju ovih podataka njih 2 da se vidi da li se prepoznaje obrazac. Ovaj ohrabrujući rezultat bio je smjerokaz prema mogućnosti modeliranja studentskog uspjeha temeljem sažetih podataka u frekvencijskoj domeni.



*Slika 15: PC prikaz faktorskih bodova PC1 i PC2 komponente vremenskog niza logova za studente upisane u predmet Napredne WEB tehnologije i servisi. Uspješnost studenata u 3 kategorije: vrlo dobri (zelene zvjezdice), prosječni (plavi kvadratići) i loši (crveni trokutići).*

*Elipse na slici definiraju granice pouzdanosti od 95%*

- a) PC prikaz faktorskih bodova za vremenske nizove logova*
- b) za frekvencijske spektre*

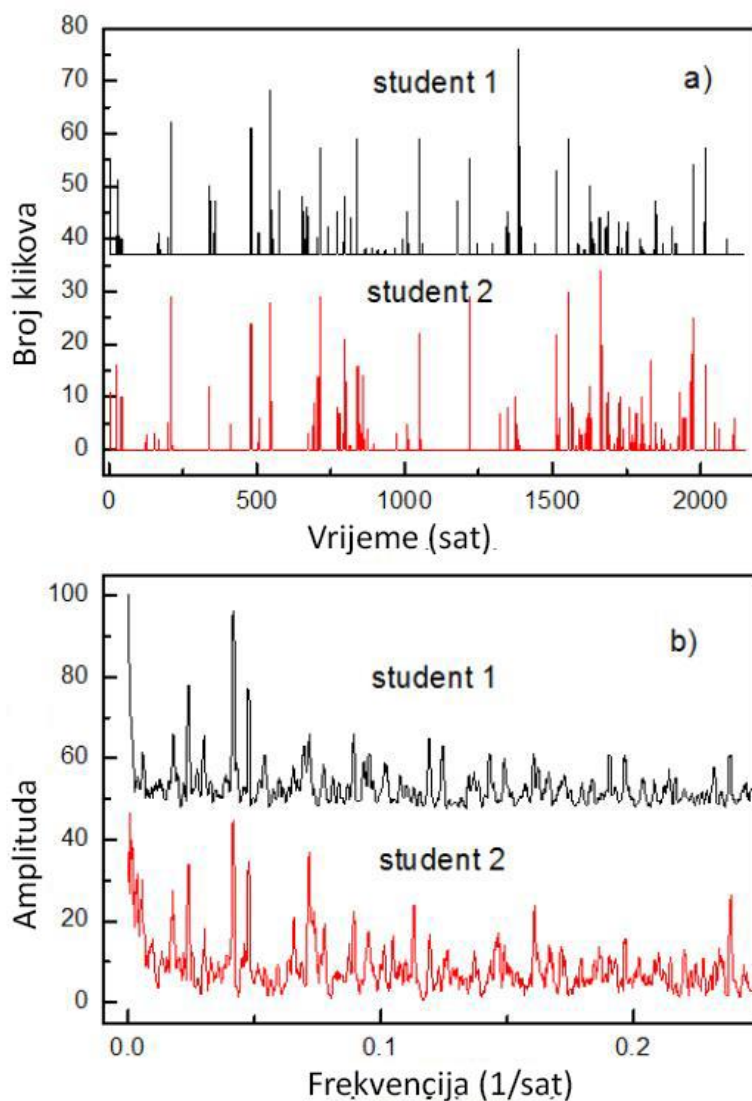
*Izvor: Gamulin, Gamulin i Kermek (2016)*

Kada je pomoću PCA pokazano da postoji razlika između kategorija studenata u promatranim skupovima podataka, pristupilo se modeliranju na način prikazan u dijelu 4.

## **5.2. Modeliranje korištenjem podataka u frekvencijskoj domeni**

Razlika između podataka o logovima u vremenskoj i u frekvencijskoj domeni, prikazana je na slici 16. Na slici su prikazani vremenski nizovi i periodogrami nastali primjenom DFT za dva slučajno odabrana studenta. Na slici 16(a) prikazani su vremenski nizovi koji izgledaju sasvim slučajno. Nasuprot tome, na slici 16(b) prikazani su periodogrami dobiveni primjenom DFT na gore prikazane vremenske nizove. Na slici 16(b) mogu se uočiti zajednički, istaknuti koeficijenti. Na slici 16(b) prikazana je samo prva polovica periodograma kako bi se naglasili zajednički dijelovi (koeficijenti). Pretpostavka je da su informacije potrebne za modeliranje studentskog uspjeha pohranjene upravo u tim istaknutim koeficijentima. U ovom istraživanju se potencijalno

zanimljivi koeficijenti protežu preko velikog dijela frekvencijskog spektra te je zato odabran pristup korištenja najboljih  $k$  Fourierovih koeficijenata.



*Slika 16: a) vremenski nizovi logova za dva slučajno odabrana studenta upisana na predmet NWTIS<sub>sat</sub> 2012; b) DFT periodogrami izračunati iz vremenskih nizova prikazanih na slici a). Zbog bolje preglednosti prikazana je samo donja polovina periodograma*

*Izvor: Gamulin, Gamulin i Kermek (2016)*

### 5.2.1. Usporedba točnosti predikcije po algoritmima

U radu su korištena 4 algoritma za predikciju uspjeha studenata, a uspoređivani su s obzirom na točnost klasifikacije. Dobivene točnosti po algoritmima su uspoređene Friedmanovim testom kako bi se prihvatila ili odbacila nulta hipoteza. Prema nultoj hipotezi primijenjenog Friedmanovog testa nema razlike između točnosti dobivenih korištenim algoritmima.

Uspoređena su sva 4 algoritma na skupovima podataka sa zavisnom varijablom diskretiziranom na 2 kategorije, vrijeme kumulacije je 1 sat, na vremenskom nizu pune dužine predmeta (2800 sati za FIB 2K<sub>sat</sub> 13, FIB 2K<sub>sat</sub> 12, NWTIS 2K<sub>sat</sub> 13, NWTIS 2K<sub>sat</sub> 12, WDIP 2K<sub>sat</sub> 13, WDIP 2K<sub>sat</sub> 12 i 120 sati za FMD 2K<sub>sat</sub> 13, FMD 2K<sub>sat</sub> 12), tablica 6. Promatrani modeli imaju smanjen broj koeficijenata na način opisan u dijelu 4.3. U tablici 6 je prezentirana maksimalna točnost dobivena takvim iterativnim postupkom smanjivanja broja koeficijenata. Od sada nadalje kada se u tekstu spominje smanjeni broj koeficijenata, misli se na ovdje i u dijelu 4.3. opisani postupak.

*Tablica 6: Točnost - 4 algoritma, zavisna varijabla diskretizirana na 2 kategorije, vrijeme kumulacije je 1 sat, na vremenskom nizu pune dužine predmeta, primijenjena tehnika sažimanja*

	ANN	SVM	kNN	NB
FIB 2K <sub>sat</sub> 13 - 2800 sati	83,80	86,02	<b>88,17</b>	77,42
FIB 2K <sub>sat</sub> 12 - 2800 sati	83,70	78,25	<b>84,78</b>	79,35
NWTIS 2K <sub>sat</sub> 13 - 2800 sati	93,25	93,70	93,33	<b>99,05</b>
NWTIS 2K <sub>sat</sub> 12 - 2800 sati	89,47	89,35	<b>94,77</b>	87,50
WDIP 2K <sub>sat</sub> 13 - 2800 sati	88,17	<b>96,18</b>	92,03	80,25
WDIP 2K <sub>sat</sub> 12 - 2800 sati	<b>95,83</b>	87,52	87,50	87,50
FMD 2K <sub>sat</sub> 13 - 120 sati	85,19	<b>86,42</b>	<b>86,42</b>	83,95
FMD 2K <sub>sat</sub> 12 - 120 sati	<b>93,42</b>	80,26	90,75	86,84
<b>Medijan</b>	88,82	86,97	<b>89,46</b>	85,40

U Friedmanovom testu vrijednost Hi-kvadrat statistike s 3 stupnja slobode bila je 5,27 (p=0,153). Na razini statističke značajnosti p<0,05 ne možemo odbaciti nultu hipotezu da ne postoji statistički značajna razlika u medijanima točnosti predikcije pomoću analiziranih algoritama.



Potom su uspoređena sva 4 algoritma na skupovima podataka sa zavisnom varijablom diskretiziranom na 2 kategorije, vrijeme kumulacije je 1 dan, na vremenskom nizu pune dužine predmeta (120 dana za FIB 2K<sub>dan</sub> 13, FIB 2K<sub>dan</sub> 12, NWTIS 2K<sub>dan</sub> 13, NWTIS 2K<sub>dan</sub> 12, WDIP 2K<sub>dan</sub> 13, WDIP 2K<sub>dan</sub> 12), tablica 7. Promatrani modeli imaju smanjen broj koeficijenata na način opisan u dijelu 4.3. U tablici 7 je prezentirana maksimalna točnost dobivena takvim iterativnim postupkom smanjivanja broja koeficijenata.

*Tablica 7: Točnost - 4 algoritma, zavisna varijabla diskretizirana na 2 kategorije, vrijeme kumulacije je 1 dan, na vremenskom nizu pune dužine predmeta, primijenjena tehnika sažimanja*

	ANN	SVM	kNN	NB
FIB 2K <sub>dan</sub> 13 - 120 dana	81,72	<b>84,95</b>	<b>84,95</b>	78,45
FIB 2K <sub>dan</sub> 12 - 120 dana	82,61	73,85	<b>84,79</b>	78,26
NWTIS 2K <sub>dan</sub> 13 - 120 dana	93,33	93,24	93,32	<b>93,56</b>
NWTIS 2K <sub>dan</sub> 12 - 120 dana	92,55	94,74	<b>94,77</b>	89,47
WDIP 2K <sub>dan</sub> 13 - 120 dana	88,58	92,15	<b>96,03</b>	92,25
WDIP 2K <sub>dan</sub> 12 - 120 dana	<b>91,67</b>	87,25	87,50	83,33
<b>Medijan</b>	90,13	89,70	<b>90,41</b>	86,40

U Friedmanovom testu vrijednost Hi-kvadrat statistike s 3 stupnja slobode bila je 4,627 (p=0,201). Na razini statističke značajnosti p<0,05 ne možemo odbaciti nultu hipotezu da ne postoji statistički značajna razlika u medijanima točnosti predikcije pomoću analiziranih algoritama.

Zatim su uspoređena 3 algoritma na skupovima podataka sa zavisnom varijablom diskretiziranom na 3 kategorije, vrijeme kumulacije 1 sat, na vremenskom nizu pune dužine predmeta (2800 sati za FIB 3K<sub>sat</sub> 13, FIB 3K<sub>sat</sub> 12, NWTIS 3K<sub>sat</sub> 13, NWTIS 3K<sub>sat</sub> 12, WDIP 3K<sub>sat</sub> 13, WDIP 3K<sub>sat</sub> 12). Algoritam SVM nije korišten za skupove podataka sa zavisnom varijablom diskretiziranom na 3 kategorije (tablica 8). Promatrani modeli imaju smanjen broj koeficijenata na način opisan u dijelu 4.3.

*Tablica 8: Točnost - 3 algoritma, zavisna varijabla diskretizirana na 3 kategorije, vrijeme kumulacije je 1 sat, na vremenskom nizu pune dužine predmeta, primijenjena tehnika sažimanja*

	<b>ANN</b>	<b>kNN</b>	<b>NB</b>
FIB 3K <sub>sat</sub> 13 - 2800 sati	70,97	<b>72,04</b>	61,15
FIB 3K <sub>sat</sub> 12 - 2800 sati	<b>77,83</b>	76,09	60,85
NWTIS 3K <sub>sat</sub> 13 - 2800 sati	86,67	<b>93,36</b>	93,33
NWTIS 3K <sub>sat</sub> 12 - 2800 sati	84,21	<b>89,45</b>	89,37
WDIP 3K <sub>sat</sub> 13 - 2800 sati	76,05	72,03	<b>80,95</b>
WDIP 3K <sub>sat</sub> 12 - 2800 sati	75,26	<b>87,25</b>	79,17
<b>Medijan</b>	76,94	<b>81,67</b>	80,06

U Friedmanovom testu vrijednost Hi-kvadrat statistike s 2 stupnja slobode bila je 2,333 ( $p=0,311$ ). Na razini statističke značajnosti  $p<0,05$  ne možemo odbaciti nultu hipotezu da ne postoji statistički značajna razlika u medijanima točnosti predikcije pomoću analiziranih algoritama.

Zatim su uspoređena 3 algoritma na skupovima podataka sa zavisnom varijablom diskretiziranom na 3 kategorije, vrijeme kumulacije 1 dan, na vremenskom nizu pune dužine predmeta (120 dana za FIB 3K<sub>dan</sub> 13, FIB 3K<sub>dan</sub> 12, NWTIS 3K<sub>dan</sub> 13, NWTIS 3K<sub>dan</sub> 12, WDIP 3K<sub>dan</sub> 13, WDIP 3K<sub>dan</sub> 12). Algoritam SVM nije korišten za skupove podataka sa zavisnom varijablom diskretiziranom na 3 kategorije (tablica 9). Promatrani modeli imaju smanjen broj koeficijenata na način opisan u dijelu 4.3.

Tablica 9: Točnost - 3 algoritma, zavisna varijabladiskretizirana na 3 kategorije, vrijeme kumulacije je 1 dan, na vremenskom nizu pune dužine predmeta, primijenjena tehnika sažimanja

	ANN	kNN	NB
FIB 3K <sub>dan</sub> 13 - 120 dana	68,82	<b>69,87</b>	61,29
FIB 3K <sub>dan</sub> 12 - 120 dana	68,84	<b>76,09</b>	61,96
NWTIS 3K <sub>dan</sub> 13 - 120 dana	86,67	<b>93,33</b>	86,55
NWTIS 3K <sub>dan</sub> 12 - 120 dana	<b>84,71</b>	84,25	78,95
WDIP 3K <sub>dan</sub> 13 - 120 dana	<b>68,25</b>	68,05	68,22
WDIP 3K <sub>dan</sub> 12 - 120 dana	<b>87,51</b>	79,16	83,25
<b>Medijan</b>	76,78	<b>77,63</b>	73,59

U Friedmanovom testu vrijednost Hi-kvadrat statistike s 2 stupnja slobode bila je 4,333 ( $p=0,115$ ). Na razini statističke značajnosti  $p<0,05$  ne možemo odbaciti nultu hipotezu da ne postoji statistički značajna razlika u medijanima točnosti predikcije pomoću analiziranih algoritama.

Postupak je ponovljen i za skupove podataka o vremenskim nizovima za 2/3 trajanja predmeta. Uspoređena su sva 4 algoritma na skupovima podataka sa zavisnom varijablom diskretiziranom na 2 kategorije, vrijeme kumulacije 1 sat, na vremenskom nizu za 2/3 trajanja predmeta (1800 sati za FIB 2K<sub>sat</sub> 13, FIB 2K<sub>sat</sub> 12, NWTIS 2K<sub>sat</sub> 13, NWTIS 2K<sub>sat</sub> 12, WDIP 2K<sub>sat</sub> 13, WDIP 2K<sub>sat</sub> 12 i 80 sati za FMD 2K<sub>sat</sub> 13, FMD 2K<sub>sat</sub> 12), tablica 10. Promatrani modeli imaju smanjen broj koeficijenata na način opisan u dijelu 4.3.

Tablica 10: Točnost - 4 algoritma, zavisna varijabla diskretizirana na 2 kategorije, vrijeme kumulacije je 1 sat, na vremenskom nizu za 2/3 trajanja predmeta, primijenjena tehnika sažimanja

	ANN	SVM	kNN	NB
FIB 2K <sub>sat</sub> 13 - 1800 sati	83,87	<b>84,95</b>	83,87	76,34
FIB 2K <sub>sat</sub> 12 - 1800 sati	84,78	78,26	<b>84,79</b>	78,26
NWTIS 2K <sub>sat</sub> 13 - 1800 sati	93,45	93,45	93,32	<b>99,06</b>
NWTIS 2K <sub>sat</sub> 12 - 1800 sati	89,47	91,67	93,33	<b>94,68</b>
WDIP 2K <sub>sat</sub> 13 - 1800 sati	92,25	<b>96,63</b>	96,02	92,05
WDIP 2K <sub>sat</sub> 12 - 1800 sati	91,67	91,67	<b>95,83</b>	91,57
FMD 2K <sub>sat</sub> 13 - 80 sati	83,95	<b>88,89</b>	87,55	85,19
FMD 2K <sub>sat</sub> 12 - 80 sati	90,79	78,95	<b>90,81</b>	81,58
<b>Medijan</b>	90,13	90,28	<b>92,07</b>	88,38

U Friedmanovom testu vrijednost Hi-kvadrat statistike s 3 stupnja slobode bila je 3,237 ( $p=0,357$ ). Na razini statističke značajnosti  $p<0,05$  ne možemo odbaciti nultu hipotezu da ne postoji statistički značajna razlika u medijanima točnosti predikcije pomoću analiziranih algoritama.

Zatim su uspoređena sva 4 algoritma na skupovima podataka sa zavisnom varijablom diskretiziranom na 2 kategorije, vrijeme kumulacije 1 dan, na vremenskom nizu za 2/3 trajanja predmeta (80 dana za FIB 2K<sub>dan</sub> 13, FIB 2K<sub>dan</sub> 12, NWTIS 2K<sub>dan</sub> 13, NWTIS 2K<sub>dan</sub> 12, WDIP 2K<sub>dan</sub> 13, WDIP 2K<sub>dan</sub>), tablica 11. Promatrani modeli imaju smanjen broj koeficijenata na način opisan u dijelu 4.3.

Tablica 11: Točnost - 4 algoritma, zavisna varijabla diskretizirana na 2 kategorije, vrijeme kumulacije je 1 dan, na vremenskom nizu za 2/3 trajanja predmeta, primijenjena tehnika sažimanja

	ANN	SVM	kNN	NB
FIB 2K <sub>dan</sub> 13 - 80 dana	81,72	81,65	<b>83,87</b>	79,57
FIB 2K <sub>dan</sub> 12 - 80 dana	83,70	73,91	<b>84,78</b>	79,35
NWTIS 2K <sub>dan</sub> 13 - 80 dana	93,33	93,25	93,36	<b>93,43</b>
NWTIS 2K <sub>dan</sub> 12 - 80 dana	92,23	<b>94,74</b>	<b>94,74</b>	94,63
WDIP 2K <sub>dan</sub> 13 - 80 dana	92,05	88,35	92,03	<b>96,15</b>
WDIP 2K <sub>dan</sub> 12 - 80 dana	<b>87,54</b>	87,50	83,31	87,45
<b>Medijan</b>	89,80	87,93	88,41	<b>90,44</b>

U Friedmanovom testu vrijednost Hi-kvadrat statistike s 3 stupnja slobode bila je 1,983 ( $p=0,576$ ). Na razini statističke značajnosti  $p<0,05$  ne možemo odbaciti nultu hipotezu da ne postoji statistički značajna razlika u medijanima točnosti predikcije pomoću analiziranih algoritama.

Potom su uspoređena 3 algoritma na skupovima podataka sa zavisnom varijablom diskretiziranom na 3 kategorije, vrijeme kumulacije 1 sat, na vremenskom nizu 2/3 trajanja predmeta (1800 sati za FIB 3K<sub>sat</sub> 13, FIB 3K<sub>sat</sub> 12, NWTIS 3K<sub>sat</sub> 13, NWTIS 3K<sub>sat</sub> 12, WDIP 3K<sub>sat</sub> 13, WDIP 3K<sub>sat</sub> 12). Algoritam SVM nije korišten za skupove podataka sa zavisnom varijablom diskretiziranom na 3 klase (tablica 12). Promatrani modeli imaju smanjen broj koeficijenata na način opisan u dijelu 4.3.

Tablica 12: Točnost - 3 algoritma, zavisna varijabla diskretizirana na 3 kategorije, vrijeme kumulacije je 1 sat, na vremenskom nizu za 2/3 trajanja predmeta, primijenjena tehnika sažimanja

	ANN	kNN	NB
FIB 3K <sub>sat</sub> 13 - 1800 sati	65,59	<b>70,97</b>	60,22
FIB 3K <sub>sat</sub> 12 - 1800 sati	67,39	<b>71,74</b>	60,87
NWTIS 3K <sub>sat</sub> 13 - 1800 sati	<b>93,33</b>	93,30	93,23
NWTIS 3K <sub>sat</sub> 12 - 1800 sati	<b>78,95</b>	<b>78,95</b>	78,90
WDIP 3K <sub>sat</sub> 13 - 1800 sati	<b>76,21</b>	68,01	76,05
WDIP 3K <sub>sat</sub> 12 - 1800 sati	83,45	<b>87,50</b>	83,45
<b>Medijan</b>	<b>77,58</b>	75,35	77,48

U Friedmanovom testu vrijednost Hi-kvadrat statistike s 2 stupnja slobode bila je 5,545 ( $p=0,062$ ). Na razini statističke značajnosti  $p<0,05$  ne možemo odbaciti nultu hipotezu da ne postoji statistički značajna razlika u medijanima točnosti predikcije pomoću analiziranih algoritama.

Zatim su uspoređena 3 algoritma na skupovima podataka sa zavisnom varijablom diskretiziranom na 3 kategorije, vrijeme kumulacije 1 dan, na vremenskom nizu 2/3 trajanja predmeta (80 dana za FIB 3K<sub>dan</sub> 13, FIB 3K<sub>dan</sub> 12, NWTIS 3K<sub>dan</sub> 13, NWTIS 3K<sub>dan</sub> 12, WDIP 3K<sub>dan</sub> 13, WDIP 3K<sub>dan</sub> 12). Algoritam SVM nije korišten za skupove podataka sa zavisnom varijablom diskretiziranom na 3 kategorije (tablica 13). Promatrani modeli imaju smanjen broj koeficijenata na način opisan u dijelu 4.3.

Tablica 13: Točnost - 3 algoritma, zavisna varijabla diskretizirana na 3 kategorije, vrijeme kumulacije je 1 dan, na vremenskom nizu za 2/3 trajanja predmeta, primijenjena tehnika sažimanja

	ANN	kNN	NB
FIB 3K <sub>dan</sub> 13 - 80 dana	63,55	<b>66,67</b>	58,03
FIB 3K <sub>dan</sub> 12 - 80 dana	69,57	<b>70,65</b>	61,96
NWTIS 3K <sub>dan</sub> 13 - 80 dana	86,67	<b>93,35</b>	86,45
NWTIS 3K <sub>dan</sub> 12 - 80 dana	78,95	78,95	<b>84,34</b>
WDIP 3K <sub>dan</sub> 13 - 80 dana	<b>72,18</b>	64,02	68,15
WDIP 3K <sub>dan</sub> 12 - 80 dana	83,33	79,15	<b>87,50</b>
<b>Medijan</b>	75,57	74,80	<b>76,25</b>

U Friedmanovom testu vrijednost Hi-kvadrat statistike s 2 stupnja slobode bila je 0,261 ( $p=0,878$ ). Na razini statističke značajnosti  $p<0,05$  ne možemo odbaciti nultu hipotezu da ne postoji statistički značajna razlika u medijanima točnosti predikcije pomoću analiziranih algoritama.

Temeljem ovih testova možemo zaključiti da nema statistički značajne razlike u točnosti predikcije dobivenih različitim algoritmima. Algoritam kNN u pravilu daje najbolji rezultat, a algoritam NB u pravilu daje najlošiji rezultat. S obzirom na to da algoritam NB polazi od pretpostavke da nema korelacije među atributima, vjerojatno je to razlog nešto niže točnosti.

Zaključak da primjena različitih algoritama za predikciju ne daje statistički značajno različite točnosti predikcije, ima određeni značaj jer je korištena jednostruka unakrsna validacija. Zbog izuzetno velikog broja modela, bilo je vremenski prezahtjevno provoditi kod svih izračuna vrednovanje s 10 preklapanja. Tehnike koje služe za sprečavanje prenaučivosti modela su ili unakrsna validacija (Larose i Larose, 2015, str. 163) ili komparacija modela, koja je korištena ovdje. Kako je ovdje korištena i komparacija modela za izbjegavanje prenaučivosti, napravljen je kompromis te je provedeno jednostruko vrednovanje.

### 5.2.2. Usporedba točnosti predikcije po dužini promatranog vremenskog niza

Točnost predikcije uspoređivana je za svaki skup podataka i za svaki algoritam posebno i prema promatranim dužinama vremenskih nizova. Za predmete koji traju jedan semestar, diskretizirani su na dvije kategorije, a vrijeme kumulacije je sat, uspoređivane su točnosti temeljene na skupu podataka o prvih 1000 sati, 1400 sati, 1800 sati, 2000 sati, 2150 sati, 2500 sati, 2800 sati i maksimalan broj sati (tablica 14).

*Tablica 14: Točnosti predikcije algoritma ANN za predmete koji traju jedan semestar, dvije kategorije, vrijeme kumulacije 1 sat – za različite dužine vremenskih nizova, primijenjena tehnika sažimanja*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	79,57	82,80	83,87	83,90	81,72	76,34	83,80	80,65
FIB 2K <sub>sat</sub> 12	82,61	81,52	84,78	82,61	81,50	81,50	83,70	83,70
NWTIS 2K <sub>sat</sub> 13	93,35	93,30	93,45	93,33	93,32	93,25	93,25	93,33
NWTIS 2K <sub>sat</sub> 12	84,21	89,47	89,47	94,74	89,47	89,47	89,47	84,21
WDIP 2K <sub>sat</sub> 13	92,15	88,54	92,25	96,05	92,55	88,17	88,17	92,15
WDIP 2K <sub>sat</sub> 12	87,50	87,50	91,67	91,67	91,67	91,67	95,83	91,67

Postupak je ponovljen za predmete koji traju jedan semestar, zavisna varijabla je diskretizirana na dvije kategorije, a vrijeme kumulacije je dan, uspoređivane su točnosti temeljene na skupu podataka o prvih 60 dana, 80 dana, 100 dana, 120 dana, 150 dana i maksimalan broj dana (tablica 15).



*Tablica 15: Točnosti predikcije algoritma ANN za predmete koji traju jedan semestar, dvije kategorije, vrijeme kumulacije 1 dan – za različite dužine vremenskih nizova, primijenjena tehnika sažimanja*

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 2K <sub>dan</sub> 13	82,80	81,72	82,08	81,72	82,91	83,87
FIB 2K <sub>dan</sub> 12	80,44	83,70	84,48	82,61	82,61	84,78
NWTIS 2K <sub>dan</sub> 13	93,33	93,33	93,33	93,33	93,33	93,33
NWTIS 2K <sub>dan</sub> 12	92,14	92,23	97,35	92,55	94,47	98,13
WDIP 2K <sub>dan</sub> 13	88,78	92,05	92,10	88,58	88,15	88,25
WDIP 2K <sub>dan</sub> 12	83,33	87,54	87,15	91,67	91,50	n/a

Isto je napravljeno i za predmete koji traju jedan semestar, zavisna varijabla je diskretizirana na tri kategorije, a vrijeme kumulacije je sat, uspoređivane su točnosti temeljene na skupu podataka o prvih 1000 sati, 1400 sati, 1800 sati, 2000 sati, 2150 sati, 2500 sati, 2800 sati i maksimalan broj sati (tablica 16).

*Tablica 16: Točnosti predikcije algoritma ANN za predmete koji traju jedan semestar, tri kategorije, vrijeme kumulacije 1 sat – za različite dužine vremenskih nizova, primijenjena tehnika sažimanja*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 3K <sub>sat</sub> 13	63,44	62,37	65,59	65,59	65,59	65,59	70,97	66,67
FIB 3K <sub>sat</sub> 12	65,22	65,22	67,39	68,48	67,39	71,74	77,83	79,35
NWTIS 3K <sub>sat</sub> 13	93,33	93,33	93,33	86,67	86,67	86,67	86,67	86,67
NWTIS 3K <sub>sat</sub> 12	84,41	78,95	78,95	78,95	84,21	78,95	84,21	84,21
WDIP 3K <sub>sat</sub> 13	76,11	72,41	76,21	80,23	68,41	72,36	76,05	68,15
WDIP 3K <sub>sat</sub> 12	79,17	83,33	83,45	79,2	84,22	75,22	75,26	84,21

Postupak je ponovljen za predmete koji traju jedan semestar, zavisna varijabla je diskretizirana na tri kategorije, a vrijeme kumulacije je dan, uspoređivane su točnosti temeljene na skupu

podataka o prvih 60 dana, 80 dana, 100 dana, 120 dana, 150 dana i maksimalan broj dana (tablica 17).

*Tablica 17: Točnosti predikcije algoritma ANN za predmete koji traju jedan semestar, tri kategorije, vrijeme kumulacije 1 dan – za različite dužine vremenskih nizova, primijenjena tehnika sažimanja*

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 3K <sub>dan</sub> 13	63,44	63,55	67,74	68,82	67,74	67,74
FIB 3K <sub>dan</sub> 12	66,30	69,57	67,39	68,84	68,48	70,65
NWTIS 3K <sub>dan</sub> 13	93,34	86,67	86,67	86,67	93,33	93,33
NWTIS 3K <sub>dan</sub> 12	78,95	78,95	78,95	84,71	84,21	84,21
WDIP 3K <sub>dan</sub> 13	72,15	72,18	72,15	68,25	68,25	n/a
WDIP 3K <sub>dan</sub> 12	79,17	83,33	79,17	87,51	76,21	n/a

Konačno, usporedba je napravljena za predmet koji traje jedan tjedan, zavisna varijabla je diskretizirana na dvije kategorije, a vrijeme kumulacije je sat, uspoređivane su točnosti temeljene na skupu podataka o prvih 60 sati, 80 sati, 100 sati, 120 sati, 150 sati i maksimalan broj sati (tablica 18).

*Tablica 18: Točnosti predikcije algoritma ANN za predmet koji traje jedan tjedan, dvije kategorije, vrijeme kumulacije 1 sat – za različite dužine vremenskih nizova, primijenjena tehnika sažimanja*

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	83,95	86,42	85,19	83,95	83,95
FMD 2K <sub>sat</sub> 12	93,11	90,79	90,79	93,42	90,79	89,47

Postupak je ponovljen i za algoritme SVM (tablice 72, 76 i 80), kNN (tablice 84, 88, 92, 96 i 100) i NB (tablice 104, 108, 112, 116 i 120). Navedene tablice se nalaze na kraju rada u Prilozima. Po grupama predmeta koji su usporedivi s obzirom na dužine vremenskih nizova, broj

klasa i vrijeme kumulacije, proveden je Friedmanov test kako bi se provjerilo ima li unutar grupe statistički značajne razlike između medijana točnosti predikcije za različite dužine vremenskog niza. Formirane su sljedeće grupe:

2K <sub>sat</sub>
2K <sub>dan</sub>
2K <sub>sat</sub> FMD
3K <sub>sat</sub>
3K <sub>dan</sub>

Od ukupno 18 rezultata, u 6 slučajeva je (tablica 19) Friedmanov test pokazao da ima razlike u točnosti ako se uspoređuju točnosti predikcije za različite promatrane dužine vremenskog niza. Puni rezultati provedenih Friedmanovih testova navedeni su u Prilozima (tablica 124). Drugim riječima, u približno 67% slučajeva ne možemo odbaciti tvrdnju da modeli postižu jednako dobru točnost predikcije bez obzira temelji li se izračun na skupu podataka za cijelo vrijeme trajanja predmeta ili na skupu podataka koji pokriva puno kraće vremensko razdoblje. To je dobar pokazatelj da će rana predikcija biti moguća, a što je predmet Hipoteze 3.

*Tablica 19: Usporedba točnosti predikcije po dužini promatranog vremenskog niza i grupama skupova podataka, p-vrijednost Friedmanovog testa (bijela polja=nema razlike, siva polja=ima razlike)*

	ANN	SVM	kNN	NB
2K <sub>sat</sub>	0,0166	0,1349	0,4247	0,0405
2K <sub>dan</sub>	0,3212	0,0187	0,0355	0,4876
2K <sub>sat</sub> FMD	0,4726	0,6290	0,0004	0,9490
3K <sub>sat</sub>	0,7782	n/a	0,0370	0,7204
3K <sub>dan</sub>	0,5249	n/a	0,0642	0,6847

### 5.2.3. Usporedba točnosti predikcije po vremenu kumulacije

Zanimljivo je bilo usporediti točnosti predikcije po vremenu kumulacije. Korišten je Wilcoxonov test usklađenih parova kako bi se ispitalo postoji li statistički značajna razlika između točnosti izračunate na skupovima podataka s vremenom kumulacije jedan sat, diskretiziranim u 2 kategorije (tablica 20) i diskretiziranim u 3 kategorije (tablica 21) te na skupovima podataka s vremenom kumulacije jedan dan, diskretiziranim u 2 kategorije (tablica 22) i diskretiziranim u 3 kategorije (tablica 23). Modeli su temeljeni na podacima o punoj dužini trajanja predmeta. Testirane su točnosti dobivene upotrebom odgovarajućih algoritama u tablicama 20 i 22 te tablicama 21 i 23.

*Tablica 20: Točnost predikcije po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 sat, 2 kategorije, primijenjena tehnika sažimanja*

	2800 sati			
	ANN	SVM	kNN	NB
FIB 2K <sub>sat</sub> 13	83,80	86,02	88,17	77,42
FIB 2K <sub>sat</sub> 12	83,70	78,25	84,78	79,35
NWTIS 2K <sub>sat</sub> 13	93,25	93,70	93,33	99,05
NWTIS 2K <sub>sat</sub> 12	89,47	89,35	94,77	87,50
WDIP 2K <sub>sat</sub> 13	88,17	96,18	92,03	80,25
WDIP 2K <sub>sat</sub> 12	95,83	87,52	87,50	87,50

*Tablica 21: Točnost predikcije po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 sat, 3 kategorije, primijenjena tehnika sažimanja*

	2800 sati		
	ANN	kNN	NB
FIB 3K <sub>sat</sub> 13	70,97	72,04	61,15
FIB 3K <sub>sat</sub> 12	77,83	76,09	60,85
NWTIS 3K <sub>sat</sub> 13	86,67	93,36	93,33
NWTIS 3K <sub>sat</sub> 12	84,21	89,45	89,37
WDIP 3K <sub>sat</sub> 13	76,05	72,03	80,95
WDIP 3K <sub>sat</sub> 12	75,26	87,25	79,17

*Tablica 22: Točnost predikcije po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 dan, 2 kategorije, primijenjena tehnika sažimanja*

	120 dana			
	ANN	SVM	kNN	NB
FIB 2K <sub>dan</sub> 13	81,72	84,95	84,95	78,45
FIB 2K <sub>dan</sub> 12	82,61	73,85	84,79	78,26
NWTIS 2K <sub>dan</sub> 13	93,33	93,24	93,32	93,56
NWTIS 2K <sub>dan</sub> 12	92,55	94,74	94,77	89,47
WDIP 2K <sub>dan</sub> 13	88,58	92,15	96,03	92,25
WDIP 2K <sub>dan</sub> 12	91,67	87,25	87,50	83,33

*Tablica 23: Točnost predikcije po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 dan, 3 kategorije, primijenjena tehnika sažimanja*

	120 dana		
	ANN	kNN	NB
FIB 3K <sub>dan</sub> 13	68,82	69,87	61,29
FIB 3K <sub>dan</sub> 12	68,84	76,09	61,96
NWTIS 3K <sub>dan</sub> 13	86,67	93,33	86,55
NWTIS 3K <sub>dan</sub> 12	84,71	84,25	78,95
WDIP 3K <sub>dan</sub> 13	68,25	68,05	68,22
WDIP 3K <sub>dan</sub> 12	87,51	79,16	83,25

Rezultati Wilcoxonovog testa usklađenih parova (tablica 24) pokazuju da nema statistički značajne razlike u točnosti kada po pojedinim algoritmima usporedimo rezultate temeljene na skupovima podataka s vremenom kumulacije 1 sat i 1 dan. Jedini ograničavajući čimbenik je gustoća pristupa LMS-u. Gustoću podataka u ovom slučaju definiramo kao ukupan broj klikova na predmetu podijeljen s umnoškom broja studenata i dana trajanja predmeta. Ako je gustoća premala, nije moguće napraviti model predikcije koji bi davao zadovoljavajuće rezultate.

*Tablica 24: Rezultati Wilcoxonovog testa usklađenih parova – usporedba točnosti po algoritmima između rezultata temeljenih na skupovima podataka s vremenom kumulacije 1 sat i 1 dan, puna dužina trajanja predmeta, p-vrijednost (bijelo polje=nema razlike, sivo polje=ima razlike)*

	2 kategorije	3 kategorije
<b>ANN</b>	0,8182	0,8550
<b>SVM</b>	0,8182	n/a
<b>kNN</b>	0,9134	0,5087
<b>NB</b>	0,7835	0,8182

Zatim je isti postupak ponovljen i za modele temeljene na podacima o prve 2/3 trajanja predmeta.

Dakle, korišten je Wilcoxonov test usklađenih parova kako bi se ispitalo postoji li statistički značajna razlika između točnosti izračunate na skupovima podataka s vremenom kumulacije jedan sat, diskretiziranim u 2 kategorije (tablica 25) i diskretiziranim u 3 kategorije (tablica 26) te na skupovima podataka s vremenom kumulacije jedan dan diskretiziranim u 2 kategorije (tablica 27) i diskretiziranim u 3 kategorije (tablica 28). Testirane su točnosti dobivene upotrebom odgovarajućih algoritama u tablicama 25 i 27 te tablicama 26 i 28.

*Tablica 25: Točnost predikcije po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 sat, 2 kategorije, primijenjena tehnika sažimanja*

	1800 sati			
	ANN	SVM	kNN	NB
FIB 2K <sub>sat</sub> 13	83,87	84,95	83,87	76,34
FIB 2K <sub>sat</sub> 12	84,78	78,26	84,79	78,26
NWTIS 2K <sub>sat</sub> 13	93,45	93,45	93,32	99,06
NWTIS 2K <sub>sat</sub> 12	89,47	91,67	93,33	94,68
WDIP 2K <sub>sat</sub> 13	92,25	96,63	96,02	92,05
WDIP 2K <sub>sat</sub> 12	91,67	91,67	95,83	91,57

Tablica 26: Točnost predikcije po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 sat, 3 kategorije, primijenjena tehnika sažimanja

	1800 sati		
	ANN	kNN	NB
FIB 3K <sub>sat</sub> 13	65,59	70,97	60,22
FIB 3K <sub>sat</sub> 12	67,39	71,74	60,87
NWTIS 3K <sub>sat</sub> 13	93,33	93,30	93,23
NWTIS 3K <sub>sat</sub> 12	78,95	78,95	78,90
WDIP 3K <sub>sat</sub> 13	76,21	68,01	76,05
WDIP 3K <sub>sat</sub> 12	83,45	87,50	83,45

Tablica 27: Točnost predikcije po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 dan, 2 kategorije, primijenjena tehnika sažimanja

	80 dana			
	ANN	SVM	kNN	NB
FIB 2K <sub>dan</sub> 13	81,72	81,65	83,87	79,57
FIB 2K <sub>dan</sub> 12	83,70	73,91	84,78	79,35
NWTIS 2K <sub>dan</sub> 13	93,33	93,25	93,36	93,43
NWTIS 2K <sub>dan</sub> 12	92,23	94,74	94,74	94,63
WDIP 2K <sub>dan</sub> 13	92,05	88,35	92,03	96,15
WDIP 2K <sub>dan</sub> 12	87,54	87,50	83,31	87,45

Tablica 28: Točnost predikcije po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 dan, 3 kategorije, primijenjena tehnika sažimanja

	80 dana		
	ANN	kNN	NB
FIB 3K <sub>dan</sub> 13	63,55	66,67	58,03
FIB 3K <sub>dan</sub> 12	69,57	70,65	61,96
NWTIS 3K <sub>dan</sub> 13	86,67	93,35	86,45
NWTIS 3K <sub>dan</sub> 12	78,95	78,95	84,34
WDIP 3K <sub>dan</sub> 13	72,18	64,02	68,15
WDIP 3K <sub>dan</sub> 12	83,33	79,15	87,50

*Tablica 29: Rezultati Wilcoxonovog testa usklađenih parova – usporedba točnosti po algoritmima između rezultata temeljenih na skupovima podataka s vremenom kumulacije 1 sat i 1 dan, prve 2/3 trajanja predmeta, p-vrijednost (bijelo polje=nema razlike, sivo polje=ima razlike)*

	<b>2 kategorije</b>	<b>3 kategorije</b>
<b>ANN</b>	0,6991	0,8528
<b>SVM</b>	0,5628	n/a
<b>kNN</b>	0,3312	0,6147
<b>NB</b>	0,9372	0,9372

Rezultati Wilcoxonovog testa usklađenih parova (tablica 29) također pokazuju da nema statistički značajne razlike u točnosti kada po pojedinim algoritmima usporedimo rezultate temeljene na skupovima podataka s vremenom kumulacije 1 sat i 1 dan na modelima temeljenim na podacima o prve 2/3 trajanja predmeta.



### 5.3. Rezultati testiranja hipoteza

#### 5.3.1. Prva hipoteza

*H1: Točnost predikcije uspjeha studenata (na konačnom ispitu iz određenog predmeta) pomoću modela izgrađenog dubinskom analizom podataka iz hibridnog okruženja za učenje, a primjenom analize vremenskih nizova logova studenata u LMS-u u frekvencijskoj domeni, jednaka je ili veća od točnosti predikcije pomoću sličnih modela opisanih u literaturi.*

Za potrebe ove hipoteze, najprije će biti pobliže definirani slični modeli opisani u literaturi.

Točnost predikcije uspjeha studenata na konačnom ispitu iz određenog predmeta česta je zadaća brojnih autora. Ako se, međutim, područje interesa suzi na uspjeh studenata (ocjena) kao zavisnu varijablu te ako uspoređujemo modele temeljene na hibridnom okruženju za učenje, broj sličnih modela opisanih u literaturi znatno se smanjuje. No i tada su to samo slični modeli jer su u ovom radu razmatrani modeli temeljeni na vremenskim nizovima klikova studenata na LMS-u, a u sličnim modelima radi se o aktivnosti na forumima (Romero i sur., 2013); rezultatima zadataka, kvizova i aktivnosti na forumima (Zafra i sur., 2011); rezultatima zadataka, aktivnosti na forumima i vremenu (Mineai-Bidgoli i sur., 2003); segmentaciji aktivnosti po vremenskim razdobljima (Delgado i sur., 2006); rezultatima kvizova, testova i vremenu potrebnom za rješavanje (Gamulin i sur., 2014). U radovima Romero i sur., 2013; Zafra i sur., 2011; Delgado i sur., 2006; Gamulin i sur., 2014, opisani su slučajevi gdje studenti pristupaju platformi Moodle (2017), a u radu Minaei-Bidgoli i sur. (2003), radi se o platformi LON-CAPA (2017).

U tablici 30 nalaze se točnosti predikcije za modele opisane u ovom radu, a temeljem podataka u punoj dužini trajanja sva četiri promatrana predmeta, kada je zavisna varijabla diskretizirana na 2 kategorije, za 2012./2013. godinu. Iz tablice je vidljivo da podaci za predmet Fizika i biofizika, koji su diskretizirani na 2 kategorije, s vremenom kumulacije 1 sat, za 2012. godinu, primjenom sva 4 algoritma postižu najnižu točnost. To se može objasniti činjenicom da je gustoća pristupa na tom predmetu niska. Predmet se nalazi na samom početku prve godine što može uzrokovati nesnalaženje, a studenti nisu obavezni pristupati materijalima na LMS-u. Nasuprot tome, podaci za predmet Napredne WEB tehnologije i servisi postižu najveću točnost primjenom 2 algoritma, a primjenom trećeg algoritma dijele najveću točnost s predmetom WEB dizajn i programiranje. Predmet Napredne WEB tehnologije i servisi ima visoku gustoću pristupa, a pristup dijelu materijala je obavezan.

Tablica 30: Točnosti predikcije za modele opisane u ovom radu, a temeljem podataka u punoj dužini trajanja sva četiri predmeta, zavisna varijabla diskretizirana na 2 kategorije, po algoritmima, za 2012. godinu, primijenjena tehnika sažimanja. Crveno je označena najniža vrijednost, a zeleno najviša vrijednost za pojedini algoritam

	ANN	SVM	kNN	NB
FIB 2K <sub>sat</sub> 12 - 2800 sati	83,70	78,25	84,78	79,35
NWTIS 2K <sub>sat</sub> 12 - 2800 sati	89,47	89,35	94,77	87,50
WDIP 2K <sub>sat</sub> 12 - 2800 sati	95,83	87,52	87,50	87,50
FMD 2K <sub>sat</sub> 12 - 120 sati	93,42	80,26	90,75	86,84
MIN	83,70	78,25	84,78	79,35
MED	91,45	83,89	89,13	86,84
MAX	95,83	89,35	94,77	87,50

Raspon točnosti po algoritmima:

ANN – točnost 84-96%
SVM – točnost 78-89%
kNN – točnost 85-95%
NB – točnost 79-88%

U radu Delgado i sur. (2006) provedeno je istraživanje na uzorku od 240 studenata upisanih u predmet Metodologija i tehnologija programiranja Sveučilišta u Cordobi, Španjolska. Ciljna varijabla su ocjene na kraju predmeta diskretizirane u 2 klase (prošao/ pao). Atributa ima 10: broj službenih registracija u predmet, broj provjera znanja, ocjena, ukupan broj pristupa Moodle-u tijekom trajanja predmeta, ukupan broj pregleda resursa, postotak pregleda resursa u odnosu na ukupan broj pristupa, broj pregleda resursa po različitim kategorijama (teorija, primjeri), postotak broja pregleda resursa po kategorijama, segmentacija broja pristupa po mjesecima, segmentacija postotka pristupa po mjesecima. Podaci se odnose na period od 11 mjeseci. Varijacijom postavki i primjenom RBF ANN (engl. *radial basis functions*, *RBF*) postignuta je

točnost predviđanja 64,58-80,21%. Kako je u ovdje prikazanim modelima postignuta točnost za ANN algoritam 83,70 - 95,83%, može se zaključiti da se u ovom slučaju H1 prihvaća.

U radu autora Romero i sur. (2013) je opisano predviđanje studentskog uspjeha temeljem aktivnosti na online forumima. Uzorak čine podaci o 114 preddiplomskih studenata na predmetu Osnove računalstva, na Sveučilištu Cordoba u Španjolskoj. Zavisna varijabla je diskretizirana na 2 klase. Autori su, kao i u ovom radu, studente koji nisu prošli predmet i one s ocjenom dovoljan smjestili u istu kategoriju. Studenti s ocjenom dovoljan također su rizični i na njih također treba usmjeriti pomoć. Za predviđanje studentskog uspjeha koriste klasifikaciju i klasifikaciju pomoću klasteriranja (gupiranja). Podaci su prikupljeni na sredini i na kraju predmeta. Atributa ima 9: broj poruka koje je student napisao, broj tema koje je inicirao određeni student, broj riječi koje je student napisao, broj rečenica koje je student napisao, broj pročitanih poruka, vrijeme u minutama provedeno na forumu, rezultat vrednovanja studentove aktivnosti na forumu, stupanj centralnosti studenta (broj odgovora/ ukupan broj studenata -1), stupanj ugleda (prestiza) studenta (broj odgovora koje je primio od drugih studenata/ ukupan broj studenata -1). Izračunata je točnost i F vrijednost (harmonijska sredina preciznosti i opoziva).

Kada su uzeti podaci za cijelu dužinu predmeta u radu Romero i sur. (2013) dobivena je točnost za algoritam ANN 84-87%, (ovdje 84-96%), SVM 87-89% (ovdje 78-89%). Za oba navedena algoritma Romero i sur. (2013) ima nešto uži raspon točnosti no većim dijelom se rezultati poklapaju, s time što je ovdje za ANN bolja gornja granica točnosti.

Za algoritam kNN – dobivena je točnost 81-87% (ovdje 85-95%). Za navedeni algoritam su rezultati u ovom radu bolji. Za algoritam NB dobivena je točnost 86-90% (ovdje 79-88%), što znači da su rezultati za navedeni algoritam bolji kod Romero i sur. (2013).

Na 2 algoritma je bolji Romero i sur. (2013), a na 2 algoritma su bolju točnost imaju modeli opisani u ovom radu. Kako je prije (u dijelu 4.2.2.) pokazano da se na većem broju modela ne razlikuje točnost dobivena različitim algoritmima, može se zaključiti da je točnost modela predikcije opisanih u ovom radu i modela predikcije u radu Romero i sur. (2013) – barem jednaka te da se H1 prihvaća.

U radu autora Zafra i sur. (2011) opisano je istraživanje provedeno na Sveučilištu Cordoba u Španjolskoj, tijekom jedne akademske godine tj. od rujna do lipnja. Obrađeni su podaci o 7 predmeta i ukupno 419 studenata, a radi se o hibridnom okruženju za učenje. Studenti su

pristupali Moodle platformi. U radu su korištene informacije o kvizovima, forumima i zadacima (broj i vrijeme u sekundama, ukupno 9 atributa). Uspoređivana je točnost predikcije uspjeha na završnom ispitu (prošao/ pao – 2 klase), a korišteni su algoritmi za klasično učenje tj. učenje pomoću jedne instance (engl. *single instance learning*) te s učenjem pomoću više instanci (engl. *multiple instance learning*, MIL).

U pristupu učenjem pomoću jedne instance korišteni su algoritmi kNN (točnost 70%), NB (točnost 69%), ANN (točnost 69%) i SVM (točnost 70%).

U pristupu učenjem pomoću više instanci od usporedivih korišteni su algoritmi SVM (točnost 68%) i NB (točnost 68%). Korišteni su i drugi algoritmi, te je ukupno gledano, pokazano da je postignuta bolja točnost pristupom učenjem pomoću više instanci (MIL). U pristupu učenjem s jednom instancom postignuta je točnost 58,1-69,52%, a u pristupu učenjem s više instanci postignuta je točnost 65,71-73,57%. Sve navedene točnosti su manje od točnosti dobivenih modelima opisanim u ovom radu te se H1 prihvaća.

U radu autora Minaei-Bidgoli i sur. (2003) opisano je modeliranje na završnu ocjenu iz ispita diskretiziranu u 2 klase, 3 klase i 9 klasa. Istraživanje je provedeno na 227 studenata upisanih u uvodni predmet Fizika na Državnom sveučilištu Michigan, SAD. Radi se o hibridnom okruženju za učenje, a studenti pristupaju platformi LON-CAPA. Kao nezavisne varijable uzet je broj točnih odgovora na zadaće, rješavanje problema u prvom pokušaju, ukupan broj pokušaja, vrijeme od prvog pokušaja do rješenja, vrijeme provedeno na zadatku bez obzira je li došlo do rješenja, sudjelovanje u komunikacijskim mehanizmima, čitanje dodatnih materijala prije rješavanja ili obrnuto, puno pokušaja u kratkom vremenu bez konzultiranja dodatnih materijala ili obrnuto, odustajanje od problema, vrijeme prvog pristupa na zadatak (na početku, u sredini tjedna ili u zadnjoj minuti). Korištenjem genetskog algoritma za odabir atributa za 10-12% je poboljšán učinak klasifikatora. Značajno poboljšanje učinka klasifikatora je postignuto korištenjem tehnike kombinacija više klasifikatora (engl. *combination of multiple classifiers*, CMC).

Točnost postignuta tehnikom kombinacije više klasifikatora na 2 klase je 94,09%, na 3 klase je 72,13% i na 9 klasa je 62,25%. Kako navedena tehnika ovdje nije korištena, za usporedbu su uzeti rezultati dobiveni primjenom pojedinačnih algoritama, pa je tako primjenom kNN algoritma dobivena točnost 82,3% (ovdje 85-95%), a NB 76,4% (ovdje 79-88%).

Ovime se H1 prihvaća u ovom slučaju.

U radu Gamulin i sur. (2014) opisani su modeli predikcije studentskog uspjeha na primjeru 302 studenta upisana u predmet Fizika i biofizika na Medicinskom fakultetu. Radi se o hibridnom okruženju za učenje, a studenti su pristupali Moodle platformi. Zavisna varijabla bila je diskretizirana na 2 klase. Atributa je bilo 10: rezultati kolokvija, rezultati formativnih web-kvizova za vrijeme vježbi, rezultati formativnih web-kvizova za vrijeme seminara, rezultati formativnih web-kvizova za vrijeme seminara i vježbi, broj ukupnih pristupa predmetu na Moodle-u, sposobnost (engl. *ability*) – izračunato temeljem rezultata kolokvija, broj pristupa predmetu na Moodle-u – samo materijali za učenje, broj pristupa predmetu na Moodle-u – samo informacije o predmetu, sposobnost – izračunato temeljem rezultata testa za samoprocjenu, vrijeme potrebno za rješavanje testa za samoprocjenu. Uz primjenu genetskog algoritma za odabir atributa, algoritmom ANN dobivena je točnost 90,7% (ovdje 84-96%), algoritmom SVM dobivena je točnost 90,7% (ovdje 78-89%), a algoritmom kNN dobivena je točnost 86,05% (ovdje 85-95%).

Na dva algoritma je točnost predikcije modela opisanih u ovom radu veća ili jednaka točnosti predikcije iz rada Gamulin i sur., (2014), a na jednom algoritmu bolju točnost ima model opisan u radu Gamulin i sur., (2014). Kako je prije (u dijelu 4.2.2.) pokazano da se na većem broju modela ne razlikuje točnost dobivena različitim algoritmima, može se zaključiti da je točnost modela predikcije opisanih u ovom radu i modela predikcije u radu Gamulin i sur., (2014) – barem jednaka te da se H1 prihvaća.

Ako definiciju sličnih modela predikcije opisanih u literaturi proširimo i na klasično i na online okruženje za učenje te ako uspjeh ne definiramo samo kao ocjenu, nego i kao neodustajanje/odustajanje od studija, broj radova znatno se povećava.

U radu autora Lykourantzou i sur. (2009) opisani su modeli za predviđanje odustajanja od studija. Problem odustajanja od studija posebice je karakterističan upravo za online okruženja za učenje. Skup podataka su podaci o 193 studenta upisana na dva online predmeta uvodne razine, Računalne mreže i komunikacije te Web dizajn. Predmete drži Multimedijски tehnološki laboratorij Nacionalnog tehničkog sveučilišta u Ateni u Grčkoj (Medialab, 2008). U modelu su korišteni atributi koji ne ovise o vremenu kao što su spol, prebivalište, broj godina radnog iskustva te prethodna akademska dostignuća tj. stupanj obrazovanja. Također su korišteni atributi koji ovise o vremenu: bodovi na testovima višestrukog izbora, ocjena projekta, rok u kojem su

predali projekt, dani kašnjenja, aktivnost na LMS-u. Podaci su uzeti za 3 ciklusa (3 semestra ) održavanja predmeta. Od ukupno 193 studenta, čak 44% je odustalo.

U radu su korištene tri popularne tehnike strojnog učenja: neuronske mreže bez povratnih veza (engl. *feed forward*, FFNNs), metoda potpornih vektora (SVM) i PESFAM (engl. *probabilistic ensemble simplified fuzzy*, ARTMAP). Korištene su tri sheme odlučivanja kako bi se sumirizirali rezultati dobiveni uz pomoć sve tri tehnike. Računata je točnost, osjetljivost i preciznost i dobiveni su rezultati bolji nego u relevantnoj literaturi. Postignuta je točnost 75-88% nakon prvog dijela predavanja (od ukupno 7). Na sredini ciklusa postignuta je točnost od 97% i takva je ostala do kraja. Točnost je postignuta korištenjem sheme 1 pri korištenju sve tri tehnike jer se niti jedna tehnika sama po sebi nije pokazala boljom od ostalih. Podaci o točnosti dobiveni korištenjem pojedinačnih algoritama nisu dani u radu (Lykourantzou i sur., 2009), pa se ne mogu uspoređivati s točnostima u ovom radu. Korištenje određene sheme odlučivanja kako bi se na najbolji način integrirali rezultati dobiveni pojedinim algoritmima, svakako zaslužuje pažnju kao sljedeći korak u istraživanju.

Rad autora Kotsiantis i sur., (2004) bavi se predviđanjem studentskog uspjeha u učenju na daljinu korištenjem tehnika strojnog učenja. Ciljna varijabla je ocjena na završnom ispitu diskretizirana u 2 kategorije. Uspoređivano je 6 algoritama za strojno učenje na podacima 354 studenta upisana u predmet Uvod u informatiku s Grčkog otvorenog sveučilišta. Kao atributi modela za predviđanje studentskog uspjeha korišteni su: dob, spol, bračni status, broj djece, zanimanje, prebivalište, računalna pismenost te ocjene na 4 pismena zadatka i prisustvovanje sastancima licem-u-lice. Predmet traje 11 mjeseci.

Tehnikom ANN – BP (eng. *back propagation*, povratna propagacija pogreške) postignuta je točnost 62-82% (ovdje 84-96%), tehnikom NB postignuta je točnost 63-82 % (ovdje 79-88%), tehnikom kNN dobivena je točnost 59–77% (ovdje 85-95%), a SVM tehnikom postignuta je točnost 64-81% (ovdje 78-89%). Naivni Bayesov klasifikator pokazao se kao najprikladniji algoritam, jer je imao više nego zadovoljavajuću točnost, vrlo zadovoljavajuću osjetljivost i pokazao se kao najlakši od algoritama za primjenu. Iako se u radu Kotsiantis i sur.(2004) radi o online okruženju koje u pravilu raspolaže većim brojem podataka o studentima, točnosti postignute promatranim algoritmima se niže od točnosti dobivenih u ovom radu.

Rad autora Dekker i sur. (2009) bavi se predviđanjem odustajanja od studija na programu Elektro inženjerstva Sveučilišta tehnologije u Eindhovenu u Nizozemskoj. Radi se klasičnom okruženju za učenje, a problem koji je potaknuo istraživanje predstavlja premali broj upisanih studenata te stopa odustajanja na prvoj godini od čak 40%.

Korišteno je 14 atributa. Jedan dio atributa odnosi se na ocjene i druge podatke o prethodnom obrazovanju, a drugi skup atributa se odnosi na ocjene, upisane predmete, broj pokušaja na ispitima i sl. Primijenjene su sljedeće tehnike: stabla odlučivanja, naivni Bayesov klasifikator, logistički modeli, učenje temeljeno na pravilima (engl. *rule based learners*), slučajna šuma. Postigli su točnost 75-80% temeljem podataka o pred sveučilišnom obrazovanju i podacima prikupljenim tijekom prvog semestra. Za algoritam NB dobivena je točnost 75% (ovdje 79-88%)

U radu autora Herzog (2006) nastoji se predvidjeti odustajanje brucoša od studija temeljem podataka prikupljenih u prvom semestru njihovog studija. Korišten je algoritam ANN s 3 različite topologije. Uzorak je bio velik i sastojao se od 8018 studenata. Korišteno je 40 atributa koji su se odnosili na ocjene iz prethodnog obrazovanja, demografske podatke, podatke prikupljene tijekom prvog semestra studija te na financijske pokazatelje. Za algoritam ANN postignuta je točnost predikcije od oko 75%, podjednako za sve tri topologije (ovdje 87-96%).

U radu autora Luan (2006) modelira se na zavisnu varijablu diskretiziranu u dvije kategorije (prijelaznik na četverogodišnji studij/ nije prijelaznik) a uzorak je 15000 studenata upisanih u višu školu. Kao atributi koriste se demografski podaci, upisani predmeti, sakupljen broj bodova i financijska pomoć. Korištenjem algoritma ANN dobivena je točnost predikcije od 72% (ovdje 84-96% ).

U radu autorica Divjak i Oreški (2009) opisano je vjerojatno hibridno okruženje, ali ta činjenica nije bitna za spomenuto istraživanje. Uzorak se sastoji od 223 studenta upisana na Fakultet organizacije i informatike Sveučilišta u Zagrebu. Modelira se na varijablu uspjeh studenata diskretiziranu u dvije klase (manje uspješni i uspješniji studenti). Korišteno je 30 atributa, demografskog karaktera i podataka o prosjeku i dužini studiranja te podataka vezanih uz percepciju studenata o studiju i učenju. Kao metoda korištena je diskriminantna analiza. Postignuta je točnost oko 75%.

U radu autora Thai-Nghe i sur. (2007) opisano je istraživanje provedeno na Sveučilištu Can Tho u Vijetnamu, velikom nacionalnom sveučilištu i na Azijskom institutu za tehnologiju u Vijetnamu, koji je mali međunarodni poslijediplomski institut sa studentima iz 86 različitih zemalja. Korišteno je 15 atributa koji se odnose na demografske podatke i podatke o uspjehu na prve dvije godine preddiplomskog studija za predviđanje uspjeha na kraju treće godine studija, te na uspjeh temeljem kojeg se student upisao na poslijediplomski. U istraživanju se predviđa uspjeh na kraju studija. Na uzorku od 20492 studenata, korištenjem algoritma stabla odlučivanja postignuta je točnost 73% za 4 klase uspjeha, a za 2 klase 93-94%.

Autori Jovanović i sur. (2012) kreirali su model za predviđanje uspjeha studenata u online okruženju za učenje. Koristili su 9 atributa: broj riješenih zadataka, broj rješavanih kvizova, broj položenih kvizova, broj nepoloženih kvizova, broj poruka poslanih na forum, broj poruka pročitanih na forumu, vrijeme utrošeno na zadatke, vrijeme utrošeno na kvizove te ukupno vrijeme provedeno na forumu. Uzorak su činili podaci o 260 studenata upisanih na Fakultet organizacijskih znanosti Sveučilišta u Beogradu. Platforma kojoj studenti pristupaju je Moodle. Zavisna varijabla, tj. uspjeh studenata, je diskretizirana na 2 klase prema potrebama autora (najbolji studenti/ ostali). Tehnikom ANN postignuta je točnost 91% (ovdje 84-96%), a tehnikom NB postignuta je točnost 53% (ovdje 79-88%).

U radu autora Paliwal i Kumar (2009) opisano je klasično okruženje za učenje, a modelira se na uspjeh studenata diskretiziran u dvije klase (uspješni/ granični studenti). Uzorak čine podaci o 244 studenta upisana u jednu indijsku poslovnu školu. Autori su uzeli 4 atributa: ocjene iz prethodnog obrazovanja, rezultati testova, rezultat diskusija i intervjua te radno iskustvo. Tehnikom ANN postignuta je točnost 67-78% (ovdje 84-96%).

U radu autora Kovacic (2012) opisan je model predviđanja odustajanja od studija temeljem 9 atributa koji se odnose na demografske pokazatelje i ocjene na testovima. Uzorak se odnosi na podatke o 450 studenata koji studiraju u online okruženju i upisani su na predmet Informacijski sistemi na Otvorenoj politehnici na Novom Zelandu. Tehnikom stabla odlučivanja postignuta je točnost predikcije 59-61%, logističkom regresijom postignuta je točnost od 63%.

U radu autora Kotsiantis i sur. (2010) opisan je model previđanja studentskog uspjeha na jednom modulu studija informatike na Grčkom otvorenom sveučilištu, dakle u online okruženju za



učenje. Uzorak se sastoji od podataka o 1347 studenata. Atributi su ocjene na pismenim testovima. Zavisna varijabla je diskretizirana na dvije klase (prošao/ pao). Tehnikom ANN postignuta je točnost 81% (ovdje 84-96%), tehnikom SVM 82% (ovdje 78-89%), tehnikom kNN postignuta je točnost 79% (ovdje 85-95%), a tehnikom NB postignuta je točnost 81% (ovdje 79-88%).

U prije spomenutom radu Lara i sur. (2014) prikazan je model za predviđanje odustajanja od studija studenata u online okruženju za učenje. Uzorak su podaci o 600 studenata upisanih na Otvoreno sveučilište Madrida, u Španjolskoj. Studenti pristupaju platformi Moodle. Iz Moodle-a su prikupljeni podaci o: broju pristupa na predmet (na Moodle-u) tjedno po studentu, o broju različitih dana u tjednu u kojima je određeni student pristupio Moodle-u, je li ili nije određeni student pristupio određenom resursu u određenom tjednu, koliko puta je student pristupio određenom resursu u određenom tjednu. Koristeći Sistem za edukacijsko rudarenje podataka (engl. *system for educational data mining*, SEDM), koji autori predlažu, na polovici trajanja predmeta postignuta je točnost predikcije od 94%. U tu svrhu bili su potrebni povijesni podaci o dvije generacije studenata kako bi se model mogao testirati na trećoj generaciji od 50 studenata. Predmeti traju po 20 tjedana. Stopa odustajanja od studija je 70-80%. Model za predmet s prve godine studija ima puno nižu točnost predikcije od modela koji se odnose na predmete s ostale tri godine studija. Pomoću SEDM dobivena je točnost predikcije od 94% u desetom tjednu nastave, tehnikom ANN 82% (ovdje 84-96%), a tehnikom SVM dobivena je točnost od 50% (ovdje 78-89%).

Kao što se može vidjeti iz ovog pregleda, predikcija uspjeha studenata je predmet istraživanja mnogih autora. Ipak, te je modele teško uspoređivati jer se odnose na različita okruženja za učenje, zbog različitog odabira atributa, zbog mnoštva dostupnih algoritama, a čije se performanse mogu razlikovati i zbog proizvoljnog odabira parametara.

### 5.3.2. Druga hipoteza

*H2: Primjenom tehnike sažimanja vremenskih nizova postiže se bolja točnost predikcije u odnosu na modele temeljene na punom skupu podataka.*

Uspoređena je točnost predikcije modela izgrađenih prije opisanom tehnikom sažimanja vremenskih nizova (dio 4.3.) za sve skupove podataka u punoj dužini trajanja predmeta i po promatranim algoritmima s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata tj. na kojima nije provedena tehnika sažimanja.

Nad podacima o točnosti iz tablica 31, 33, 35 i 37 i odgovarajućim podacima o točnosti iz tablica 32, 34, 36 i 38, a prema odgovarajućim algoritmima, provedeni su Wilcoxonovi testovi usklađenih parova (tablica 39). Podaci iz navedenih tablica prikazani su i radarskim dijagramima (slike 17, 18, 19 i 20). Konačno, na svih 92 modela uspoređena je točnost predikcije dobivena sa smanjenim brojem koeficijenata i točnost predikcije s punim brojem koeficijenata.

*Tablica 31: Točnost i broj koeficijenata po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 sat, 2 kategorije, primijenjena tehnika sažimanja*

	Točnost					Broj koeficijenata				
	ANN	SVM	kNN	NB	MED	ANN	SVM	kNN	NB	Prosj. br. koef.
FIB 2K <sub>sat</sub> 13 - 2800 sati	83,80	86,02	88,17	77,42	<b>84,91</b>	7	172	146	312	<b>159</b>
FIB 2K <sub>sat</sub> 12 - 2800 sati	83,70	78,25	84,78	79,35	<b>81,53</b>	78	144	164	5	<b>98</b>
NWTIS 2K <sub>sat</sub> 13 - 2800 sati	93,25	93,70	93,33	99,05	<b>93,52</b>	4	4	10	5	<b>6</b>
NWTIS 2K <sub>sat</sub> 12 - 2800 sati	89,47	89,35	94,77	87,50	<b>89,41</b>	5	6	25	11	<b>12</b>
WDIP 2K <sub>sat</sub> 13 - 2800 sati	88,17	96,18	92,03	80,25	<b>90,10</b>	11	139	136	261	<b>137</b>
WDIP 2K <sub>sat</sub> 12 - 2800 sati	95,83	87,52	87,50	87,50	<b>87,51</b>	55	7	208	11	<b>70</b>
FMD 2K <sub>sat</sub> 13 - 120 sati	85,19	86,42	86,42	83,95	<b>85,81</b>	5	7	14	4	<b>8</b>
FMD 2K <sub>sat</sub> 12 - 120 sati	93,42	80,26	90,75	86,84	<b>88,80</b>	32	13	7	6	<b>15</b>
<b>MED</b>	<b>88,82</b>	<b>86,97</b>	<b>89,46</b>	<b>85,40</b>	<b>88,15</b>	<b>9</b>	<b>10</b>	<b>81</b>	<b>9</b>	

Tablica 32: Točnost i broj koeficijenata po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 sat, 2 kategorije, pun broj koeficijenata

	Točnost					Broj koeficijenata				
	ANN	SVM	kNN	NB	MED	ANN	SVM	kNN	NB	Prosj. br. koef.
FIB 2K <sub>sat</sub> 13 - 2800 sati	72,04	78,49	75,27	68,82	<b>73,66</b>	349	353	353	353	<b>352</b>
FIB 2K <sub>sat</sub> 12 - 2800 sati	75,00	75,00	79,35	70,65	<b>75,00</b>	349	353	403	403	<b>377</b>
NWTIS 2K <sub>sat</sub> 13 - 2800 sati	86,67	100,00	100,00	93,33	<b>96,67</b>	313	311	309	311	<b>311</b>
NWTIS 2K <sub>sat</sub> 12 - 2800 sati	84,21	68,42	89,47	75,00	<b>79,61</b>	313	323	323	368	<b>332</b>
WDIP 2K <sub>sat</sub> 13 - 2800 sati	88,00	88,00	80,00	80,00	<b>84,00</b>	403	353	353	403	<b>378</b>
WDIP 2K <sub>sat</sub> 12 - 2800 sati	79,17	62,50	79,17	75,00	<b>77,09</b>	163	353	353	368	<b>309</b>
FMD 2K <sub>sat</sub> 13 - 120 sati	81,48	82,72	83,95	59,26	<b>82,10</b>	58	58	58	58	<b>58</b>
FMD 2K <sub>sat</sub> 12 - 120 sati	86,84	77,63	90,79	59,21	<b>82,24</b>	58	57	58	58	<b>58</b>
<b>MED</b>	<b>82,85</b>	<b>78,06</b>	<b>81,98</b>	<b>72,83</b>	<b>80,85</b>	<b>313</b>	<b>338</b>	<b>338</b>	<b>361</b>	

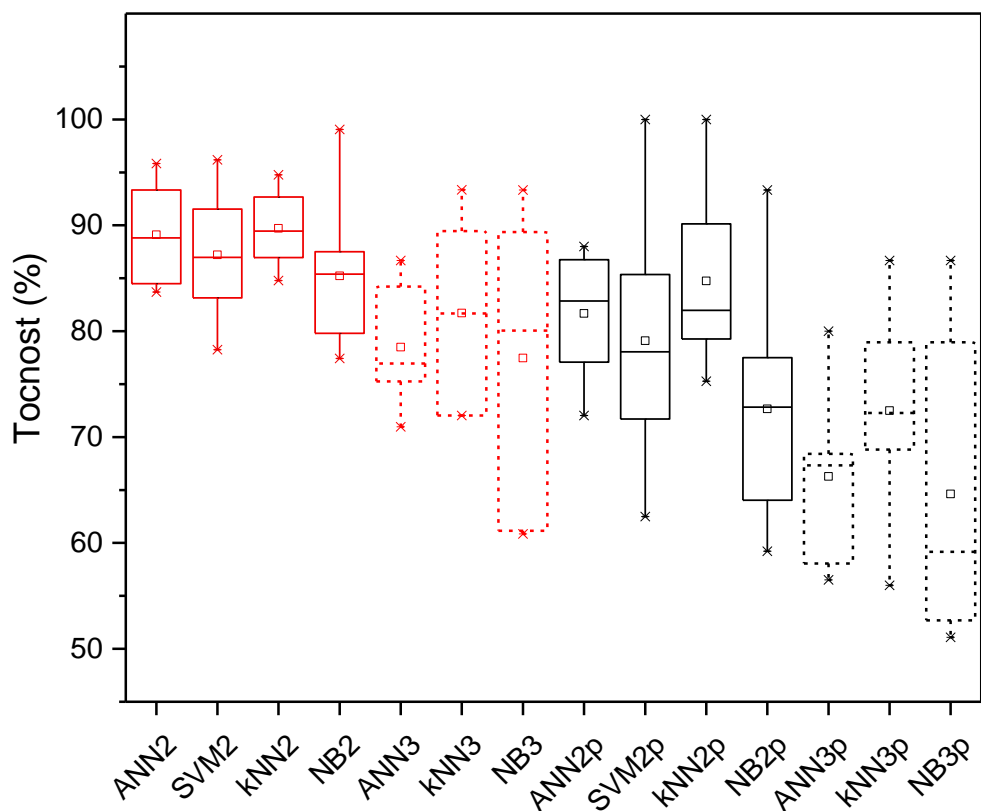
Tablica 33: Točnost i broj koeficijenata po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 sat, 3 kategorije, primijenjena tehnika sažimanja

	Točnost				Broj koeficijenata			
	ANN	kNN	NB	MED	ANN	kNN	NB	Prosj. br. koef.
FIB 3K <sub>sat</sub> 13 - 2800 sati	70,97	72,04	61,15	<b>70,97</b>	9	87	5	<b>34</b>
FIB 3K <sub>sat</sub> 12 - 2800 sati	77,83	76,09	60,85	<b>76,09</b>	11	73	6	<b>30</b>
NWTIS 3K <sub>sat</sub> 13 - 2800 sati	86,67	93,36	93,33	<b>93,33</b>	14	10	193	<b>72</b>
NWTIS 3K <sub>sat</sub> 12 - 2800 sati	84,21	89,45	89,37	<b>89,37</b>	67	273	43	<b>128</b>
WDIP 3K <sub>sat</sub> 13 - 2800 sati	76,05	72,03	80,95	<b>76,05</b>	98	153	305	<b>185</b>
WDIP 3K <sub>sat</sub> 12 - 2800 sati	75,26	87,25	79,17	<b>79,17</b>	9	208	11	<b>76</b>
<b>MED</b>	<b>76,94</b>	<b>81,67</b>	<b>80,06</b>	<b>77,63</b>	<b>13</b>	<b>120</b>	<b>27</b>	

Tablica 34: Točnost i broj koeficijenata po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 sat, 3 kategorije, pun broj koeficijenata

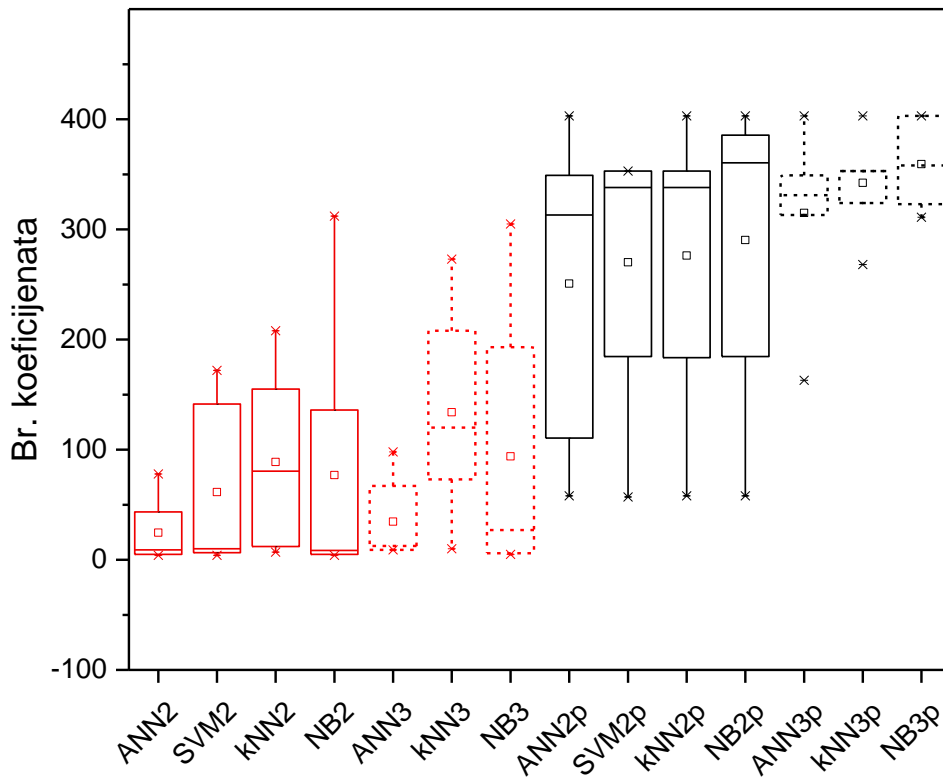
	Točnost				Broj koeficijenata			
	ANN	kNN	NB	MED	ANN	kNN	NB	Pros. br. koef.
FIB 3K <sub>sat</sub> 13 - 2800 sati	58,06	68,82	52,69	<b>58,06</b>	349	353	353	<b>352</b>
FIB 3K <sub>sat</sub> 12 - 2800 sati	56,52	69,57	51,09	<b>56,52</b>	349	403	403	<b>385</b>
NWTIS 3K <sub>sat</sub> 13 - 2800 sati	80,00	86,67	86,67	<b>86,67</b>	313	268	311	<b>297</b>
NWTIS 3K <sub>sat</sub> 12 - 2800 sati	68,42	78,95	78,95	<b>78,95</b>	313	324	323	<b>320</b>
WDIP 3K <sub>sat</sub> 13 - 2800 sati	68,00	56,00	60,00	<b>60,00</b>	403	353	403	<b>386</b>
WDIP 3K <sub>sat</sub> 12 - 2800 sati	66,67	75,00	58,33	<b>66,67</b>	163	353	363	<b>293</b>
<b>MED</b>	<b>67,34</b>	<b>72,29</b>	<b>59,17</b>	<b>63,34</b>	<b>331</b>	<b>353</b>	<b>358</b>	

Podaci iz tablica 31, 32, 33 i 34 prikazani su i box-plotovima. Na slici 17 prikazani su podaci iz navedenih tablica koji se odnose na točnost. Crne linije označavaju točnost po algoritmima za modele temeljene na punom broju koeficijenata, a crvene linije označavaju točnost po algoritmima za modele temeljene na smanjenom broju koeficijenata, vrijeme kumulacije 1 sat (pune linije - 2 kategorije; isprekidane linije - 3 kategorije).



Slika 17: Točnost za punu dužinu predmeta, vrijeme kumulacije 1 sat: crvene linije – točnost za smanjeni broj koeficijenata; crne linije - točnost za puni broj koeficijenata (pune linije - 2 kategorije; isprekidane linije - 3 kategorije)

Na slici 18 je box plot s podacima iz tablica 31, 32, 33 i 34 koji se odnose na broj korištenih koeficijenata. Crne linije označavaju pun broj korištenih koeficijenata po algoritmima, a crvene linije označavaju smanjeni broj korištenih koeficijenata po algoritmima, vrijeme kumulacije 1 sat (pune linije - 2 kategorije; isprekidane linije - 3 kategorije).



Slika 18: Broj koeficijenata za punu dužinu predmeta, vrijeme kumulacije 1 sat: crvene linije – smanjeni broj koeficijenata; crne linije - puni broj koeficijenata (pune linije - 2 kategorije; isprekidane linije - 3 kategorije)

Zatim su uspoređeni odgovarajući podaci o točnostima za vrijeme kumulacije 1 dan, a prema odgovarajućim algoritmima i skupovima podataka. Uspoređeni su podaci iz tablica 35 i 36 te 37 i 38.

Tablica 35: Točnost i broj koeficijenata po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 dan, 2 kategorije, primijenjena tehnika sažimanja

	Točnost					Broj koeficijenata				
	ANN	SVM	kNN	NB	MED	ANN	SVM	kNN	NB	Prosj. br. koef.
FIB 2K <sub>dan</sub> 13 - 120 dana	81,72	84,95	84,95	78,45	<b>83,34</b>	4	55	16	6	<b>20</b>
FIB 2K <sub>dan</sub> 12 - 120 dana	82,61	73,85	84,79	78,26	<b>80,44</b>	5	55	17	4	<b>20</b>
NWTIS 2K <sub>dan</sub> 13 - 120 dana	93,33	93,24	93,32	93,56	<b>93,33</b>	4	5	4	4	<b>4</b>
NWTIS 2K <sub>dan</sub> 12 - 120 dana	92,55	94,74	94,77	89,47	<b>93,65</b>	14	24	21	10	<b>17</b>
WDIP 2K <sub>dan</sub> 13 - 120 dana	88,58	92,15	96,03	92,25	<b>92,20</b>	4	38	20	56	<b>30</b>
WDIP 2K <sub>dan</sub> 12 - 120 dana	91,67	87,25	87,50	83,33	<b>87,38</b>	26	34	23	4	<b>22</b>
<b>MED</b>	<b>90,13</b>	<b>89,70</b>	<b>90,41</b>	<b>86,40</b>	<b>89,79</b>	<b>5</b>	<b>36</b>	<b>19</b>	<b>5</b>	

Tablica 36: Točnost i broj koeficijenata po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 dan, 2 kategorije, pun broj koeficijenata

	Točnost					Broj koeficijenata				
	ANN	SVM	kNN	NB	MED	ANN	SVM	kNN	NB	Prosj. br. koef.
FIB 2K <sub>dan</sub> 13 - 120 dana	76,34	81,72	78,49	64,52	<b>77,42</b>	58	58	60	58	<b>59</b>
FIB 2K <sub>dan</sub> 12 - 120 dana	77,17	72,83	82,61	59,78	<b>75,00</b>	58	58	57	58	<b>58</b>
NWTIS 2K <sub>dan</sub> 13 - 120 dana	93,33	93,33	100,00	86,67	<b>93,33</b>	59	58	58	58	<b>58</b>
NWTIS 2K <sub>dan</sub> 12 - 120 dana	78,95	84,21	94,74	73,68	<b>81,58</b>	58	58	58	58	<b>58</b>
WDIP 2K <sub>dan</sub> 13 - 120 dana	84,00	88,00	80,00	84,00	<b>84,00</b>	59	72	58	58	<b>62</b>
WDIP 2K <sub>dan</sub> 12 - 120 dana		79,17	79,17	83,33	<b>79,17</b>		58	58	58	<b>44</b>
<b>MED</b>	<b>78,95</b>	<b>82,97</b>	<b>81,31</b>	<b>78,51</b>	<b>80,38</b>	<b>58</b>	<b>58</b>	<b>58</b>	<b>58</b>	

Tablica 37: Točnost i broj koeficijenata po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 dan, 3 kategorije, primijenjena tehnika sažimanja

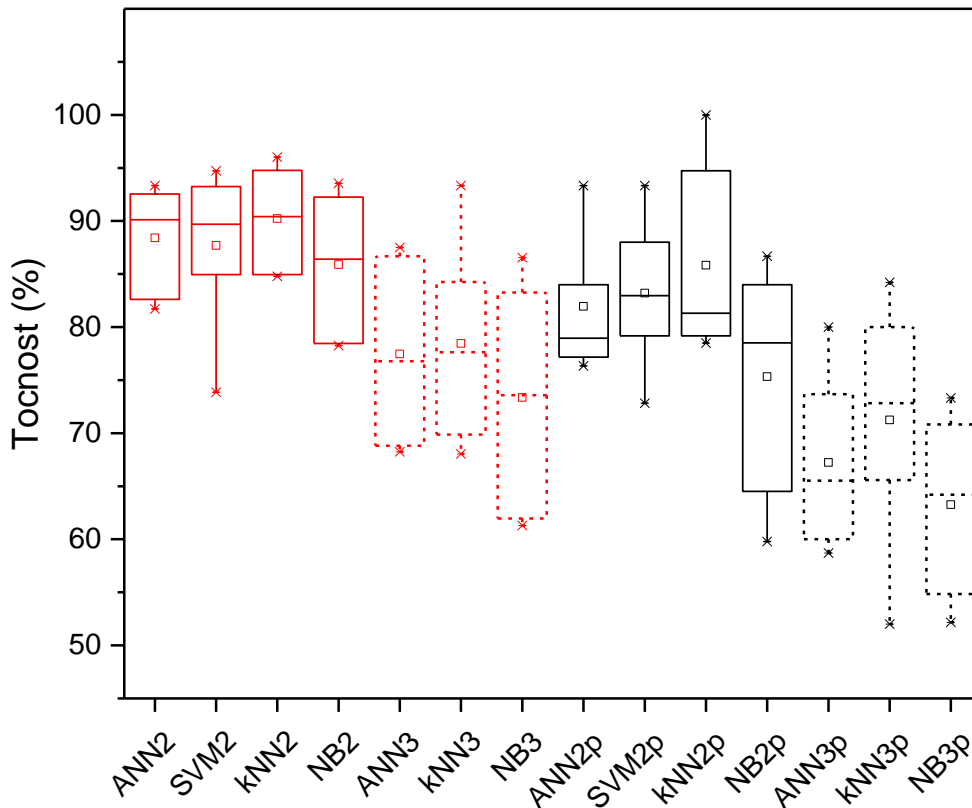
	Točnost				Broj koeficijenata			
	ANN	kNN	NB	MED	ANN	kNN	NB	Prosj. br. koef.
FIB 3K <sub>dan</sub> 13 - 120 dana	68,82	69,87	61,29	<b>68,82</b>	11	28	4	<b>14</b>
FIB 3K <sub>dan</sub> 12 - 120 dana	68,84	76,09	61,96	<b>68,84</b>	4	20	4	<b>9</b>
NWTIS 3K <sub>dan</sub> 13 - 120 dana	86,67	93,33	86,55	<b>86,67</b>	9	13	6	<b>9</b>
NWTIS 3K <sub>dan</sub> 12 - 120 dana	84,71	84,25	78,95	<b>84,25</b>	17	10	14	<b>14</b>
WDIP 3K <sub>dan</sub> 13 - 120 dana	68,25	68,05	68,22	<b>68,22</b>	4	7	13	<b>8</b>
WDIP 3K <sub>dan</sub> 12 - 120 dana	87,51	79,16	83,25	<b>83,25</b>	22	4	8	<b>11</b>
<b>MED</b>	<b>76,78</b>	<b>77,63</b>	<b>73,59</b>	<b>76,05</b>	<b>10</b>	<b>12</b>	<b>7</b>	

Tablica 38: Točnost i broj koeficijenata po algoritmima za punu dužinu trajanja predmeta, vrijeme kumulacije je 1 dan, 3 kategorije, pun broj koeficijenata

	Točnost				Broj koeficijenata			
	ANN	kNN	NB	MED	ANN	kNN	NB	Prosj. br. koef.
FIB 3K <sub>dan</sub> 13 - 120 dana	60,22	65,59	54,84	<b>60,22</b>	59	58	58	<b>58</b>
FIB 3K <sub>dan</sub> 12 - 120 dana	58,70	70,65	52,17	<b>58,70</b>	50	57	58	<b>55</b>
NWTIS 3K <sub>dan</sub> 13 - 120 dana	80,00	80,00	73,33	<b>80,00</b>	59	58	58	<b>58</b>
NWTIS 3K <sub>dan</sub> 12 - 120 dana	73,68	84,21	68,42	<b>73,68</b>	58	58	58	<b>58</b>
WDIP 3K <sub>dan</sub> 13 - 120 dana	60,00	52,00	60,00	<b>60,00</b>	59	58	58	<b>58</b>
WDIP 3K <sub>dan</sub> 12 - 120 dana	70,83	75,00	70,83	<b>70,83</b>	59	58	58	<b>58</b>
<b>MED</b>	<b>65,53</b>	<b>72,83</b>	<b>64,21</b>	<b>65,53</b>	<b>59</b>	<b>58</b>	<b>58</b>	

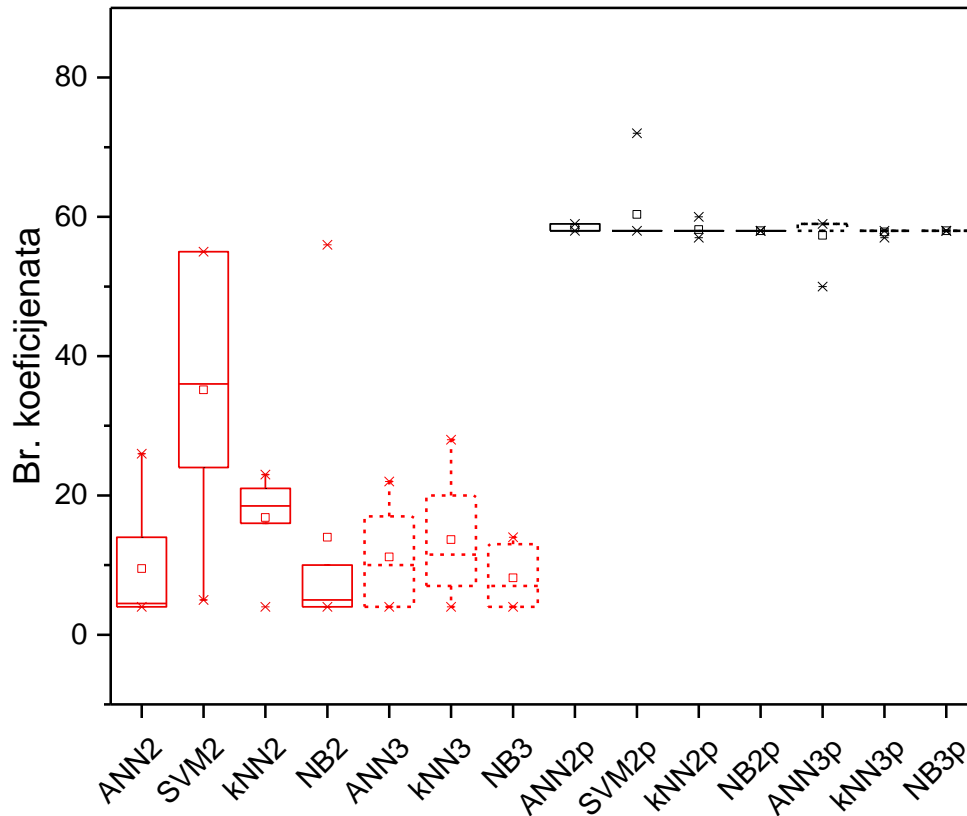
Podaci iz tablica 35, 36, 37 i 38 prikazani su i box-plotovima. Na slici 19 prikazani su podaci iz navedenih tablica koji se odnose na točnost. Crne linije označavaju točnost po algoritmima za modele temeljene na punom broju koeficijenata, a crvene linije označavaju točnost po algoritmima za modele temeljene na smanjenom broju koeficijenata, vrijeme kumulacije 1 dan (pune linije - 2 kategorije; isprekidane linije - 3 kategorije).





Slika 19: Točnost za punu dužinu predmeta, vrijeme kumulacije 1 dan: crvene linije – točnost za smanjeni broj koeficijenata; crne linije - točnost za puni broj koeficijenata (pune linije - 2 kategorije; isprekidane linije - 3 kategorije)

Na slici 20 je box-plot s podacima iz tablica 35, 36, 37 i 38 koji se odnose na broj korištenih koeficijenata. Crne linije označavaju pun broj korištenih koeficijenata po algoritmima, a crvene linije označavaju smanjeni broj korištenih koeficijenata po algoritmima, vrijeme kumulacije 1 dan (pune linije - 2 kategorije; isprekidane linije - 3 kategorije).

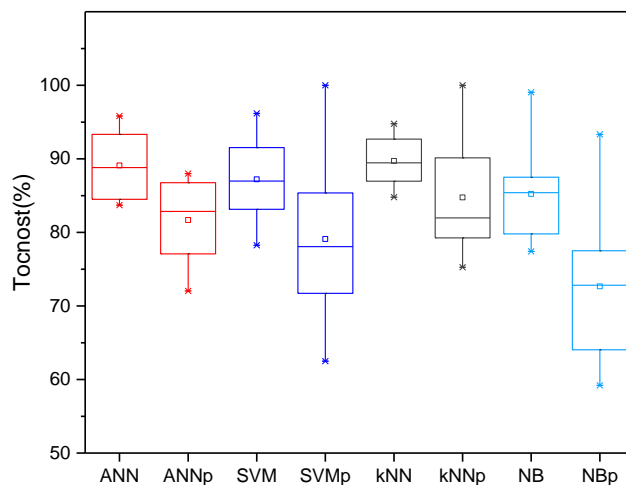


Slika 20: Broj koeficijenata za punu dužinu predmeta, vrijeme kumulacije 1 dan: crvene linije – smanjeni broj koeficijenata; crne linije - puni broj koeficijenata (pune linije - 2 kategorije; isprekidane linije - 3 kategorije)

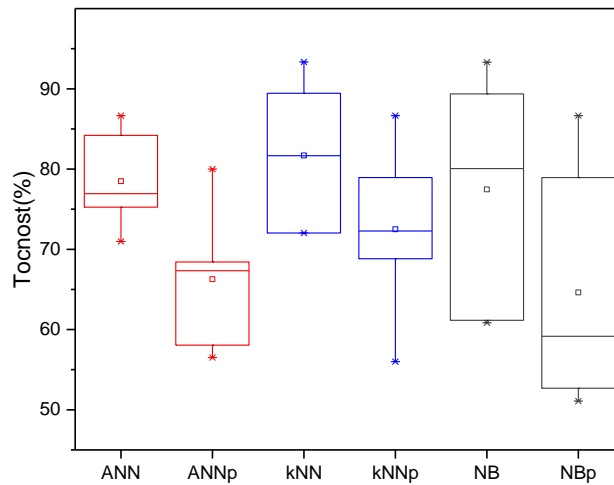
Usporedbom srednjih točnosti po algoritmima za modele izgrađene tehnikom sažimanja i modela izgrađenih s punim brojem koeficijenata, može se uočiti da modeli izgrađeni tehnikom sažimanja u pravilu imaju veću točnost predikcije. Za modele izgrađene tehnikom sažimanja srednja točnost je 83,26%, a za modele izgrađene s punim brojem koeficijenata srednja točnost je 73,68%. Radi li se i o statistički značajnoj razlici provjereno je Wilcoxonovim testom usklađenih parova.

Tablica 39: Rezultati Wilcoxonovog testa usklađenih parova – usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova za sve skupove podataka u punoj dužini trajanja predmeta i po promatranim algoritmima s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata (0=nema razlike, 1=ima razlike)

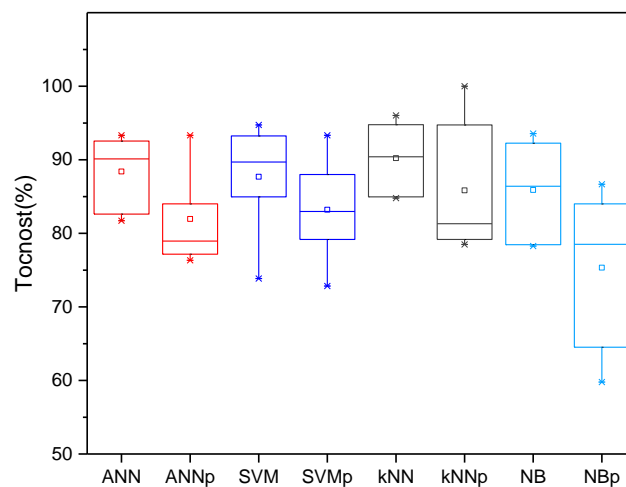
	Sati						Dani					
	2 kategorije			3 kategorije			2 kategorije			3 kategorije		
	Wilcoxon-razlika	MED grupe-točnost		Wilcoxon-razlika	MED grupe-točnost		Wilcoxon-razlika	MED grupe-točnost		Wilcoxon-razlika	MED grupe-točnost	
	Sažimanje	Pun broj koef.		Sažimanje	Pun broj koef.		Sažimanje	Pun broj koef.		Sažimanje	Pun broj koef.	
ANN	1	88,82	82,85	1	76,94	67,34	1	90,13	78,95	1	76,78	65,53
SVM	0	86,97	78,06	n/a	n/a	n/a	1	89,70	82,97	n/a	n/a	n/a
kNN	0	89,46	81,98	1	81,67	72,29	1	90,41	81,31	1	77,63	72,83
NB	1	85,40	72,83	1	80,06	59,17	1	86,40	78,51	1	73,59	64,21



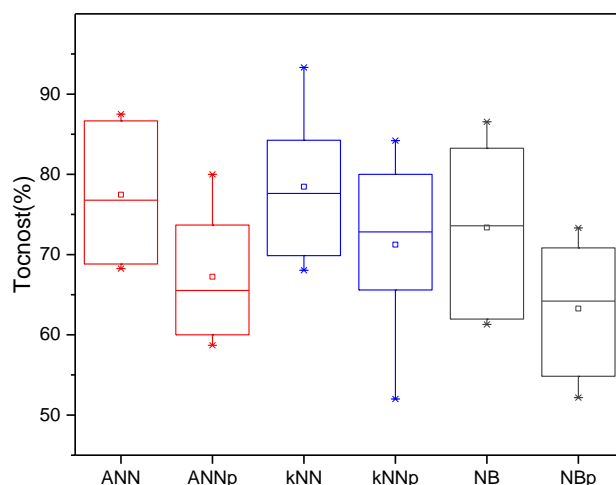
Slika 21: Usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata, puna dužina trajanja predmeta, po promatranim algoritmima, vrijeme kumulacije 1 sat, 2 kategorije



Slika 22: Usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata, puna dužina trajanja predmeta, po promatranim algoritmima, vrijeme kumulacije 1 sat, 3 kategorije



Slika 23: Usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata, puna dužina trajanja predmeta, po promatranim algoritmima, vrijeme kumulacije 1 dan, 2 kategorije



*Slika 24: Usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata, puna dužina trajanja predmeta, po promatranim algoritmima, vrijeme kumulacije 1 dan, 3 kategorije*

Rezultati Wilcoxonovog testa usklađenih parova (tablica 39) pokazuju da se u 12 od 14 promatranih grupa nulta hipoteza ne može prihvatiti, odnosno da modeli sa smanjenim brojem koeficijenata imaju i statistički značajno bolju točnost od modela s punim brojem koeficijenata. Usporedba točnosti po promatranim grupama prikazana je i grafički (slike 21, 22, 23, 24). U dva slučaja, primjenom algoritama SVM i kNN na grupu skupova podataka diskretiziranih na 2 kategorije, s vremenom kumulacije 1 sat, Wilcoxonov test usklađenih parova pokazuje da ne postoji statistički značajna razlika u točnosti predikcije modela sa smanjenim brojem koeficijenata i modela s punim brojem koeficijenata. U tablici 32 vidljivo je da skup podataka za predmet NWTIS 2K<sub>sat</sub> 13 - 2800 sati, pun broj koeficijenata, uz primjenu algoritama SVM i kNN daje točnost od 100%, što nije realno. Da je provedeno vrednovanje s 10 preklapanja, točnost bi bila manja. Kada umjesto ekstremne točnosti od 100% u grupu podataka uvrstimo prosječnu točnost algoritama ANN i NB za skup podataka NWTIS 2K<sub>sat</sub> 13 - 2800 sati, pun broj koeficijenata (90%), Wilcoxonov test usklađenih parova pokazuje da i uz primjenu algoritama

SVN i kNN postoji statistički značajna razlika u točnosti modela sa smanjenim brojem koeficijenata i modela s punim brojem koeficijenata.

Također, kada umjesto ekstremne točnosti od 100% u grupu podataka uvrstimo prosječnu točnost kNN u 8 vremenskih točaka za NWTIS 2K<sub>sat</sub> 13 - 2800 sati, pun broj koeficijenata (95,34%) i za SVM (94,17%), Wilcoxonov test usklađenih parova pokazuje da i uz primjenu algoritama SVN i kNN postoji statistički značajna razlika u točnosti predikcije modela sa smanjenim brojem koeficijenata i modela s punim brojem koeficijenata.

Ovime je Hipoteza 2 djelomično potvrđena.

Prosječni broj koeficijenata u modelima izgrađenim tehnikom sažimanja vremenskih nizova je 46, a prosječni broj koeficijenata u modelima s punim brojem koeficijenata je 188.

Usporedba točnosti napravljena je na svih 92 modela. Kod 7 modela, od ukupno 92 (manje od 8%) je točnost predikcije uz smanjen broj koeficijenata bila manja od točnosti uz puni broj koeficijenata (tablica 40).

Tablica 40: Usporedba točnosti (1=točnost uz smanjen broj koeficijenata je veća od točnosti uz puni broj koeficijenata, 0=točnost uz smanjen broj koeficijenata je manja od točnosti uz puni broj koeficijenata )

	ANN	SVM	kNN	NB
FIB 3K <sub>sat</sub> 13 - 2800 sati	1		1	1
FIB 3K <sub>sat</sub> 12 - 2800 sati	1		1	1
NWTIS 3K <sub>sat</sub> 13 - 2800 sati	1		1	1
NWTIS 3K <sub>sat</sub> 12 - 2800 sati	1		1	1
WDIP 3K <sub>sat</sub> 13 - 2800 sati	1		1	1
WDIP 3K <sub>sat</sub> 12 - 2800 sati	1		1	1
FIB 2K <sub>sat</sub> 13 - 2800 sati	1	1	1	1
FIB 2K <sub>sat</sub> 12 - 2800 sati	1	1	1	1
NWTIS 2K <sub>sat</sub> 13 - 2800 sati	1	0	0	1
NWTIS 2K <sub>sat</sub> 12 - 2800 sati	1	1	1	1
WDIP 2K <sub>sat</sub> 13 - 2800 sati	1	1	1	1
WDIP 2K <sub>sat</sub> 12 - 2800 sati	1	1	1	1
FIB 3K <sub>dan</sub> 13 - 120 dana	1		1	1
FIB 3K <sub>dan</sub> 12 - 120 dana	1		1	1
NWTIS 3K <sub>dan</sub> 13 - 120 dana	1		1	1
NWTIS 3K <sub>dan</sub> 12 - 120 dana	1		1	1
WDIP 3K <sub>dan</sub> 13 - 120 dana	1		1	1
WDIP 3K <sub>dan</sub> 12 - 120 dana	1		1	1
FIB 2K <sub>dan</sub> 13 - 120 dana	1	1	1	1
FIB 2K <sub>dan</sub> 12 - 120 dana	1	1	1	1
NWTIS 2K <sub>dan</sub> 13 - 120 dana	0	0	0	1
NWTIS 2K <sub>dan</sub> 12 - 120 dana	1	1	1	1
WDIP 2K <sub>dan</sub> 13 - 120 dana	1	1	1	1
WDIP 2K <sub>dan</sub> 12 - 120 dana	1	1	1	0
FMD 2K <sub>sat</sub> 13 - 120 sati	1	1	1	1
FMD 2K <sub>sat</sub> 12 - 120 sati	1	1	0	1

Zatim je cijeli postupak ponovljen i na skupovima podataka koji sadrže podatke o prve 2/3 dužine trajanja predmeta. Nad podacima o točnosti iz tablice 41, 43, 45 i 47 i odgovarajućim podacima o točnosti iz tablice 42, 44, 46 i 48, a prema odgovarajućim algoritmima, proveden je Wilcoxonov test usklađenih parova (tablica 49). Podaci iz navedenih tablica prikazani su i box-plotovima (slike 25, 26, 27 i 28). Konačno, na svih 92 modela uspoređena je točnost predikcije dobivena sa smanjenim brojem koeficijenata i točnost predikcije s punim brojem koeficijenata.

*Tablica 41: Točnost i broj koeficijenata po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 sat, 2 kategorije, primijenjena tehnika sažimanja*

	Točnost					MED	Broj koeficijenata				Pros. br. koef.
	ANN	SVM	kNN	NB	ANN		SVM	kNN	NB		
FIB 2K <sub>sat</sub> 13 - 1800 sati	83,87	84,95	83,87	76,34	<b>83,87</b>	92	295	15	8	<b>103</b>	
FIB 2K <sub>sat</sub> 12 - 1800 sati	84,78	78,26	84,79	78,26	<b>81,52</b>	18	242	26	4	<b>73</b>	
NWTIS 2K <sub>sat</sub> 13 - 1800 sati	93,45	93,45	93,32	99,06	<b>93,45</b>	4	4	4	4	<b>4</b>	
NWTIS 2K <sub>sat</sub> 12 - 1800 sati	89,47	91,67	93,33	94,68	<b>92,50</b>	14	22	10	19	<b>16</b>	
WDIP 2K <sub>sat</sub> 13 - 1800 sati	92,25	96,63	96,02	92,05	<b>94,14</b>	56	58	116	89	<b>80</b>	
WDIP 2K <sub>sat</sub> 12 - 1800 sati	91,67	91,67	95,83	91,57	<b>91,67</b>	15	13	193	64	<b>71</b>	
FMD 2K <sub>sat</sub> 13 - 80 sati	83,95	88,89	87,55	85,19	<b>86,37</b>	4	27	31	7	<b>17</b>	
FMD 2K <sub>sat</sub> 12 - 80 sati	90,79	78,95	90,81	81,58	<b>86,19</b>	22	38	8	6	<b>19</b>	
<b>MED</b>	<b>90,13</b>	<b>90,28</b>	<b>92,07</b>	<b>88,38</b>	<b>89,02</b>	<b>17</b>	<b>33</b>	<b>21</b>	<b>8</b>		



Tablica 42: Točnost i broj koeficijenata po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 sat, 2 kategorije, pun broj koeficijenata

	Točnost					Broj koeficijenata				
	ANN	SVM	kNN	NB	MED	ANN	SVM	kNN	NB	Prosj. broj koef.
FIB 2K <sub>sat</sub> 13 - 1800 sati	70,97	82,80	74,19	72,04	<b>73,12</b>	349	353	353	353	<b>352</b>
FIB 2K <sub>sat</sub> 12 - 1800 sati	71,74	76,09	79,35	69,57	<b>73,92</b>	349	353	403	403	<b>377</b>
NWTIS 2K <sub>sat</sub> 13 - 1800 sati	100,00	93,33	93,33	93,33	<b>93,33</b>	253	268	270	268	<b>265</b>
NWTIS 2K <sub>sat</sub> 12 - 1800 sati	73,68	70,83	86,67	78,95	<b>76,32</b>	253	315	263	563	<b>349</b>
WDIP 2K <sub>sat</sub> 13 - 1800 sati	80,00	84,00	84,00	84,00	<b>84,00</b>	403	353	353	403	<b>378</b>
WDIP 2K <sub>sat</sub> 12 - 1800 sati	79,17	70,83	83,33	79,17	<b>79,17</b>	313	313	313	313	<b>313</b>
FMD 2K <sub>sat</sub> 13 - 80 sati	79,01	82,72	82,72	70,37	<b>80,87</b>	38	38	38	38	<b>38</b>
FMD 2K <sub>sat</sub> 12 - 80 sati	89,47	77,63	86,84	50,00	<b>82,24</b>	38	38	38	38	<b>38</b>
<b>MED</b>	<b>79,09</b>	<b>80,18</b>	<b>83,67</b>	<b>75,50</b>	<b>80,02</b>	<b>283</b>	<b>314</b>	<b>292</b>	<b>333</b>	

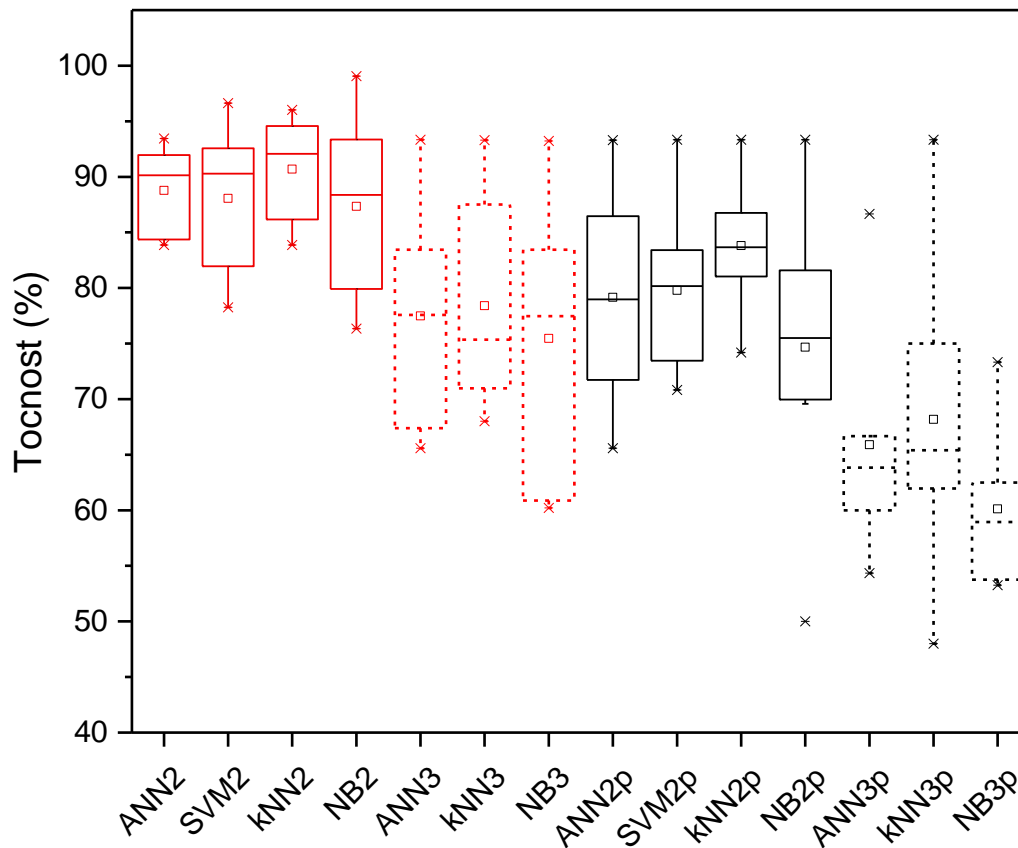
Tablica 43: Točnost i broj koeficijenata po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 sat, 3 kategorije, primijenjena tehnika sažimanja

	Točnost				Broj koeficijenata			
	ANN	kNN	NB	MED	ANN	kNN	NB	Prosj. broj koef.
FIB 3K <sub>sat</sub> 13 - 1800 sati	65,59	70,97	60,22	<b>65,59</b>	6	113	7	<b>42</b>
FIB 3K <sub>sat</sub> 12 - 1800 sati	67,39	71,74	60,87	<b>67,39</b>	8	363	381	<b>251</b>
NWTIS 3K <sub>sat</sub> 13 - 1800 sati	93,33	93,30	93,23	<b>93,30</b>	32	10	50	<b>31</b>
NWTIS 3K <sub>sat</sub> 12 - 1800 sati	78,95	78,95	78,90	<b>78,95</b>	40	83	30	<b>51</b>
WDIP 3K <sub>sat</sub> 13 - 1800 sati	76,21	68,01	76,05	<b>76,05</b>	20	117	282	<b>140</b>
WDIP 3K <sub>sat</sub> 12 - 1800 sati	83,45	87,50	83,45	<b>83,45</b>	4	112	16	<b>44</b>
<b>MED</b>	<b>77,58</b>	<b>75,35</b>	<b>77,48</b>	<b>77,50</b>	<b>14</b>	<b>113</b>	<b>40</b>	

Tablica 44: Točnost i broj koeficijenata po algoritmima za prve 2/3trajanja predmeta, vrijeme kumulacije je 1 sat, 3 kategorije, pun broj koeficijenata

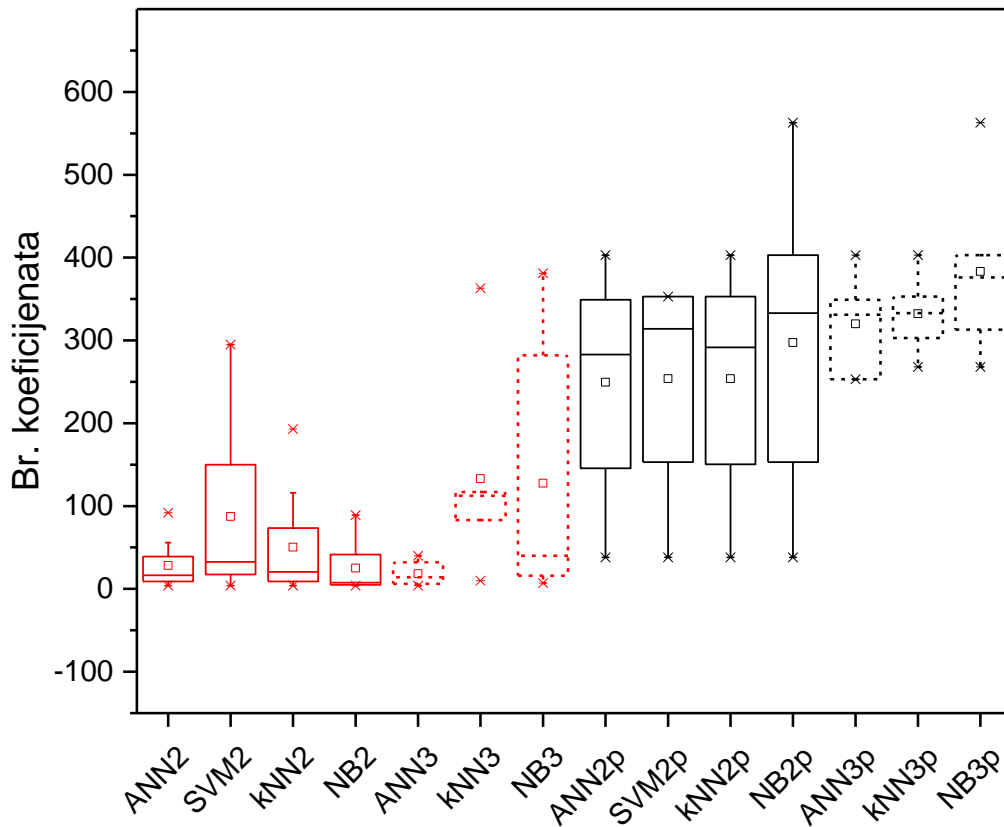
	Točnost				Broj koeficijenata			
	ANN	kNN	NB	<b>MED</b>	ANN	kNN	NB	<b>Prosj. broj koef.</b>
FIB 3K <sub>sat</sub> 13 - 1800 sati	64,52	62,37	53,76	<b>62,37</b>	349	353	349	<b>350</b>
FIB 3K <sub>sat</sub> 12 - 1800 sati	54,35	61,96	53,26	<b>54,35</b>	349	403	403	<b>385</b>
NWTIS 3K <sub>sat</sub> 13 - 1800 sati	86,67	93,33	73,33	<b>86,67</b>	253	268	268	<b>263</b>
NWTIS 3K <sub>sat</sub> 12 - 1800 sati	63,16	68,42	57,89	<b>63,16</b>	253	303	563	<b>373</b>
WDIP 3K <sub>sat</sub> 13 - 1800 sati	60,00	48,00	60,00	<b>60,00</b>	403	353	403	<b>386</b>
WDIP 3K <sub>sat</sub> 12 - 1800 sati	66,67	75,00	62,50	<b>66,67</b>	313	313	313	<b>313</b>
<b>MED</b>	<b>63,84</b>	<b>65,40</b>	<b>58,95</b>	<b>62,77</b>	<b>331</b>	<b>333</b>	<b>376</b>	

Podaci iz tablica 41, 42, 43 i 44 prikazani su i box-plotovima. Na slici 25 prikazani su podaci iz navedenih tablica koji se odnose na točnost. Crne linije označavaju točnost po algoritmima za modele temeljene na punom broju koeficijenata, a crvene linije označavaju točnost po algoritmima za modele temeljene na smanjenom broju koeficijenata, vrijeme kumulacije 1 sat (pune linije - 2 kategorije; isprekidane linije - 3 kategorije).



Slika 25: Točnost za prve 2/3 trajanja predmeta, vrijeme kumulacije 1 sat: crvene linije – točnost za smanjeni broj koeficijenata; crne linije - točnost za puni broj koeficijenata (pune linije - 2 kategorije; isprekidane linije - 3 kategorije)

Na slici 26 je box-plot s podacima iz tablica 41, 42, 43 i 44 koji se odnose na broj korištenih koeficijenata. Crne linije označavaju pun broj korištenih koeficijenata po algoritmima, a crvene linije označavaju smanjeni broj korištenih koeficijenata po algoritmima, vrijeme kumulacije 1 sat (pune linije - 2 kategorije; isprekidane linije - 3 kategorije).



*Slika 26: Broj koeficijenata za prve 2/3 trajanja predmeta, vrijeme kumulacije 1 sat: crvene linije –smanjeni broj koeficijenata; crne linije - puni broj koeficijenata (pune linije - 2 kategorije; isprekidane linije - 3 kategorije)*

Zatim su uspoređeni odgovarajući podaci o točnostima za vrijeme kumulacije 1 dan, a prema odgovarajućim algoritmima i skupovima podataka. Uspoređeni su podaci iz tablica 45 i 46 te 47 i 48.

Tablica 45: Točnost i broj koeficijenata po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 dan, 2 kategorije, primijenjena tehnika sažimanja

	Točnost					Broj koeficijenata				Prosje. br. koef.
	ANN	SVM	kNN	NB	MED	ANN	SVM	kNN	NB	
FIB 2K <sub>dan</sub> 13 - 80 dana	81,72	81,65	83,87	79,57	<b>81,69</b>	7	6	13	4	<b>8</b>
FIB 2K <sub>dan</sub> 12 - 80 dana	83,70	73,91	84,78	79,35	<b>81,53</b>	4	20	25	6	<b>14</b>
NWTIS 2K <sub>dan</sub> 13 - 80 dana	93,33	93,25	93,36	93,43	<b>93,35</b>	4	4	4	5	<b>4</b>
NWTIS 2K <sub>dan</sub> 12 - 80 dana	92,23	94,74	94,74	94,63	<b>94,69</b>	11	16	27	17	<b>18</b>
WDIP 2K <sub>dan</sub> 13 - 80 dana	92,05	88,35	92,03	96,15	<b>92,04</b>	10	13	7	9	<b>10</b>
WDIP 2K <sub>dan</sub> 12 - 80 dana	87,54	87,50	83,31	87,45	<b>87,48</b>	23	32	8	23	<b>22</b>
<b>MED</b>	<b>89,80</b>	<b>87,93</b>	<b>88,41</b>	<b>90,44</b>	<b>89,76</b>	<b>9</b>	<b>15</b>	<b>11</b>	<b>8</b>	

Tablica 46: Točnost i broj koeficijenata po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 dan, 2 kategorije, pun broj koeficijenata

	Točnost					Broj koeficijenata				Prosje. br. koef.
	ANN	SVM	kNN	NB	MED	ANN	SVM	kNN	NB	
FIB 2K <sub>dan</sub> 13 - 80 dana	74,19	77,42	82,80	67,74	<b>75,81</b>	38	38	35	38	<b>37</b>
FIB 2K <sub>dan</sub> 12 - 80 dana	79,35	71,74	81,52	63,04	<b>75,55</b>	27	38	37	38	<b>35</b>
NWTIS 2K <sub>dan</sub> 13 - 80 dana	80,00	100,00	100,00	93,33	<b>96,67</b>	39	38	38	38	<b>38</b>
NWTIS 2K <sub>dan</sub> 12 - 80 dana	84,21	78,95	89,47	89,47	<b>86,84</b>	38	38	38	38	<b>38</b>
WDIP 2K <sub>dan</sub> 13 - 80 dana	80,00	80,00	80,00	84,00	<b>80,00</b>	39	38	38	38	<b>38</b>
WDIP 2K <sub>dan</sub> 12 - 80 dana	75,00	75,00	75,00	79,17	<b>75,00</b>	39	38	38	38	<b>38</b>
<b>MED</b>	<b>79,68</b>	<b>78,19</b>	<b>82,16</b>	<b>81,59</b>	<b>77,90</b>	<b>39</b>	<b>38</b>	<b>38</b>	<b>38</b>	

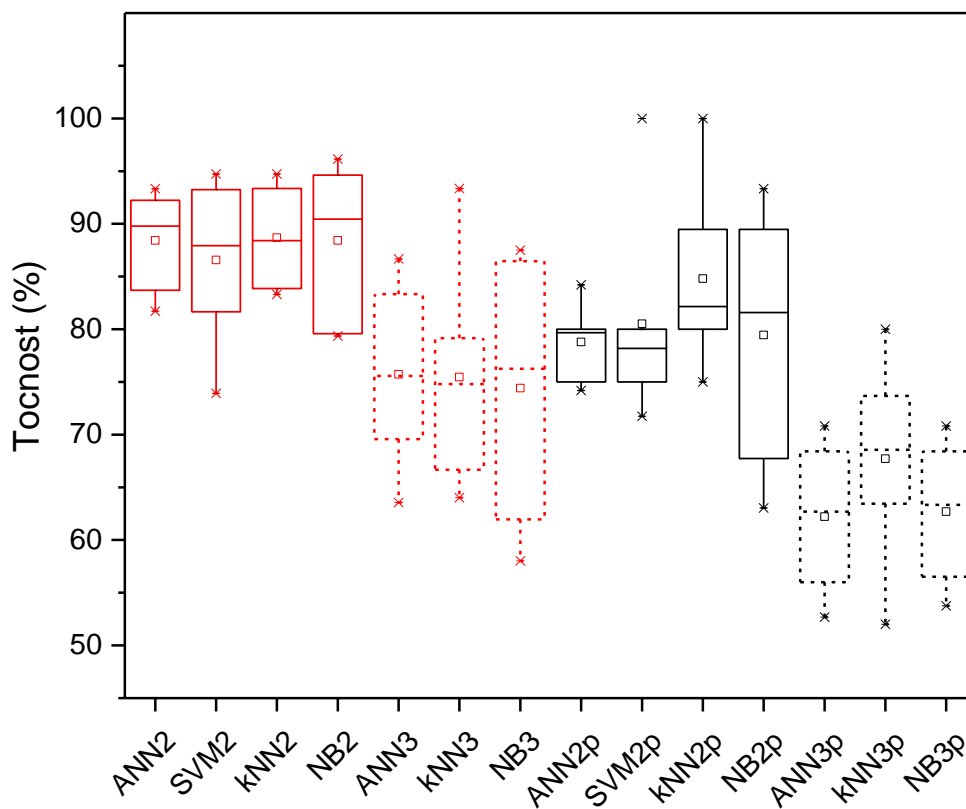
Tablica 47: Točnost i broj koeficijenata po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 dan, 3 kategorije, primijenjena tehnika sažimanja

	Točnost				Broj koeficijenata			Prosje. br. koef.
	ANN	kNN	NB	MED	ANN	kNN	NB	
FIB 3K <sub>dan</sub> 13 - 80 dana	63,55	66,67	58,03	<b>63,55</b>	7	4	5	<b>5</b>
FIB 3K <sub>dan</sub> 12 - 80 dana	69,57	70,65	61,96	<b>69,57</b>	12	5	5	<b>7</b>
NWTIS 3K <sub>dan</sub> 13 - 80 dana	86,67	93,35	86,45	<b>86,67</b>	4	10	22	<b>12</b>
NWTIS 3K <sub>dan</sub> 12 - 80 dana	78,95	78,95	84,34	<b>78,95</b>	7	15	27	<b>16</b>
WDIP 3K <sub>dan</sub> 13 - 80 dana	72,18	64,02	68,15	<b>68,15</b>	10	19	6	<b>12</b>
WDIP 3K <sub>dan</sub> 12 - 80 dana	83,33	79,15	87,50	<b>83,33</b>	23	5	5	<b>11</b>
<b>MED</b>	<b>75,57</b>	<b>74,80</b>	<b>76,25</b>	<b>74,26</b>	<b>9</b>	<b>8</b>	<b>6</b>	

Tablica 48: Točnost i broj koeficijenata po algoritmima za prve 2/3 trajanja predmeta, vrijeme kumulacije je 1 dan, 3 kategorije, pun broj koeficijenata

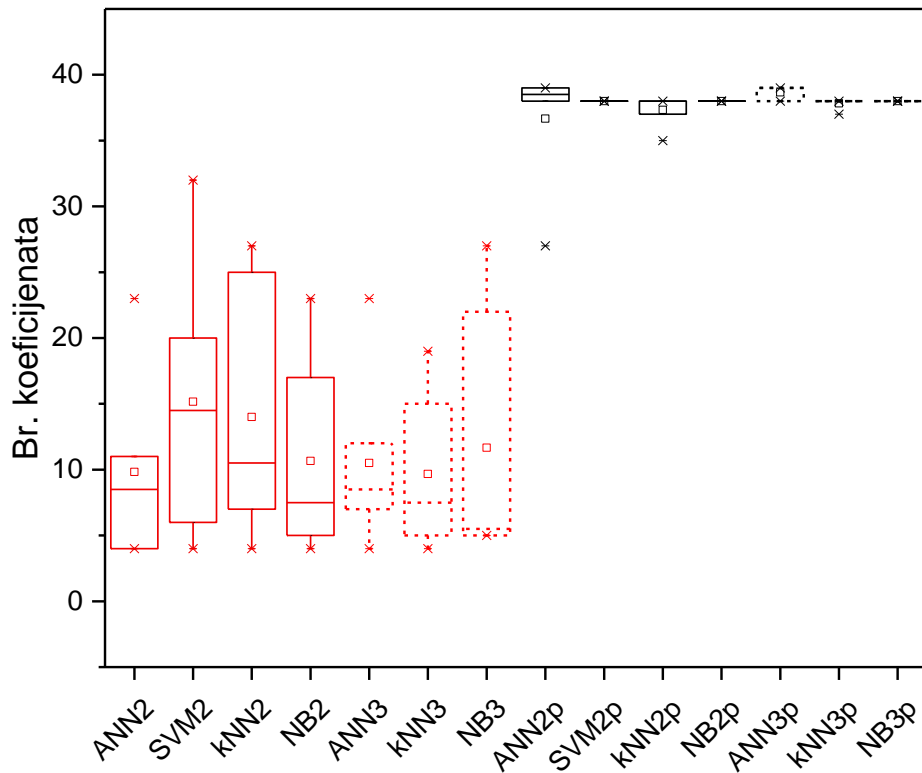
	Točnost				Broj koeficijenata			
	ANN	kNN	NB	<b>MED</b>	ANN	kNN	NB	<b>Pros. br. koef.</b>
FIB 3K <sub>dan</sub> 13 - 80 dana	52,69	63,44	53,76	<b>53,76</b>	39	38	38	<b>38</b>
FIB 3K <sub>dan</sub> 12 - 80 dana	58,70	66,30	56,52	<b>58,70</b>	38	37	38	<b>38</b>
NWTIS 3K <sub>dan</sub> 13 - 80 dana	66,67	80,00	66,67	<b>66,67</b>	39	38	38	<b>38</b>
NWTIS 3K <sub>dan</sub> 12 - 80 dana	68,42	73,68	68,42	<b>68,42</b>	38	38	38	<b>38</b>
WDIP 3K <sub>dan</sub> 13 - 80 dana	56,00	52,00	60,00	<b>56,00</b>	39	38	38	<b>38</b>
WDIP 3K <sub>dan</sub> 12 - 80 dana	70,83	70,83	70,83	<b>70,83</b>	39	38	38	<b>38</b>
<b>MED</b>	<b>62,69</b>	<b>68,57</b>	<b>63,34</b>	<b>62,69</b>	<b>39</b>	<b>38</b>	<b>38</b>	

Podaci iz tablica 45, 46, 47 i 48 prikazani su i box-plotovima. Na slici 27 prikazani su podaci iz navedenih tablica koji se odnose na točnost. Crne linije označavaju točnost po algoritmima za modele temeljene na punom broju koeficijenata, a crvene linije označavaju točnost po algoritmima za modele temeljene na smanjenom broju koeficijenata, vrijeme kumulacije 1 dan (pune linije - 2 kategorije; isprekidane linije - 3 kategorije).



Slika 27: Točnost za prve 2/3 trajanja predmeta, vrijeme kumulacije 1 dan: crvene linije – točnost za smanjeni broj koeficijenata; crne linije - točnost za puni broj koeficijenata (pune linije - 2 kategorije; isprekidane linije - 3 kategorije)

Na slici 28 je box-plot s podacima iz tablica 45, 46, 47 i 48 koji se odnose na broj korištenih koeficijenata. Crne linije označavaju pun broj korištenih koeficijenata po algoritmima, a crvene linije označavaju smanjeni broj korištenih koeficijenata po algoritmima, vrijeme kumulacije 1 dan (pune linije - 2 kategorije; isprekidane linije - 3 kategorije).



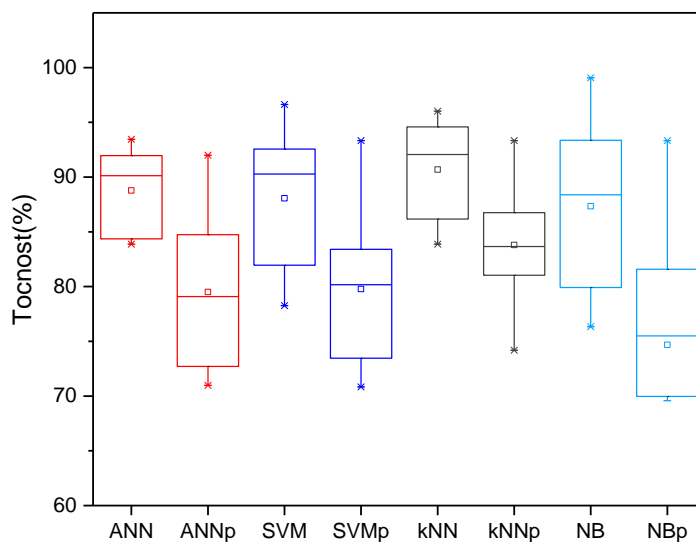
Slika 28: Broj koeficijenata za prve 2/3 trajanja predmeta, vrijeme kumulacije 1 dan: crvene linije – smanjeni broj koeficijenata; crne linije - puni broj koeficijenata (pune linije - 2 kategorije; isprekidane linije - 3 kategorije)

Usporedbom srednjih točnosti po algoritmima za modele izgrađene tehnikom sažimanja i modela izgrađenih s punim brojem koeficijenata, može se uočiti da modeli izgrađeni tehnikom sažimanja u pravilu imaju veću točnost predikcije. Za modele izgrađene tehnikom sažimanja srednja točnost je 82,76%, a za modele izgrađene s punim brojem koeficijenata srednja točnost je 72,94%. Radi li se i o statistički značajnoj razlici provjereno je Wilcoxonovim testom usklađenih parova.

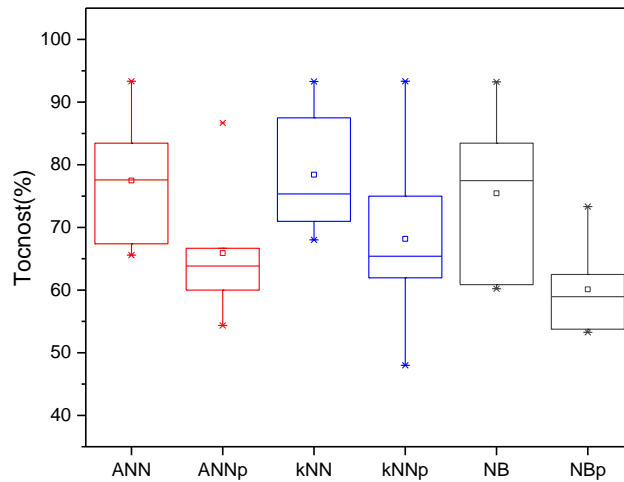


Tablica 49: Rezultati Wilcoxonovog testa usklađenih parova – usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova za sve skupove podataka, za prve 2/3 trajanja predmeta i po promatranim algoritmima s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata (0=nema razlike, 1=ima razlike)

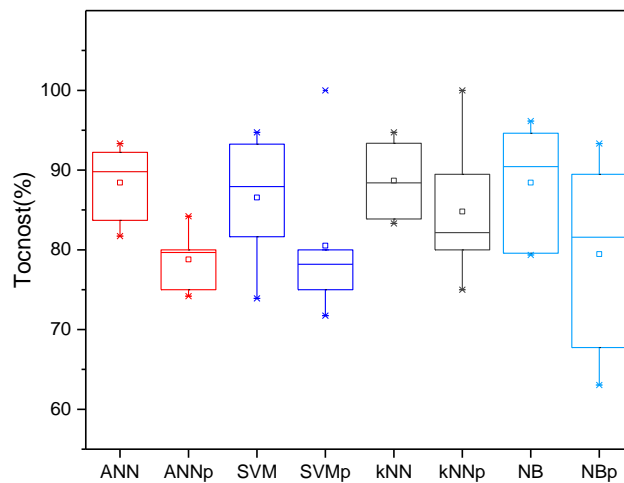
	Sati						Dani					
	2 kategorije			3 kategorije			2 kategorije			3 kategorije		
	Wilcoxon-razlika	MED grupe-točnost		Wilcoxon-razlika	MED grupe-točnost		Wilcoxon-razlika	MED grupe-točnost		Wilcoxon-razlika	MED grupe-točnost	
		Sažimanje	Pun broj koef.		Sažimanje	Pun broj koef.		Sažimanje	Pun broj koef.		Sažimanje	Pun broj koef.
ANN	1	90,13	79,09	1	77,58	63,84	1	89,80	79,68	1	75,57	62,69
SVM	1	90,28	80,18	n/a	n/a	n/a	0	87,93	78,19	n/a	n/a	n/a
kNN	1	92,07	83,67	1	75,35	65,40	0	88,41	82,16	1	74,80	68,57
NB	1	88,38	75,50	1	77,48	58,95	1	90,44	81,59	1	76,25	63,34



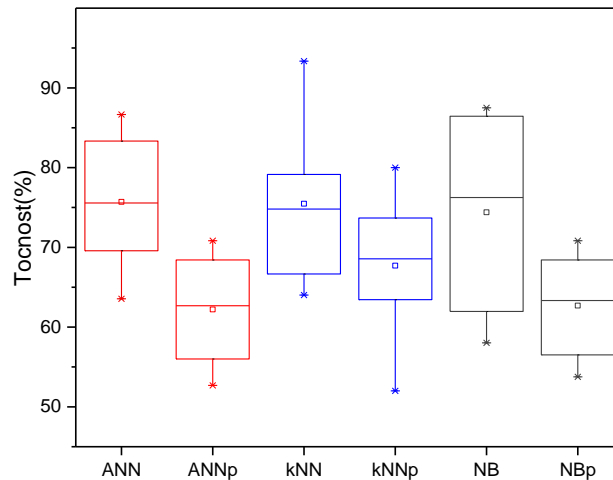
Slika 29: Usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata, prve 2/3 trajanja predmeta, po promatranim algoritmima, vrijeme kumulacije 1 sat, 2 kategorije



Slika 30: Usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata, prve 2/3 trajanja predmeta, po promatranim algoritmima, vrijeme kumulacije 1 sat, 3 kategorije



Slika 31: Usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata, prve 2/3 trajanja predmeta, po promatranim algoritmima, vrijeme kumulacije 1 dan, 2 kategorije



*Slika 32: Usporedba točnosti predikcije modela izgrađenih tehnikom sažimanja vremenskih nizova s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata, prve 2/3 trajanja predmeta, po promatranim algoritmima, vrijeme kumulacije 1 dan, 3 kategorije*

Rezultati Wilcoxonovog testa usklađenih parova (tablica 49) pokazuju da se u 12 od 14 promatranih grupa nulta hipoteza odbacuje, odnosno da modeli sa smanjenim brojem koeficijenata imaju i statistički značajno bolju točnost predikcije od modela s punim brojem koeficijenata. Usporedba točnosti po promatranim grupama prikazana je i grafički (slike 29, 30, 31, 32). U dva slučaja, primjenom algoritama SVM i kNN na grupu skupova podataka diskretiziranih na 2 kategorije, s vremenom kumulacije 1 dan, Wilcoxonov test usklađenih parova pokazuje da ne postoji statistički značajna razlika u točnosti modela predikcije sa smanjenim brojem koeficijenata i modela s punim brojem koeficijenata. U tablici 46 vidljivo je da skup podataka za predmet NWTIS 2K<sub>dan</sub> 13 - 80 dana, pun broj koeficijenata, uz primjenu algoritama SVM i kNN daje točnost od 100%, što nije realno. Da je provedeno vrednovanje s 10 preklapanja, točnost bi bila manja. Kada umjesto ekstremne točnosti od 100% u grupu podataka uvrstimo prosječnu točnost algoritama ANN i NB za skup podataka NWTIS 2K<sub>dan</sub> 13 - 80 dana, pun broj koeficijenata (90%), Wilcoxonov test usklađenih parova pokazuje da i uz primjenu algoritama SVN i kNN postoji statistički značajna razlika u točnosti modela sa smanjenim brojem koeficijenata i modela s punim brojem koeficijenata.

Međutim, kada umjesto ekstremne točnosti od 100% u grupu podataka uvrstimo prosječnu točnost kNN u 8 vremenskih točaka za NWTIS 2K<sub>dan</sub> 13 - 80 dana, pun broj koeficijenata (95,15%) i za SVM (94,44%), Wilcoxonov test usklađenih parova i dalje pokazuje da uz primjenu algoritama SVN i kNN ne postoji statistički značajna razlika u točnosti modela sa smanjenim brojem koeficijenata i modela s punim brojem koeficijenata. Takav rezultat može se objasniti karakteristikama SVM i kNN algoritama. Na složenost SVM modela ne utječe broj atributa. Broj potpornih vektora koje SVM algoritam odabire obično je mali. Zbog toga su SVM prikladni za situacije u kojim je broj atributa velik u odnosu na broj slučajeva (Kotsiantis i sur., 2006). Tako se može objasniti relativno velika točnost koju SVM algoritam postiže na modelima izgrađenim s punim brojem koeficijenata, odnosno statistički beznačajna razlika u točnosti modela kada se primjenjuje SVM algoritam uz primjenu tehnike sažimanja. Kod kNN se radi o neparometrijskoj metodi koja ne traži formulu odnosa između atributa i ciljne varijable niti određenu distribuciju skupa podataka (Altman, 1992). Kako je pokazano u dijelu 5.2.1., iako primjena određenog algoritma ne daje statistički značajno veću točnost, u najvećem broju promatranih slučajeva je upravo algoritam kNN dao najveću točnost.

Ovime je Hipoteza 2 djelomično prihvaćena i na skupu podataka o samo prve 2/3 trajanja predmeta.

Prosječni broj koeficijenata u modelima izgrađenim tehnikom sažimanja vremenskih nizova je 41, a prosječni broj koeficijenata u modelima s punim brojem koeficijenata je 178. Prosječna točnost u modelima izgrađenim tehnikom sažimanja vremenskih nizova je 83%, a prosječna točnost u modelima s punim brojem koeficijenata je 73%.

Usporedba točnosti napravljena je na svih 92 modela. Kod 5 modela, od ukupno 92 (približno 5%) je točnost predikcije uz smanjen broj koeficijenata bila manja od točnosti uz puni broj koeficijenata (tablica 50).

Tablica 50: Usporedba točnosti (1=točnost uz smanjen broj koeficijenata je veća od točnosti uz puni broj koeficijenata, 0=točnost uz smanjen broj koeficijenata je manja od točnosti uz puni broj koeficijenata), za modele temeljene na podacima o prve 2/3 trajanja predmeta

	ANN	SVM	kNN	NB
FIB 3K <sub>sat</sub> 13 - 1800 sati	1		1	1
FIB 3K <sub>sat</sub> 12 - 1800 sati	1		1	1
NWTIS 3K <sub>sat</sub> 13 - 1800 sati	1		0	1
NWTIS 3K <sub>sat</sub> 12 - 1800 sati	1		1	1
WDIP 3K <sub>sat</sub> 13 - 1800 sati	1		1	1
WDIP 3K <sub>sat</sub> 12 - 1800 sati	1		1	1
FIB 2K <sub>sat</sub> 13 - 1800 sati	1	1	1	1
FIB 2K <sub>sat</sub> 12 - 1800 sati	1	1	1	1
NWTIS 2K <sub>sat</sub> 13 - 1800 sati	0	1	0	1
NWTIS 2K <sub>sat</sub> 12 - 1800 sati	1	1	1	1
WDIP 2K <sub>sat</sub> 13 - 1800 sati	1	1	1	1
WDIP 2K <sub>sat</sub> 12 - 1800 sati	1	1	1	1
FIB 3K <sub>dan</sub> 13 - 80 dana	1		1	1
FIB 3K <sub>dan</sub> 12 - 80 dana	1		1	1
NWTIS 3K <sub>dan</sub> 13 - 80 dana	1		1	1
NWTIS 3K <sub>dan</sub> 12 - 80 dana	1		1	1
WDIP 3K <sub>dan</sub> 13 - 80 dana	1		1	1
WDIP 3K <sub>dan</sub> 12 - 80 dana	1		1	1
FIB 2K <sub>dan</sub> 13 - 80 dana	1	1	1	1
FIB 2K <sub>dan</sub> 12 - 80 dana	1	1	1	1
NWTIS 2K <sub>dan</sub> 13 - 80 dana	1	0	0	1
NWTIS 2K <sub>dan</sub> 12 - 80 dana	1	1	1	1
WDIP 2K <sub>dan</sub> 13 - 80 dana	1	1	1	1
WDIP 2K <sub>dan</sub> 12 - 80 dana	1	1	1	1
FMD 2K <sub>sat</sub> 13 - 80 sati	1	1	1	1
FMD 2K <sub>sat</sub> 12 - 80 sati	1	1	1	1

### 5.3.3. Treća hipoteza

*H3: Točnost predikcije modela je jednaka ili veća od 75% kada se modelira temeljem podataka o prve 2/3 trajanja predmeta.*

Naznaka da će biti moguće potvrditi ovu hipotezu opisana je u dijelu 5.2.2. gdje su uspoređivane točnosti u 8 vremenskih točaka tijekom i po završetku predmeta. U dijelu 5.2.2. pokazano je da u čak 2/3 svih ispitanih slučajeva modeli postižu jednako dobru točnost predikcije bez obzira temelji li se izračun na skupu podataka za cijelo vrijeme trajanja predmeta ili na skupu podataka koji pokriva puno kraće vremensko razdoblje.

S obzirom na to da su FIB, NWTIS i WDIP jednosemestralni predmeti, kao 2/3 trajanja predmeta uzeto je 1800 sati, odnosno 80 dana. Predmet FMD traje 5 dana, pa je kao 2/3 trajanja predmeta uzeto 80 sati. Točnosti predikcije za modele temeljene na podacima o prve 2/3 trajanja predmeta prikazane su u tablici 51. Crveno su označeni redovi u kojima je srednja točnost manja od 75%.

Iz tablice 51 je vidljivo da je točnost predikcije uvijek veća od 75% za modele u kojima je zavisna varijabla diskretizirana na 2 kategorije. U tim slučajevima se H3 se prihvaća.

Kada je zavisna varijabla diskretizirana na 3 kategorije, H3 nije istinita za sve slučajeve. To je očekivano jer je iz relevantne literature vidljivo da točnost predviđanja studentskog modela ovisi i o broju klasa ciljne varijable. U pravilu, što je veći broj klasa ciljne varijable, točnost je manja. U radu autora Minaei-Bidgoli i sur. (2003) postignuta točnost tehnikom kombinacije više klasifikatora na 2 klase je 94,09%, na 3 klase je 72,13% i na 9 klasa je 62,25%. Autori Thai-Nghe i sur. (2007) su na uzorku od 20492 studenata, korištenjem algoritma stabla odlučivanja postigli točnost 73% za 4 klase uspjeha, a za 2 klase 93-94%. Autori Romero, Ventura, Espejo i Hervas (2008) pri diskretizaciji na 4 klase ne postižu točnost veću od 70%, a koristili su čak 24 algoritma na uzorku od 438 studenata. Autori Gamulin i sur. (2014) na uzorku od 302 studenta i primjenom 5 algoritama postigli su točnost od 84-94% kada je ciljna varijabla diskretizirana u 2 kategorije, a 58-77% kada je diskretizirana u 3 kategorije. Drugim riječima, bilo kakvi se modeli i algoritmi koristili, točnost je niža što je ciljna varijabla diskretizirana u više kategorija.

Tablica 51: Točnost predikcije za modele temeljene na podacima o prve 2/3 trajanja predmeta, primijenjena tehnika sažimanja. Crveno su označeni redovi u kojima je srednja točnost manja od

75%

	Točnost				Srednja točnost
	ANN	SVM	kNN	NB	
FIB 3K <sub>sat</sub> 13 - 1800 sati	65,59		70,97	60,22	65,59
FIB 3K <sub>sat</sub> 12 - 1800 sati	67,39		71,74	60,87	66,67
NWTIS 3K <sub>sat</sub> 13 - 1800 sati	93,33		93,3	93,23	93,29
NWTIS 3K <sub>sat</sub> 12 - 1800 sati	78,95		78,95	78,9	78,93
WDIP 3K <sub>sat</sub> 13 - 1800 sati	76,21		68,01	76,05	73,42
WDIP 3K <sub>sat</sub> 12 - 1800 sati	83,45		87,5	83,45	84,80
FIB 2K <sub>sat</sub> 13 - 1800 sati	83,87	84,95	83,87	76,34	82,26
FIB 2K <sub>sat</sub> 12 - 1800 sati	84,78	78,26	84,79	78,26	81,52
NWTIS 2K <sub>sat</sub> 13 - 1800 sati	93,45	93,45	93,32	99,06	94,82
NWTIS 2K <sub>sat</sub> 12 - 1800 sati	89,47	91,67	93,33	94,68	92,29
WDIP 2K <sub>sat</sub> 13 - 1800 sati	92,25	96,63	96,02	92,05	94,24
WDIP 2K <sub>sat</sub> 12 - 1800 sati	91,67	91,67	95,83	91,57	92,69
FIB 3K <sub>dan</sub> 13 - 80 dana	63,55		66,67	58,03	62,75
FIB 3K <sub>dan</sub> 12 - 80 dana	69,57		70,65	61,96	67,39
NWTIS 3K <sub>dan</sub> 13 - 80 dana	86,67		93,35	86,45	88,82
NWTIS 3K <sub>dan</sub> 12 - 80 dana	78,95		78,95	84,34	80,75
WDIP 3K <sub>dan</sub> 13 - 80 dana	72,18		64,02	68,15	68,12
WDIP 3K <sub>dan</sub> 12 - 80 dana	83,33		79,15	87,5	83,33
FIB 2K <sub>dan</sub> 13 - 80 dana	81,72	81,65	83,87	79,57	81,70
FIB 2K <sub>dan</sub> 12 - 80 dana	83,70	73,91	84,78	79,35	80,44
NWTIS 2K <sub>dan</sub> 13 - 80 dana	93,33	93,25	93,36	93,43	93,34
NWTIS 2K <sub>dan</sub> 12 - 80 dana	92,23	94,74	94,74	94,63	94,09
WDIP 2K <sub>dan</sub> 13 - 80 dana	92,05	88,35	92,03	96,15	92,15
WDIP 2K <sub>dan</sub> 12 - 80 dana	87,54	87,5	83,31	87,45	86,45
FMD 2K <sub>sat</sub> 13 - 80 sati	83,95	88,89	87,55	85,19	86,40
FMD 2K <sub>sat</sub> 12 - 80 sati	90,79	78,95	90,81	81,58	85,53
<b>Srednja vrijednost</b>	<b>83,08</b>	<b>87,42</b>	<b>83,88</b>	<b>81,86</b>	<b>82,76</b>

Modeli za predmet FIB - u obje akademske godine, diskretizirano u 3 kategorije, vrijeme kumulacije dan i sat, imaju srednju točnost između 63% i 68%. Predmet WDIP 2013,

diskretizirano u 3 kategorije, vrijeme kumulacije dan i sat, ima srednju točnost između 68% i 73%.

Čak i za punu dužinu trajanja predmeta, navedeni predmeti (FIB2012 i FIB2013 i WDIP2013) imaju točnost graničnu ili manju od 75%. Predmet FIB ima za punu dužinu trajanja predmeta srednju točnost između 68% i 72%, a predmet WDIP2013 ima srednju točnost za punu dužinu trajanja predmeta između 68% i 76%. Temeljem toga može se zaključiti da je rana predikcija moguća no točnost vjerojatno dosta ovisi o gustoći podataka. Ako kao gustoću podataka uzmemo ukupan broj klikova na predmetu podijeljen s umnoškom broja studenata i dana trajanja predmeta, može se vidjeti da upravo predmeti FIB2012 i FIB2013 imaju gustoću podataka ispod 10 (FIB2012=2,9 i FIB2013=2,7) dok ostali predmeti imaju gustoću podataka od 12 do 16. Predmet WDIP2013 ima zadovoljavajuću gustoću, no uvidom u podatke o elementima i aktivnostima na LMS-u te usporedbom tih podataka za predmet WDIP2013 i WDIP2012 (koji ima zadovoljavajuću točnost i za podatke diskretizirane u tri kategorije) može se primijetiti da se struktura tih podataka razlikuje. Na predmetu WDIP2013, od svih pristupa predmetu bilo je 12% klikova na materijale koji se odnose na zadaće i projektne zadatke, a 9% klikova na forume. Na predmetu WDIP2012 bilo je 17% klikova na materijale koji se odnose na zadaće i projektne zadatke, a 3% klikova na forume, dok je struktura ostalih klikova ostala ista. Je li tako izmijenjena struktura klikova utjecala na točnost bit će predmet budućih istraživanja.



## 6. SINTEZA REZULTATA

### 6.1. Sinteza rezultata testiranja hipoteza

Usporedbom točnosti predikcije po algoritmima pokazano je da primjena različitih algoritama za izračunavanje točnosti predikcije ne daje statistički značajno različite rezultate. Testirane su razlike u točnosti predikcije za modele sa smanjenim brojem koeficijenata, dakle, gdje je primijenjena tehnika sažimanja, i to za punu dužinu predmeta, a potom i za podatke o prve dvije trećine trajanja predmeta. Algoritam kNN u pravilu daje najbolji rezultat, a algoritam NB u pravilu daje najlošiji rezultat. Budući da algoritam NB polazi od pretpostavke da nema korelacije među atributima, vjerojatno je to razlog nešto niže točnosti.

U drugom koraku provedena je usporedba točnosti predikcije po dužini vremenskog niza. Uspoređivane su točnosti za 8 vremenskih točaka kada je vrijeme kumulacije 1 sat, tj. za prvih 1000 sati, 1400 sati, 1800 sati, 2000 sati, 2150 sati, 2500 sati, 2800 sati i maksimalan broj sati. Također su uspoređivane točnosti za 6 vremenskih točaka kada je vrijeme kumulacije 1 dan, tj. za prvih 60 dana, 80 dana, 100 dana, 120 dana, 150 dana i maksimalan broj dana. Za predmet Fizika medicinske dijagnostike, koji je specifičan zbog svog kratkog trajanja, uspoređivane su točnosti za 6 vremenskih točaka, tj. za prvih 60 sati, 80 sati, 100 sati, 120 sati, 150 sati i maksimalan broj sati. Sve je to napravljeno i za skupove kod kojih su zavisne varijable diskretizirane u 2 i u 3 kategorije. Postupak je ponovljen za sva 4 algoritma. Po grupama predmeta koji su usporedivi s obzirom na dužine vremenskih nizova, broj klasa i vrijeme kumulacije, proveden je Friedmanov test kako bi se provjerilo ima li unutar grupe statistički značajne razlike u točnosti predikcije za različite dužine vremenskog niza. Od ukupno 18 rezultata, u 6 slučajeva je (tablica 19) Friedmanov test pokazao da postoje razlike u točnosti ako se uspoređuju točnosti predikcije za različite promatrane dužine vremenskog niza. Drugim riječima, u približno 67% slučajeva modeli postižu jednako dobru točnost predikcije bez obzira temelji li se izračun na skupu podataka za cijelo vrijeme trajanja predmeta ili na skupu podataka koji pokriva puno kraće vremensko razdoblje. To je bio dobar pokazatelj da će rana predikcija biti moguća, a što je predmet Hipoteze 3.

Treći korak bio je usporedba točnosti predikcije po vremenu kumulacije. Ispitano je postoji li statistički značajna razlika između točnosti izračunate na skupovima podataka s vremenom kumulacije jedan sat i na skupovima podataka s vremenom kumulacije jedan dan. Modeli su

temeljeni na podacima o punoj dužini trajanja predmeta. Rezultati Wilcoxonovog testa usklađenih parova pokazali su da nema statistički značajne razlike u točnosti kada po pojedinim algoritmima usporedimo rezultate temeljene na skupovima podataka s vremenom kumulacije 1 sat i 1 dan. Ograničavajući čimbenik je gustoća pristupa LMS-u. Ako je gustoća premala, nije moguće napraviti model predikcije koji bi davao zadovoljavajuće rezultate.

Zatim je isti postupak ponovljen i za modele temeljene na podacima o prve 2/3 trajanja predmeta i također je pokazao da nema statistički značajne razlike u točnosti kada po pojedinim algoritmima usporedimo rezultate temeljene na skupovima podataka s vremenom kumulacije 1 sat i 1 dan.

Za potrebe hipoteze 1, koja glasi:

*Točnost predikcije uspjeha studenata (na konačnom ispitu iz određenog predmeta) pomoću modela izgrađenog dubinskom analizom podataka iz hibridnog okruženja za učenje, a primjenom analize vremenskih nizova logova studenata u LMS-u u frekvencijskoj domeni, jednaka je ili veća od točnosti predikcije pomoću sličnih modela opisanih u literaturi.*

-slični modeli su definirani kao oni koji imaju uspjeh studenata (ocjena) kao zavisnu varijablu i koji se odnose na hibridno okruženje za učenje. Poblježe je opisano 5 modela koji zadovoljavaju navedene uvjete. Pri tom je važno naglasiti da su to samo slični modeli jer su u ovom radu razmatrani modeli temeljeni na vremenskim nizovima klikova studenata na LMS-u, a u sličnim modelima radi se aktivnosti na forumima (Romero i sur., 2013) ili rezultatima zadataka, kvizova i aktivnosti na forumima (Zafra i sur., 2011), rezultatima zadataka, aktivnosti na forumima i vremenu (Mineai-Bidgoli i sur., 2003), segmentaciji aktivnosti po vremenskim razdobljima (Delgado i sur., 2006), rezultatima kvizova, testova i vremenu potrebnom za rješavanje (Gamulin i sur., 2014).

Zbog preglednosti i lakše usporedbe, kao referentne vrijednosti za ovaj rad odabrane su točnosti predikcije za modele sagrađene temeljem podataka o punoj dužini trajanja sva četiri promatrana predmeta, gdje je zavisna varijabla diskretizirana na 2 kategorije, a podaci se odnose 2012./2013. godinu.

Usporedbom s gore navedenim radovima prihvaćena je H1, odnosno pokazano je da je točnost predikcije uspjeha studenata (na konačnom ispitu iz određenog predmeta) pomoću modela

izgrađenog dubinskom analizom podataka iz hibridnog okruženja za učenje, a primjenom analize vremenskih nizova logova studenata u LMS-u u frekvencijskoj domeni, jednaka ili veća od točnosti predikcije pomoću sličnih modela opisanih u literaturi.

U nastavku je na pregledu modela opisanih u literaturi, a koji ne zadovoljavaju usku definiciju iz H1, pokazano da je točnost ovdje predloženih modela usporediva i s točnošću modela koji se odnose na online okruženje za učenje ili modela koji predviđaju neku drugu kategoriju studenata.

Za potrebe hipoteze 2, koja glasi:

*Primjenom tehnike sažimanja vremenskih nizova postiže se bolja točnost predikcije u odnosu na modele temeljene na punom skupu podataka.*

-uspoređena je točnost predikcije modela izgrađenih predloženom tehnikom sažimanja vremenskih nizova i to za sve skupove podataka u punoj dužini trajanja predmeta i po promatranim algoritmima s točnošću predikcije odgovarajućih modela koji sadrže pun broj koeficijenata tj. na kojima nije provedena tehnika sažimanja. Pri tom je uz svaku izračunatu točnost naveden i primijenjeni broj koeficijenata. Usporedbom srednjih točnosti po algoritmima za modele izgrađene tehnikom sažimanja i modela izgrađenih s punim brojem koeficijenata, može se uočiti da modeli izgrađeni tehnikom sažimanja u pravilu imaju veću točnost predikcije. Za modele izgrađene tehnikom sažimanja srednja točnost je 83,26%, a za modele izgrađene s punim brojem koeficijenata srednja točnost je 73,68%. Wilcoxonovim testom usklađenih parova provjereno je radi li se i o statistički značajnoj razlici.

Rezultati Wilcoxonovog testa usklađenih parova pokazuju da se u 12 od 14 promatranih grupa nulta hipoteza treba odbaciti, odnosno da modeli sa smanjenim brojem koeficijenata imaju i statistički značajno bolju točnost od modela s punim brojem koeficijenata. U dva slučaja, primjenom algoritama SVM i kNN na grupu skupova podataka diskretiziranih na 2 kategorije, s vremenom kumulacije 1 sat, Wilcoxonov test usklađenih parova pokazuje da ne postoji statistički značajna razlika u točnosti predikcije modela sa smanjenim brojem koeficijenata i modela s punim brojem koeficijenata. Budući da skup podataka za predmet NWTIS 2K<sub>sat</sub> 13 - 2800 sati, pun broj koeficijenata, uz primjenu algoritama SVM i kNN daje točnost od 100%, što nije realno, u grupu podataka uvrštena je prosječna točnost algoritama ANN i NB za skup podataka NWTIS 2K<sub>sat</sub> 13 - 2800 sati, pun broj koeficijenata (90%). Tada Wilcoxonov test usklađenih parova pokazuje da i uz primjenu algoritama SVN i kNN postoji statistički značajna

razlika u točnosti modela sa smanjenim brojem koeficijenata i modela s punim brojem koeficijenata.

Također, kada umjesto ekstremne točnosti od 100% u grupu podataka uvrstimo prosječnu točnost kNN u 8 vremenskih točaka za NWTIS 2K<sub>sat</sub> 13 - 2800 sati, pun broj koeficijenata (95,34%) i za SVM (94,17%), Wilcoxonov test usklađenih parova pokazuje da i uz primjenu algoritama SVN i kNN postoji statistički značajna razlika u točnosti predikcije modela sa smanjenim brojem koeficijenata i modela s punim brojem koeficijenata.

Prosječni broj koeficijenata u modelima izgrađenim tehnikom sažimanja vremenskih nizova je 46, a prosječni broj koeficijenata u modelima s punim brojem koeficijenata je 188.

Zatim je cijeli postupak ponovljen i na skupovima podataka koji sadrže podatke o prve 2/3 dužine trajanja predmeta. Prosječni broj koeficijenata u modelima izgrađenim tehnikom sažimanja vremenskih nizova je 41, a prosječni broj koeficijenata u modelima s punim brojem koeficijenata je 178. Prosječna točnost u modelima izgrađenim tehnikom sažimanja vremenskih nizova je 83%, a prosječna točnost u modelima s punim brojem koeficijenata je 73%.

Rezultati Wilcoxonovog testa usklađenih parova pokazuju da se u 12 od 14 promatranih grupa nulta hipoteza treba odbaciti, odnosno da modeli sa smanjenim brojem koeficijenata imaju i statistički značajno bolju točnost predikcije od modela s punim brojem koeficijenata. U dva slučaja, primjenom algoritama SVM i kNN na grupu skupova podataka diskretiziranih na 2 kategorije, s vremenom kumulacije 1 dan, Wilcoxonov test usklađenih parova pokazuje da ne postoji statistički značajna razlika u točnosti predikcije modela sa smanjenim brojem koeficijenata i modela s punim brojem koeficijenata.

Time je djelomično potvrđena H2, tj. da se primjenom tehnike sažimanja vremenskih nizova postiže bolja točnost predikcije u odnosu na modele temeljene na punom skupu podataka.

Za hipotezu 3, koja glasi:

*Točnost predikcije modela je jednaka ili veća od 75% kada se modelira temeljem podataka o prve 2/3 trajanja predmeta.*

-postojala je naznaka da će biti moguće potvrditi ovu hipotezu i to kada su uspoređivane točnosti u 8 vremenskih točaka tijekom i po završetku predmeta. U dijelu 5.2.2. pokazano je da u čak 2/3 svih ispitanih slučajeva modeli postižu jednako dobru točnost predikcije bez obzira temelji li se izračun na skupu podataka za cijelo vrijeme trajanja predmeta ili na skupu podataka koji pokriva puno kraće vremensko razdoblje.

Kako su FIB, NWTIS i WDIP jednosemestralni predmeti, kao 2/3 trajanja predmeta uzeto je 1800 sati, odnosno 80 dana. Predmet FMD traje 5 dana, pa je kao 2/3 trajanja predmeta uzeto 80 sati.

Iz rezultata je vidljivo da je točnost predikcije uvijek veća od 75% za modele u kojima je zavisna varijabla diskretizirana na 2 kategorije. U tim slučajevima se H3 se prihvaća.

Kada je zavisna varijabla diskretizirana na 3 kategorije, H3 nije potvrđena za sve slučajeve. Rezultat je očekivan jer je iz relevantne literature vidljivo je da točnost predviđanja studentskog modela ovisi i o broju klasa ciljne varijable. U pravilu, što je veći broj klasa ciljne varijable, točnost je manja što je vidljivo iz relevantne literature (Minaei-Bidgoli i sur., 2003; Thai-Nghe i sur., 2007; Romero, Ventura, Espejo i Hervas, 2008.)

Predmeti (skupovi podataka) kod kojih je uočena točnost niža od 75% čak i za punu dužinu trajanja predmeta, imaju točnost graničnu ili manju od 75%. Predmet FIB ima za punu dužinu trajanja predmeta srednju točnost između 68% i 72%, a predmet WDIP2013 ima srednju točnost za punu dužinu trajanja predmeta između 68% i 76%. Temeljem toga može se zaključiti da je rana predikcija moguća no točnost ovisi i o gustoći podataka. Ako gustoću podataka definiramo kao ukupan broj klikova na predmetu podijeljen s umnoškom broja studenata i dana trajanja predmeta, može se vidjeti da upravo predmeti FIB2012 i FIB2013 imaju gustoću podataka ispod 10 (FIB2012=2,9 i FIB2013=2,7) dok ostali predmeti imaju gustoću podataka od 12 do 16. Predmet WDIP2013 ima zadovoljavajuću gustoću, no uvidom u podatke o elementima i aktivnostima na LMS-u te usporedbom tih podataka za predmet WDIP2013 i WDIP2012 (koji ima zadovoljavajuću točnost i za podatke diskretizirane u tri kategorije) može se primijetiti da se struktura tih podataka razlikuje. Na predmetu WDIP2013, od svih pristupa predmetu bilo je 12% klikova na materijale koji se odnose na zadaće i projektne zadatke, a 9% klikova na forume. Na predmetu WDIP2012 bilo je 17% klikova na materijale koji se odnose na zadaće i projektne

zadatke, a 3% klikova na forume, dok je struktura ostalih klikova ostala ista. Je li tako izmijenjena struktura klikova utjecala na točnost bit će predmet budućih istraživanja.

## 6.2. Doprinosi i ograničenja istraživanja

Doprinosi:

U ovom istraživanju korištena su 2 vremena kumulacije  $T$ ,  $T_{sat}=1$  sat, i  $T_{dan}=1$  dan. To znači da je bilježen broj studentskih klikova u jednom satu, odnosno u jednom danu. U relevantnoj literaturi je pronađena vremenska segmentacija atributa u periodu od mjesec dana (Delgado i sur., 2006), tjedan dana (Lara i sur., 2014; Dias i Diniz, 2013), ali ne po danu ili satu. Xie i sur. (2014) su koristili vrijeme kumulacije od 1 minute no radilo se o samo jednom predmetu i samo 20 studenata. Dobashi (2015) je koristio vrijeme kumulacije od 5 minuta no također na podacima za jedan predmet i 61 studenta. Na tako malim uzorcima još se ne pojavljuju problemi vezani uz preveliku (teško savladivu) količinu podataka, pa se može ustvrditi da se radi o doprinosu na ovom području istraživanja.

Za potrebe ovog rada obrađen je veliki broj podataka iz LMS-a na Sveučilištu u Zagrebu, s Medicinskog fakulteta te s Fakulteta organizacije i informatike. Veličina ukupnog uzorka je oko 1400 studenata. Radi se o različitim tipovima studenata na dva fakulteta s različitom frekvencijom pristupa. Jedino što je zajedničko je postojanje i korištenje LMS-a u klasičnoj nastavi. Pregledom relevantne literature ustanovljeno je da se radi o jednom od većih uzoraka iz hibridnog okruženja za učenje.

Također, uzorak sadrži podatke o studentima koji su po svojim interesima i navikama u hibridnom okruženju različiti. Najbolji pokazatelj toga je ukupan broj logova po predmetu. Na Fakultetu organizacije i informatike radi se o približno 100.000 logova po predmetu, baš kao i na jednosemestralnom predmetu s Medicinskog fakulteta, iako je na Fakultetu organizacije i informatike upisano višestruko manje studenata u promatrane predmete. To se može objasniti činjenicom da su studenti Fakulteta organizacije i informatike više usmjereni na korištenje informacijsko-komunikacijskih tehnologija. Time je pokazano da su predloženi modeli robusni i

da funkcioniraju na fundamentalno različitim podacima, što također nije čest primjer u literaturi i predstavlja doprinos ovog istraživanja.

Također, autorica nije pronašla rad u kojem je spektralna analiza, odnosno analiza vremenskih nizova u frekvencijskoj domeni korištena na skupu podataka iz hibridnog okruženja za učenje. Prebacivanjem podataka iz vremenske u frekvencijsku domenu te primjenom tehnike za uklanjanje šuma, smanjena je količina podataka potrebna za uspješno modeliranje, a također su otkrivene periodičnosti karakteristične za pojedine studentske grupe. To su periodičnosti koje u vremenskoj domeni nisu vidljive, a koje su ključ modeliranja za ovdje predložene modele.

Najvažniji je doprinos što je pokazano da je moguće napraviti model predikcije uspjeha studenata temeljem podataka iz LMS-a primijenjenog u hibridnom okruženju za učenje i to rudarenjem podataka o vremenskim nizovima pristupa studenata LMS-u te uz primjenu analize vremenskih nizova u frekvencijskoj domeni. Pokazano je da je predloženi model usporediv ili bolji od do sada postojećih modela te da daje kvalitetnu predikciju uspjeha studenata. Modelira se temeljem podataka koji se automatski prikupljaju tijekom nastave, dakle bez zadiranja u nastavni proces.

Ograničenja:

Kategorije uspješnosti studenata su formirane tako da ocjene 1 i 2 upadaju u istu kategoriju. To je učinjeno kako bi distribucija ispitanika po kategorijama bila što ujednačenija. Iako takvo formiranje kategorija može predstavljati izvjesno ograničenje, smještanje studenata s ocjenama 1 i 2 istu kategoriju ima svoje opravdanje jer i studenti s ocjenom dovoljan predstavljaju rizične studente i potrebna im je pomoć. Stoga navedeno ograničenje može, zbog moguće praktične primjene, imati vrijednost.

U radu je ispitivana samo točnost kao mjera vrednovanje klasifikatora. Skupovi prezentirani u radu su ujednačeni, pa analiziranje isključivo točnosti kao pokazatelja ima svoje opravdanje.

U budućim istraživanjima potrebno je uključiti u ispitivanje i druge mjere (preciznost, odziv, specifičnost), odnosno F –mjera, koja predstavlja harmonijsku sredinu preciznosti i odziva.

Također, potrebno je sustavno provesti unakrsno vrednovanje s 5 ili 10 preklapanja kako bi se dobila dobra procjena pogreške (engl. *error estimation*). To je ovdje bilo nemoguće jer je zbog potpune neobrađenosti područja bilo potrebno provesti vrlo velik broj izračuna. U radu su ustanovljene temeljne smjernice, a preciznije pokazatelje bi trebalo tek napraviti.

Budući da su algoritmi za klasifikaciju, zbog opsežnosti istraživanja, korišteni sa zadanim postavkama, potrebno je detaljnije ispitati mogućnosti boljeg podešavanja parametara algoritama, kao i korištenja algoritama koji nisu upotrijebljeni u ovom radu.

U svrhu praktične primjene, potrebno je usavršiti i što više automatizirati proces prethodne obrade podataka i učiniti ga što prihvatljivijim i jednostavnijim za korisnika.



## **7. ETIČKI ASPEKTI ISTRAŽIVANJA**

Za potrebe ovog istraživanja etičkim povjerenstvima Fakulteta organizacije i informatike i Medicinskog fakulteta Sveučilišta u Zagrebu upućene su zamolbe za korištenjem podataka o dnevničkim zapisima studenata i njihovim ocjenama. Autorica ovog rada je pri tom dala izjavu da će osobni podaci biti obrađeni na način koji neće omogućavati identifikaciju osobe na koju se osobni podaci odnose, a u skladu sa Zakonom o zaštiti osobnih podataka (NN103/2003), članak 11. Navedena etička povjerenstva su izdala mišljenja da je istraživanje etički prihvatljivo.

## **8. ZAKLJUČAK**

U ovom istraživanju ispitane su osnovne postavke modela: je li predloženi model uopće moguće napraviti te postiže li točnost predikcije zbog koje bi vrijedilo dalje raditi na usavršavanju modela u smislu razvijanja tehnike sažimanja vremenskih nizova i u smislu praktične primjene tj., da je dostupan rezultat predikcije prije kraja nastave, kada još ima vremena za pomoć studentima ili za korekcije u nastavnom procesu. Pokazano je da je model predikcije uspjeha studenata dubinskom analizom podataka iz hibridnog okruženja za učenje, a primjenom analize vremenskih nizova pristupa studenata LMS-u u frekvencijskoj domeni, moguć. Djelomično je potvrđeno da predložena tehnika sažimanja koeficijenata daje bolju točnost i olakšava izračun. Djelomično je potvrđeno je da je moguća predikcija s točnošću od barem 75% upotrebom podataka o prve 2/3 trajanja predmeta što modelu može dati praktičnu upotrebljivost. To su bili potrebni koraci kako bi se u budućim istraživanjima moglo pristupiti proučavanju utjecaja pojedinih Fourierovih koeficijenata i njihovom povezivanju s ritmom kolegija, kako bi se mogli uspoređivati rezultati modeliranja na različitim tipovima kolegija te kako bi se moglo pristupiti traženju najprikladnijeg algoritma i podešavanju postavki algoritma. Ovdje predloženi model predikcije uspjeha studenata dubinskom analizom podataka iz hibridnog okruženja za učenje primjenom analize vremenskih nizova pristupa studenata LMS-u u frekvencijskoj domeni pruža dobar temelj za navedena buduća istraživanja, daje pretpostavke za nov uvid i bolje razumijevanje procesa podučavanja u hibridnom okruženju te slijedno tome bolje planiranje i odabir intervencija.

## 9. SMJERNICE ZA BUDUĆA ISTRAŽIVANJA

U budućim istraživanjima trebalo bi ispitati primjenu shema odlučivanja pri korištenju nekoliko klasifikatora kako bi se na najbolji način integrirali rezultati dobiveni pojedinim algoritmima. U literaturi navedeni primjeri (Lykourantzou i sur., 2009; Minaei-Bidgoli i sur., 2003) ukazuju na mogućnost da se time dodatno poboljšaju rezultati predikcije.

U ovom radu su algoritmi za klasifikaciju, zbog opsežnosti istraživanja, korišteni sa zadanim postavkama. To ostavlja prostor za daljnja poboljšanja u smislu boljeg podešavanja parametara algoritama, kao i korištenja algoritama koji nisu upotrijebljeni u ovom radu.

Prije navedena ograničenja rada istovremeno daju i smjer budućim istraživanjima, pa je tako, osim točnosti, za vrednovanje klasifikatora potrebno ispitati i druge mjere (preciznost, odziv, specifičnost).

Također, potrebno je sustavno provesti unakrsno vrednovanje s 5 ili 10 preklapanja. To je ovdje bilo nemoguće jer je zbog potpune neobrađenosti područja bilo potrebno provesti vrlo velik broj izračuna. U radu su ustanovljene temeljne smjernice, a preciznije pokazatelje bi trebalo tek napraviti.

Potrebno je i sustavno ispitati odnos potrebne gustoće podataka (ukupan broj klikova na predmetu podijeljen s umnoškom broja studenata i brojem dana trajanja predmeta) i točnosti predikcije.

Također je potrebno ispitati utjecaj strukture klikova po vrstama nastavnih materijala na točnost predikcije. U analizu dnevnčkih zapisa u ovom radu uključeni su svi zapisi o aktivnostima studenata. Unutar tog skupa nalaze se i tzv. neproduktivni zapisi, a to su oni koji se odnose na prolazak kroz izbornike u LMS-u (engl. *course-view*). Ti zapisi ne pridonose podizanju znanja, kompetencija, stava i/ili vještina. Njihov udio u ukupnom broju zapisa u akademskoj godini 2012./2013. iznosi po predmetima od 34-60%, a slično je i u akademskoj godini 2013./2014. U nekom novom istraživanju bi trebalo ispitati kakve bi bile procjene kada bi se analizirali dnevnički zapisi bez tzv. neproduktivnih zapisa.

Jedna od važnijih smjernica je usavršavanje i automatiziranje ovdje opisanog procesa prethodne obrade podataka. To je posebice značajno za praktičnu primjenu opisanog procesa. Potrebno ga je učiniti što jednostavnijim i prihvatljivijim za budućeg korisnika.

Zbog praktične primjene potrebno je istražiti mogućnost predikcije korištenjem povijesnih modela za predmet u aktualnoj akademskoj godini. Preliminarno su napravljeni izračuni tako da su obrađeni podaci o klikovima studenata u jednoj godini ubačeni u model generiran temeljem podataka o istom predmetu iz prethodne godine. Dobivena točnost predikcije je manja za 5-10 postotnih točaka od točnosti predikcije temeljem podataka za istu godinu.

Najvažniji smjer za budućnost istraživanja je rad na povezivanju najznačajnijih koeficijenata s ritmom kolegija. Rezultati tog istraživanja omogućili bi uvid u proces podučavanja u hibridnom okruženju na potpuno nov i neistražen način pružajući mogućnosti za nove i kvalitativne pomake.

## LITERATURA:

1. Abdi, H., L. J. Williams (2010) Principal component analysis, *WIREs Computational Statistics*, **2**, 433–459.
2. Agrawal, R., C. Faloutsos, A. Swami (1993) Efficient similarity search in sequence databases, in *Proceedings of the 4th FODO*, LNCS 730, p. 69-84.
3. Altman, N.S. (1992) An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression, *The American Statistician*, **46(3)**, 175-185.
4. Amoroso, D.L. (2005) Use of Online Assessment Tools to Enhance Student Performance in Large Classes, *Information Systems Education Journal*, **3 (4)**, 1-10.
5. Bagnall, A., G. Janacek (2005) Clustering time series with clipped data, *Machine Learning*, **58(2–3)**, 151–178.
6. Baker, R.S.J.D., K. Yacef (2009) The State of Educational Data Mining in 2009: A Review and Future Visions, *Journal of Educational Data Mining*, **1 (1)**, 3-17.
7. BlackBoard (2017), dostupno na <http://www.blackboard.com/>, zadnji pristup 21.6.2017.
8. Blikstein, P., M. Worsley (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks, *Journal of Learning Analytics*, **3(2)**, 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>
9. Bubaš, G., D. Kermek (2004) The Prospects for Blended Learning in Croatian Academic Institutions, in *Proceedings of the 6<sup>th</sup> CARNet Users Conference CUC 2004*, Zagreb, Croatia.
10. Chen, M.Y., B.T. Chen (2014) Online fuzzy time series analysis based on entropy discretization and a Fast Fourier Transform, *Applied Soft Computing*, **14(B)**, 156-166.
11. Chen, C.M., Y.L. Hsieh, S.H. Hsu (2007) Mining learner profile utilizing association rule for web-based learning diagnosis, *Expert Systems with Applications*, **33**, 6-22.
12. Chetchumlong, S. (2010) The effect of web-based formative assessment on students' achievement and attitudes in a large enrolment tertiary English course [doktorska disertacija], *University of Southern Queensland Faculty of Education*, Australia, dostupno na <http://eprints.usq.edu.au/8928/>, zadnji pristup 29.11.2016.
13. Cobo, G., D. García –Solórzano, E. Santamaría, J.A. Morán, J. Melenchón, C. Monzo (2011) Modeling students activity in online discussion forums: a strategy based on time series and agglomerative hierarchical clustering, in *Proceedings of the fourth*

- international conference on educational data mining*, Eindhoven, The Netherlands, 253-258.
14. Cortes, C.A. (1995) Support Vector Networks, *Machine Learning*, **45**, 5-32.
  15. Cruz, E., H. Dias, G. Kortemeyer (2011) The effect of formative assessment in Brazilian university physics courses, *Revista Brasileira de Ensino de Fisica*, **33(4)**, 1-7.
  16. Čehovin, L., Z. Bosnić (2010) Empirical evaluation of feature selection methods in classification, *Intelligent Data Analysis*, **14**, 265-281.
  17. Daniel, W.W., C.L. Cross (2013) *Biostatistics: A foundation for analysis in the health sciences*, 10th edition, John Wiley & Sons, Inc. USA.
  18. Dawson, S., D. Gašević, N. Mirriahi (2015) Challenging Assumptions in Learning Analytics, *Journal of Learning Analytics*, **2(3)**, 1-3.  
<http://dx.doi.org/10.18608/jla.2015.23.1>
  19. Dekker, W.G., M. Pechenizky, M.J. Vleeshouwers (2009) Predicting Students Drop Out: A case study, in *Proceedings of the 2nd International Conference on Educational Data Mining*, 2009, Cordoba, Spain.
  20. Delgado, M., E. Gibaja, M.C. Pegalajar, O. Pérez (2006) Predicting Students' Marks from Moodle Logs using Neural Network Models, in *Proc. International Conference on Current Developments in Technology Assisted Education*, Sevilla, Spain, 586-590.
  21. Demšar, J. (2006) Statistical comparisons of classifiers over multiple data sets, *Journal of Machine Learning Research*, **7**, 1-30.
  22. Deneui, D.L., T.L. Dodge (2006) Asynchronous Learning Networks and Student Outcomes: The Utility of Online Learning Components in Hybrid Courses, *Journal of Instructional Psychology*, **33(4)**, 256-259.
  23. Dias, S.B., J.A. Diniz (2013) FuzzyQoI model: A fuzzy logic-based modelling of users' quality of interaction with a learning management system under blended learning, *Computers & Education*, **69**, 38-59.
  24. Divjak, B., M. Maretić (2015) Learning analytics for e-Assessment: The State of the Art and One Case Study, In *Proceedings of the Central European Conference on Information and Intelligent Systems (CECIIS)*, Varaždin, Croatia.
  25. Divjak, B., D. Oreški (2009) Prediction of Academic Performance Using Discriminant Analysis, in *Proceedings of the 31st Int. Conf. on Information Technology Interfaces (ITI*

- 2009) Cavtat, Croatia.
26. Dobashi, K. (2015 ) Time series analysis of the in class page view history of digital teaching materials using cross table, In Proceedings of the 19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, *Procedia Computer Science*, **60**, 1032 – 1040, doi:10.1016/j.procs.2015.08.148.
  27. Dorça, F.A., L.V. Lima, M.A. Fernandes, C. R. Lopes (2013) Comparing strategies for modeling students learning styles through reinforcement learning in adaptive and intelligent educational systems: An experimental analysis, *Expert Systems with Applications*, **40**, 2092–2101.
  28. Dráždilová, P., G. Obadi, K. Slaninová, S. Al-Dubae, J. Martinovič, V. Snášel (2010) Computational Intelligence Methods for Data Analysis and Mining of eLearning Activities, In: *Computational Intelligence for Tech. Enhanced Learning*, Springer Berlin Heidelberg, **SCI 273**, 195–224.
  29. Duda, R.O., P.E. Hart, D.G. Stork (2000) Pattern classification, Wiley Interscience.
  30. Falakmasir, M.H., Z.A. Prados, G.J. Goerdon, P. Brusilovsky (2013) A spectral learning approach to knowledge tracing, in *Proceedings of the 6<sup>th</sup> International Conference on Educational Data Mining (EDM 2013)*, Memphis, TN, USA, 28-34.
  31. Fawcett, T. (2006) An introduction to ROC analysis, *Pattern Recognition Letters*, **27**, 861–874.
  32. Feldman, J., A. Monteserin, A. Amandi (2014) Detecting students' perception style by using games, *Computers and Education*, **71**, 14-22.
  33. Free Pascal (2017), dostupno na <<https://www.freepascal.org/>>, zadnji pristup 21.6.2017.
  34. Fu, T.C. (2011) A review on time series data mining, *Engineering Applications of Artificial Intelligence*, **24 (1)**, 164–181.
  35. Gamulin, J., O. Gamulin, D. Kermek (2016) Using Fourier coefficients in time series analysis for student performance prediction in blended learning environments, *Expert Systems*, 33-2, 176-188, DOI: 10.1111/exsy.12142.
  36. Gamulin, J., O. Gamulin, D. Kermek (2014) Comparing classification models in the final exam performance prediction, in *Proceedings of the 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2014/CE)*, Opatija, Croatia, 781-786.

37. Gamulin, J., O. Gamulin, D. Kermek (2013) Data mining in hybrid learning: possibility to predict the final exam result, in *Proceedings of the 36th International Convention, Conference Computers in Education (MIPRO 2013/CE)*, Opatija, Croatia, IEEE Conference Publications, 591 - 596.
38. Garrison, D. R., H. Kanuka (2004) Blended learning: Uncovering its transformative potential in higher education, *The Internet and Higher Education*, **7(2)**, 95–105, doi:10.1016/j.iheduc.2004.02.001
39. Gasevic, D., G. Conole, G. Siemens, P. Long (Eds). (2011) Learning analytics and educational data mining, in *Proc. LAK11: International Conference on Learning Analytics and Knowledge*, Banff, Canada, 27 February - 1 March 2011.
40. Gupta, P., A. Sharma, R. Jindal (2016) Scalable machine-learning algorithms for big data analytics: a comprehensive review, *WIREs Data Mining Knowl Discov*, **6**, 194–214. doi: 10.1002/widm.1194.
41. Hämmäläinen, W., M. Vinni (2011) *Cassifiers for educational data mining*, London, Chapman & Hall, CRC.
42. Hastie, T., R. Tibshirani, J. Friedman (2008) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, second edition.
43. Herzog, S. (2006) Estimating student retention and degree-completion time: Decision trees and neural networks vis-`a-vis regression, *New Directions For Institutional Research*, **2006(131)**, 17–33, doi: 10.1002/ir
44. Jadrić, M., Z. Garača, M. Čukušić (2010) Student dropout analysis with application of data mining methods, *Management*, **15**, 31-46.
45. Jain, A.K., M.N. Murty , P.J. Flynn (1999) Data clustering: a review, *ACM Computing Surveys* **31 (3)**, 264-323.
46. Jovanović, M., M. Vukičević, M. Milovanović, M. Minović (2012) Using data mining on student behavior and cognitive style data for improving e-learning systems: a case study, *International Journal of Computational Intelligence Systems*, **5(3)**, 597-610.
47. Khan, T. M., Clear, F., Sajadi, S. S. (2012). The relationship between educational performance and online access routines: analysis of students' access to an online discussion forum . In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* . ACM, 226-229. Doi:10.1145/2330601.2330655

48. Kliček, B., M. Zekić Sušac (2003) Toward Integrated and Revised Learning Styles Theory Supported by Web and Multimedia Technologies, in *Proceedings of the 8th Annual ELSIN Conference Hull*, England , University of Hull, 391-410.
49. Kohavi, R. (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection, *International Joint Conference on Artificial Intelligence*, 14(2) 1137-1145.
50. Kotsiantis, S., C. Pierrakeas, P. Pintelas (2004) Predicting students' performance in distance learning using machine learning techniques, *Applied Artificial Intelligence*, **18**, 411-426.
51. Kotsiantis, S.B., I.D. Zaharakis. P. E. Pintelas (2006) Machine learning: a review of classification and combining techniques, *Artif Intelligence Review*, **26(3)**, 159–190, DOI 10.1007/s10462-007-9052-3
52. Kotsiantis, S., K. Patriarcheas, M. Xenos (2010) A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education, *Knowledge-Based Systems*, **23(6)**, 529- 535.
53. Kotu, V., B. Deshpande (2014) *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*, Elsevier Science & Technology/ Morgan Kaufmann.
54. Kovacic, Z.J. (2012) Predicting student success by mining enrolment data, *Research in Higher Education Journal*, **15**, 1-20.
55. Kovačić, Z.J. (1995) Analiza vremenskih serija, Univerzitet u Beogradu Ekonomski fakultet.
56. Laerd Statistics, dostupno na <https://statistics.laerd.com/spss-tutorials/friedman-test-using-spss-statistics.php>, zadnji pristup 21.8.2017.
57. Lara, J.A., D. Lizcano, M.A. Martínez, J. Pazos, T. Riera (2014) A system for knowledge discovery in e-learning environments within the European Higher Education Area – Application to student data from Open University of Madrid, UDIMA, *Computers & Education*, **72**, 23–36.
58. Larose, D.T. (2014) *Discovering knowledge in data: an introduction to data mining*, Hoboken: John Wiley & Sons, Inc.
59. Larose, D.T., C. D. Larose (2015) *Data Mining and Predictive Analytics*, 2nd edition, John Wiley & Sons, Inc. USA, ISBN 978-1-118-11619-7.



60. Laurillard, D. (2004) *E-Learning in Higher Education, From Changing Higher Education*, (ed. Ashwin P.), Routledge Falmer, London, UK.
61. Laurillard, D., M. Oliver, B. Wasson, U. Hoppe (2009) Implementing technology-enhanced learning, In *Technology-enhanced learning* (pp. 289-306) Springer Netherlands
62. Lazarus (2017), dostupno na < <http://www.lazarus-ide.org/>>, zadnji pristup 29.11.2016.
63. Lon-Capa (2017), dostupno na < <http://www.lon-capa.org/>>, zadnji pristup 25.10.2016.
64. Luan, J. (2006) Data mining and its applications in higher education., *New Directions For Institutional Research*, **2002/113**, 17–36.
65. Lykourantzou, I., I. Giannoukos, V. Nikolopoulos, G. Mparadis, V. Loumos (2009) Dropout prediction in e-learning courses through the combination of machine learning techniques, *Computers and Education*, **53 (3)**, 950–965.
66. Matlab (2017), dostupno na <http://www.mathworks.com/>, zadnji pristup 21.6.2017.
67. Minaei-Bidgoli, B., D. A. Kashy, G. Kortmeyer, W. F. Punch (2003) Predicting student performance: an application of data mining methods with an educational Web-based system, in *Proc. 33rd International Conference on Frontiers in Education*, (Colorado, Westminster, 13-18.
68. Moodle (2017) < <https://moodle.org/>>, zadnji pristup 21.6.2017.
69. MySQL (2017) < <https://www.mysql.com/>>, zadnji pristup 21.6.2017.
70. Nasiri, M., B. Minaei, F. Vafaei (2012) Predicting GPA and academic dismissal in LMS using educational data mining: a case mining, in *Proc. 6th National and 3rd International conference of e-Learning and e-Teaching (ICELET 2012)*, 53-58.
71. Nemenyi, P.B. (1963) Distribution-free Multiple Comparisons [doktorska disertacija], *Princeton University*
72. Ocepek, U., Z. Bosnić, I. Nančovska Šerbec, J. Rugelj (2013) Exploring the relation between learning style models and preferred multimedia types, *Computers & Education*, **69**, 343–355.
73. Oreški, S., D. Oreški, G. Oreški (2012) Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment, *Expert Systems with Applications*, **39**, 12605-12617.
74. Paliwal, M., U.A. Kumar (2009) A study of academic performance of business school graduates using neural network and statistical techniques, *Expert Systems with*

- Applications*, **36(4)**, 7865–7872.
75. Powers, D. M. W. (2011) Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation, *Journal of Machine Learning Technologies*. **2** (1), 37-63.
  76. RapidMiner (2017), dostupno na <<http://rapidminer.com/>>, zadnji pristup 21.6.2017
  77. Reis, M.S., P.M. Saravia, B.R. Bakshi (2009) Denoising and Signal-to-Noise Ratio Enhancement: Wavelet Transform and Fourier Transform, *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, **2**, 25-55.
  78. Rish, I. (2001) An empirical study of the naive Bayes classifier, IBM Research Report.
  79. Romero, C., S. Ventura (2017) Educational data science in massive open online courses, *WIRES Data Mining Knowl Discov*, **7**, e1187, doi: 10.1002/widm.1187.
  80. Romero, C., M.I. López, J.M. Luna, S. Ventura, (2013) Predicting students' final performance from participation in on-line discussion forums, *Computers & Education*, **68(C)**, 458–472, doi:10.1016/j.compedu.2013.06.009
  81. Romero, C., S. Ventura (2010) Educational data mining: a review of the state-of-the-art, *IEEE Trans. Syst. Man Cybernet, C Appl. Rev.*, **40(6)**, 601–618.
  82. Romero, C., S. Ventura, P.G. Espejo, C. Hervas (2008) Data Mining Algorithms to Classify Students, in *Proc. 1st International conference on Educational Data Mining (EDM 2008)*, Montreal , Canada, p 8-17.
  83. Romero, C., S. Ventura, E. García (2008) Data mining in course management systems: moodle case study and tutorial, *Computers & Education*, **51(1)**, 368–384.
  84. Romero, C., Ventura, S. (2007) Educational Data Mining: A Survey from 1995 to 2005. *Expert Systems with Applications*, **33**, 125-146.
  85. Siemens, G., R.S.J.D. Baker (2010) Learning Analytics and Educational Data Mining: Towards Communication and Collaboration, in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, Publisher: ACM
  86. Sison, R., Shimura, M. (1998) Student modeling and machine learning, *International Journal of Artificial Intelligence in Education*, **9(1-2)**, 128-158.
  87. Statistics Solutions, dostupno na <http://www.statisticssolutions.com/assumptions-of-the-wilcox-sign-test/>, zadnji pristup 21.8.2017.
  88. Sveučilište u Splitu, Strategija Sveučilišta u Splitu 2015-2020, (strateški cilj 2, str. 14)

dostupno

na

[http://www.unist.hr/Portals/0/datoteke/dokumenti/UNIST\\_STRATEGIJA\\_2015\\_2020.pdf](http://www.unist.hr/Portals/0/datoteke/dokumenti/UNIST_STRATEGIJA_2015_2020.pdf)  
zadnji pristup, 21.6.2017.

89. Sveučilište u Zagrebu, Strategija studija i studiranja Sveuadmin/rektorat/Studiji\_studira, (dio vezan uz e-ui studiranja Sveuadmin) dostupno na [http://www.unizg.hr/fileadmin/rektorat/Studiji\\_studiranje/Studiji/e-ucenje/Strategija\\_Studiji\\_i\\_studiranje\\_08052014.pdf](http://www.unizg.hr/fileadmin/rektorat/Studiji_studiranje/Studiji/e-ucenje/Strategija_Studiji_i_studiranje_08052014.pdf), zadnji pristup 21.6.2017.
90. Tan, P.-N., S. Michael, V. Kumar (2005) Classification and Classification: Alternative Techniques. In Tan, P.-N., Michael, S., Kumar, V. (Eds) Introduction to Data Mining, pp. 145-315, Boston, MA: Addison-Wesley.
91. TeamViewer (2017) dostupno na < <https://www.teamviewer.com/>>, zadnji pristup 21.6.2017.
92. Terzis, V., A.A. Economides (2010) The acceptance and use of computer based assessment, *Computers & Education*, **56**, 1032–1044.
93. Thai-Nghe, N., P. Janecek, P. Haddawy (2007) A comparative analysis of techniques for predicting academic performance, in *Proceedings of 37th IEEE Frontiers in Education Conference (FIE'07)*, Milwaukee, USA, IEEE Xplore, T2G7–T2G12.
94. Thai-Nghe, N., A. Busche, L. Schmidt-Thieme (2009) Improving Academic Performance Prediction by Dealing with Class Imbalance, in: *Proceeding of 9th IEEE International Conference on Intelligent Systems Design and Applications (ISDA'09)*, Pisa, Italy, IEEE Computer Society, p. 878–883, 2009.
95. Thai-Nghe, N., L. Drumond, A. Krohn-Grimberghe, L. Schmidt-Thieme (2010) Recommender System for Predicting Student Performance, *Procedia Computer Science*, **1**, 2811–2819.
96. Velić, M. (2014) Metoda umjetne inteligencije za rudarenje podacima iz vremenskih serija i implementacija na ljudskom EKG signal [doktorska disertacija], *Fakultet organizacije i informatike Sveučilišta u Zagrebu*, Varaždin.
97. Vlachos, M., C. Meet, Z. Vagena (2004) Identifying similarities, periodicities and bursts for online search queries, in *Proceedings of the 23rd ACM SIGMOD International Conference on Management of Data*, Paris, France.
98. Vranić, M. (2009) Dubinska analiza podataka–određeni postupci u području deskriptivne

analize, *Fakultet elektrotehnike i računarstva Sveučilišta u Zagrebu*, Zagreb.

99. Warnakulasooriya, R., W. Galen (2012) Categorizing Students' Response Patterns using concept of Fractal Dimension, in *Proceedings of the 5th International Conference on Educational Data Mining*, Chania, Greece, 214-215.
100. Wilcoxon, F. (Dec 1945) Individual comparisons by ranking methods, *Biometrics Bulletin*, **1 (6)**, 80–83.
101. Wilson, S. J. (2017) Data representation for time series data mining: time domain approaches, *WIRES Comput Stat*, **9**, e1392. doi: 10.1002/wics.1392.
102. Wilson, K., C. Boyd, L. Chen, S. Jamal (2011) Improving student performance in a first-year geography course: Examining the importance of computer-assisted formative assessment, *Computers & Education*, **57**, 1493–1500.
103. Xie, C., Z. Zhang, S. Nourian, A. Pallant, E. Hazzard (2014) Time Series Analysis Method for Assessing Engineering Design Processes Using a CAD Tool, *International Journal of Engineering Education*, **30(1)**, 218–230.
104. Yang, Q., X. Wu (2006) 10 Challenging problems in data mining research, *International Journal of Information Technology and Decision Making*, **5 (4)**, 597–604.
105. Zafra, A., Romero, C., Ventura, S. (2011) Multiple instance learning for classifying students in learning management Systems, *Expert Systems with Applications*, **38(12)**, 15020–15031.

## PRILOZI

*Tablica 52: Točnosti - ANN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	79,57	82,80	83,87	83,90	81,72	76,34	83,80	80,65
FIB 2K <sub>sat</sub> 12	82,61	81,52	84,78	82,61	81,50	81,50	83,70	83,70
NWTIS 2K <sub>sat</sub> 13	93,35	93,30	93,45	93,33	93,32	93,25	93,25	93,33
NWTIS 2K <sub>sat</sub> 12	84,21	89,47	89,47	94,74	89,47	89,47	89,47	84,21
WDIP 2K <sub>sat</sub> 13	92,15	88,54	92,25	96,05	92,55	88,17	88,17	92,15
WDIP 2K <sub>sat</sub> 12	87,50	87,50	91,67	91,67	91,67	91,67	95,83	91,67

*Tablica 53: Minimalan broj koeficijenata za maksimalnu točnost - ANN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	25	41	92	63	14	133	7	9
FIB 2K <sub>sat</sub> 12	6	9	18	42	4	6	78	33
NWTIS 2K <sub>sat</sub> 13	7	4	4	5	5	4	4	6
NWTIS 2K <sub>sat</sub> 12	14	19	14	17	36	13	5	36
WDIP 2K <sub>sat</sub> 13	11	8	56	283	5	11	11	4
WDIP 2K <sub>sat</sub> 12	13	28	15	24	30	58	55	34

*Tablica 54: Točnosti - ANN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova – puni broj koeficijenata*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	68,82	70,97	70,97	70,97	68,82	76,34	72,04	73,12
FIB 2K <sub>sat</sub> 12	71,74	73,91	71,74	75,00	73,91	81,52	75,00	79,35
NWTIS 2K <sub>sat</sub> 13	93,33	93,33	100,00	93,33	86,67	93,33	86,67	100,00
NWTIS 2K <sub>sat</sub> 12	73,68	78,95	73,68	78,95	68,42	78,95	84,21	73,68
WDIP 2K <sub>sat</sub> 13	80,00	80,00	80,00	80,00	76,00	67,35	88,00	80,00
WDIP 2K <sub>sat</sub> 12	70,83	70,83	79,17	83,33	66,67	79,17	79,17	79,17

Tablica 55: Puni broj koeficijenata - ANN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	349	349	349	349	349	349	349	349
FIB 2K <sub>sat</sub> 12	349	349	349	349	349	349	349	349
NWTIS 2K <sub>sat</sub> 13	252	253	253	283	281	283	313	313
NWTIS 2K <sub>sat</sub> 12	252	253	253	283	281	283	313	313
WDIP 2K <sub>sat</sub> 13	385	384	403	403	403	403	403	403
WDIP 2K <sub>sat</sub> 12	313	313	313	343	343	343	163	343

Tablica 56: Točnosti - ANN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 2K <sub>dan</sub> 13	82,80	81,72	82,08	81,72	82,91	83,87
FIB 2K <sub>dan</sub> 12	80,44	83,70	84,48	82,61	82,61	84,78
NWTIS 2K <sub>dan</sub> 13	93,33	93,33	93,33	93,33	93,33	93,33
NWTIS 2K <sub>dan</sub> 12	92,14	92,23	97,35	92,55	94,47	98,13
WDIP 2K <sub>dan</sub> 13	88,78	92,05	92,10	88,58	88,15	88,25
WDIP 2K <sub>dan</sub> 12	83,33	87,54	87,15	91,67	91,50	n/a

Tablica 57: Minimalan broj koeficijenata za maksimalnu točnost - ANN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 3K <sub>dan</sub> 13	5	7	8	11	9	14
FIB 3K <sub>dan</sub> 12	4	12	6	4	18	12
NWTIS 3K <sub>dan</sub> 13	17	4	9	9	43	44
NWTIS 3K <sub>dan</sub> 12	24	7	8	17	25	13
WDIP 3K <sub>dan</sub> 13	20	10	4	4	22	n/a
WDIP 3K <sub>dan</sub> 12	18	23	32	22	25	n/a

Tablica 58: Točnosti - ANN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - puni broj koeficijenata

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 2K <sub>dan</sub> 13	74,19	74,19	76,34	76,34	76,34	77,42
FIB 2K <sub>dan</sub> 12	77,17	79,35	77,17	77,17	75,00	77,17
NWTIS 2K <sub>dan</sub> 13	86,67	80,00	93,33	93,33	93,33	86,67
NWTIS 2K <sub>dan</sub> 12	84,21	84,21	78,95	78,95	78,95	78,95
WDIP 2K <sub>dan</sub> 13	80,00	80,00	80,00	84,00	80,00	80,00
WDIP 2K <sub>dan</sub> 12	75,00	75,00	75,00	n/a	n/a	79,17

Tablica 59: Puni broj koeficijenata - ANN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 2K <sub>dan</sub> 13	27	38	48	58	73	90
FIB 2K <sub>dan</sub> 12	28	27	47	58	74	93
NWTIS 2K <sub>dan</sub> 13	29	39	49	59	74	100
NWTIS 2K <sub>dan</sub> 12	29	38	48	58	73	86
WDIP 2K <sub>dan</sub> 13	29	39	49	59	74	100
WDIP 2K <sub>dan</sub> 12	29	39	59	n/a	n/a	88

Tablica 60: Točnosti - ANN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 3K <sub>sat</sub> 13	63,44	62,37	65,59	65,59	65,59	65,59	70,97	66,67
FIB 3K <sub>sat</sub> 12	65,22	65,22	67,39	68,48	67,39	71,74	77,83	79,35
NWTIS 3K <sub>sat</sub> 13	93,33	93,33	93,33	86,67	86,67	86,67	86,67	86,67
NWTIS 3K <sub>sat</sub> 12	84,41	78,95	78,95	78,95	84,21	78,95	84,21	84,21
WDIP 3K <sub>sat</sub> 13	76,11	72,41	76,21	80,23	68,41	72,36	76,05	68,15
WDIP 3K <sub>sat</sub> 12	79,17	83,33	83,45	79,20	84,22	75,22	75,26	84,21

*Tablica 61: Minimalan broj koeficijenata za maksimalnu točnost - ANN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 3K <sub>sat</sub> 13	7	13	6	21	14	133	9	9
FIB 3K <sub>sat</sub> 12	41	45	8	14	21	7	11	149
NWTIS 3K <sub>sat</sub> 13	39	8	32	30	12	8	14	6
NWTIS 3K <sub>sat</sub> 12	5	19	40	5	4	5	67	86
WDIP 3K <sub>sat</sub> 13	98	24	20	83	43	43	98	10
WDIP 3K <sub>sat</sub> 12	5	7	4	23	11	7	9	36

*Tablica 62: Točnosti - ANN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 3K <sub>sat</sub> 13	53,76	54,84	64,52	54,84	56,99	61,29	58,06	55,91
FIB 3K <sub>sat</sub> 12	53,26	54,35	54,35	54,35	55,43	51,09	56,52	76,09
NWTIS 3K <sub>sat</sub> 13	80,00	80,00	86,67	80,00	86,67	80,00	80,00	80,00
NWTIS 3K <sub>sat</sub> 12	68,42	57,89	63,16	57,89	52,63	63,16	68,42	68,42
WDIP 3K <sub>sat</sub> 13	56,00	55,88	60,00	60,00	60,00	56,00	68,00	60,00
WDIP 3K <sub>sat</sub> 12	66,67	66,67	66,67	75,00	56,00	62,50	66,67	79,17

*Tablica 63: Puni broj koeficijenata - ANN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 3K <sub>sat</sub> 13	349	349	349	349	349	349	349	349
FIB 3K <sub>sat</sub> 12	349	349	349	349	349	349	349	349
NWTIS 3K <sub>sat</sub> 13	252	253	253	283	281	283	313	313
NWTIS 3K <sub>sat</sub> 12	252	253	253	283	281	283	313	313
WDIP 3K <sub>sat</sub> 13	385	384	403	403	403	403	403	403
WDIP 3K <sub>sat</sub> 12	313	313	313	343	343	343	163	343



Tablica 64: Točnosti - ANN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 3K <sub>dan</sub> 13	63,44	63,55	67,74	68,82	67,74	67,74
FIB 3K <sub>dan</sub> 12	66,30	69,57	67,39	68,84	68,48	70,65
NWTIS 3K <sub>dan</sub> 13	93,34	86,67	86,67	86,67	93,33	93,33
NWTIS 3K <sub>dan</sub> 12	78,95	78,95	78,95	84,71	84,21	84,21
WDIP 3K <sub>dan</sub> 13	72,15	72,18	72,15	68,25	68,25	n/a
WDIP 3K <sub>dan</sub> 12	79,17	83,33	79,17	87,51	76,21	n/a

Tablica 65: Minimalan broj koeficijenata za maksimalnu točnost - ANN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 3K <sub>dan</sub> 13	5	7	8	11	9	14
FIB 3K <sub>dan</sub> 12	4	12	6	4	18	12
NWTIS 3K <sub>dan</sub> 13	17	4	9	9	43	44
NWTIS 3K <sub>dan</sub> 12	24	7	8	17	25	13
WDIP 3K <sub>dan</sub> 13	20	10	4	4	22	n/a
WDIP 3K <sub>dan</sub> 12	18	23	32	22	25	n/a

Tablica 66: Točnosti - ANN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - puni broj koeficijenata

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 3K <sub>dan</sub> 13	60,22	52,69	55,91	60,22	60,22	60,22
FIB 3K <sub>dan</sub> 12	61,96	58,70	59,78	58,70	57,61	59,78
NWTIS 3K <sub>dan</sub> 13	86,67	66,67	66,67	80,00	80,00	73,33
NWTIS 3K <sub>dan</sub> 12	68,42	68,42	68,42	73,68	73,68	68,42
WDIP 3K <sub>dan</sub> 13	60,00	56,00	56,00	60,00	60,00	60,00
WDIP 3K <sub>dan</sub> 12	75,00	70,83	66,67	70,83	n/a	n/a

Tablica 67: Puni broj koeficijenata - ANN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 3K <sub>dan</sub> 13	29	39	49	59	72	84
FIB 3K <sub>dan</sub> 12	28	38	50	50	58	74
NWTIS 3K <sub>dan</sub> 13	29	39	49	59	75	100
NWTIS 3K <sub>dan</sub> 12	28	38	48	58	73	86
WDIP 3K <sub>dan</sub> 13	29	39	49	59	59	100
WDIP 3K <sub>dan</sub> 12	29	39	49	59	n/a	n/a

Tablica 68: Točnosti - ANN – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	83,95	86,42	85,19	83,95	83,95
FMD 2K <sub>sat</sub> 12	93,11	90,79	90,79	93,42	90,79	89,47

Tablica 69: Minimalan broj koeficijenata za maksimalnu točnost - ANN – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	4	10	5	4	22
FMD 2K <sub>sat</sub> 12	16	22	12	32	25	7

Tablica 70: Točnosti - ANN – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	79,01	79,01	81,48	80,25	81,48
FMD 2K <sub>sat</sub> 12	86,84	89,47	85,53	86,84	85,53	80,26

*Tablica 71: Puni broj koeficijenata - ANN – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova*

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	38	48	58	78	97
FMD 2K <sub>sat</sub> 12	28	38	48	58	77	93

*Tablica 72: Točnosti - SVM – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	84,95	87,10	84,95	80,43	81,52	87,10	86,02	86,02
FIB 2K <sub>sat</sub> 12	76,09	78,26	78,26	80,43	81,52	80,43	78,25	86,02
NWTIS 2K <sub>sat</sub> 13	93,25	93,33	93,45	93,33	93,65	93,15	93,70	93,35
NWTIS 2K <sub>sat</sub> 12	84,21	89,47	91,67	95,83	91,67	83,33	89,35	94,74
WDIP 2K <sub>sat</sub> 13	96,15	96,25	96,63	96,63	96,43	77,55	96,18	96,15
WDIP 2K <sub>sat</sub> 12	87,15	91,67	91,67	91,67	91,76	83,45	87,52	83,23

*Tablica 73: Minimalan broj koeficijenata za maksimalnu točnost - SVM – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	89	236	295	119	77	71	172	114
FIB 2K <sub>sat</sub> 12	322	272	242	314	176	213	144	274
NWTIS 2K <sub>sat</sub> 13	6	5	4	4	4	6	4	4
NWTIS 2K <sub>sat</sub> 12	52	16	22	23	48	10	6	27
WDIP 2K <sub>sat</sub> 13	37	11	58	223	68	28	139	117
WDIP 2K <sub>sat</sub> 12	28	5	13	19	48	10	7	5

Tablica 74: Točnosti - SVM – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	81,72	77,42	82,80	79,57	80,65	79,57	78,49	82,80
FIB 2K <sub>sat</sub> 12	75,00	73,91	76,09	75,00	75,00	75,00	75,00	75,00
NWTIS 2K <sub>sat</sub> 13	100,00	100,00	93,33	93,33	100,00	73,33	100,00	93,33
NWTIS 2K <sub>sat</sub> 12	68,42	63,16	70,83	66,67	70,83	66,67	68,42	73,68
WDIP 2K <sub>sat</sub> 13	84,00	88,00	84,00	92,00	88,00	61,22	88,00	84,00
WDIP 2K <sub>sat</sub> 12	62,50	54,17	70,83	66,67	70,83	66,67	62,50	66,67

Tablica 75: Puni broj koeficijenata - SVM – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	301	353	353	353	353	353	353	353
FIB 2K <sub>sat</sub> 12	353	353	353	353	353	353	353	353
NWTIS 2K <sub>sat</sub> 13	283	250	268	278	278	303	311	311
NWTIS 2K <sub>sat</sub> 12	323	258	315	348	353	353	323	323
WDIP 2K <sub>sat</sub> 13	353	353	353	353	353	603	353	353
WDIP 2K <sub>sat</sub> 12	311	323	313	348	353	353	353	353

Tablica 76: Točnosti - SVM – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 2K <sub>dan</sub> 13	81,72	81,65	82,80	84,95	83,67	83,57
FIB 2K <sub>dan</sub> 12	72,83	73,91	73,85	73,85	73,95	76,09
NWTIS 2K <sub>dan</sub> 13	93,15	93,25	93,33	93,24	93,53	93,33
NWTIS 2K <sub>dan</sub> 12	89,47	94,74	94,76	94,74	94,64	94,54
WDIP 2K <sub>dan</sub> 13	88,25	88,35	88,42	92,15	88,33	92,36
WDIP 2K <sub>dan</sub> 12	79,17	87,50	87,50	87,25	87,25	n/a

*Tablica 77: Minimalan broj koeficijenata za maksimalnu točnost - SVM – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 2K <sub>dan</sub> 13	11	6	18	55	11	4
FIB 2K <sub>dan</sub> 12	18	20	14	55	36	47
NWTIS 2K <sub>dan</sub> 13	5	4	5	5	4	4
NWTIS 2K <sub>dan</sub> 12	8	16	15	24	27	6
WDIP 2K <sub>dan</sub> 13	6	13	5	38	20	54
WDIP 2K <sub>dan</sub> 12	13	32	30	34	41	n/a

*Tablica 78: Točnosti - SVM – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - puni broj koeficijenata*

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 2K <sub>dan</sub> 13	79,57	77,42	77,42	81,72	84,95	81,72
FIB 2K <sub>dan</sub> 12	72,83	71,74	73,91	72,83	71,74	72,83
NWTIS 2K <sub>dan</sub> 13	93,33	100,00	93,33	93,33	93,33	93,33
NWTIS 2K <sub>dan</sub> 12	78,95	78,95	73,68	84,21	84,21	68,42
WDIP 2K <sub>dan</sub> 13	84,00	80,00	88,00	88,00	84,00	84,00
WDIP 2K <sub>dan</sub> 12	70,83	75,00	79,17	79,17	70,83	n/a

*Tablica 79: Puni broj koeficijenata - SVM – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova*

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 2K <sub>dan</sub> 13	28	38	48	58	73	88
FIB 2K <sub>dan</sub> 12	29	38	48	58	74	128
NWTIS 2K <sub>dan</sub> 13	28	38	48	58	73	143
NWTIS 2K <sub>dan</sub> 12	28	38	48	58	73	86
WDIP 2K <sub>dan</sub> 13	28	38	48	72	143	143
WDIP 2K <sub>dan</sub> 12	28	38	48	58	85	n/a

*Tablica 80: Točnosti - SVM – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	88,89	85,19	86,42	88,89	87,65
FMD 2K <sub>sat</sub> 12	77,63	78,95	78,95	80,26	78,95	81,58

*Tablica 81: Minimalan broj koeficijenata za maksimalnu točnost - SVM – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	27	11	7	10	6
FMD 2K <sub>sat</sub> 12	4	38	10	13	4	23

*Tablica 82: Točnosti - SVM – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata*

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	82,72	83,95	82,72	83,95	83,95
FMD 2K <sub>sat</sub> 12	77,63	77,63	77,63	77,63	77,63	77,63

*Tablica 83: Puni broj koeficijenata - SVM – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova*

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	38	48	58	77	183
FMD 2K <sub>sat</sub> 12	28	38	48	57	74	88

Tablica 84: Točnosti - kNN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	82,80	83,87	83,87	83,88	87,10	86,02	88,17	87,10
FIB 2K <sub>sat</sub> 12	82,61	84,78	84,79	88,04	84,78	86,96	84,78	85,87
NWTIS 2K <sub>sat</sub> 13	93,33	93,35	93,32	93,31	93,28	93,36	93,33	93,34
NWTIS 2K <sub>sat</sub> 12	89,47	89,48	93,33	94,74	94,75	94,74	94,77	94,74
WDIP 2K <sub>sat</sub> 13	92,01	96,03	96,02	92,01	92,03	83,67	92,03	96,02
WDIP 2K <sub>sat</sub> 12	91,67	91,62	95,83	91,67	87,50	91,67	87,50	91,67

Tablica 85: Minimalan broj koeficijenata za maksimalnu točnost - kNN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	76	271	15	18	109	26	146	176
FIB 2K <sub>sat</sub> 12	112	147	26	65	232	18	164	41
NWTIS 2K <sub>sat</sub> 13	4	4	4	5	11	10	10	13
NWTIS 2K <sub>sat</sub> 12	34	117	10	17	12	4	25	4
WDIP 2K <sub>sat</sub> 13	286	36	116	104	5	135	136	282
WDIP 2K <sub>sat</sub> 12	17	37	193	89	91	84	208	225

Tablica 86: Točnosti - kNN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	78,49	77,42	74,19	76,34	75,27	76,34	75,27	78,49
FIB 2K <sub>sat</sub> 12	76,09	82,61	79,35	75,00	80,43	76,09	79,35	79,35
NWTIS 2K <sub>sat</sub> 13	93,33	93,33	93,33	100,00	89,47	100,00	100,00	93,33
NWTIS 2K <sub>sat</sub> 12	78,95	78,95	86,67	84,21	89,47	94,74	89,47	84,21
WDIP 2K <sub>sat</sub> 13	84,00	88,00	84,00	80,00	84,00	75,51	80,00	80,00
WDIP 2K <sub>sat</sub> 12	79,17	75,00	83,33	83,33	79,17	83,33	79,17	83,33

Tablica 87: Puni broj koeficijenata - kNN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	353	353	353	353	353	353	353	353
FIB 2K <sub>sat</sub> 12	403	403	403	403	403	403	403	403
NWTIS 2K <sub>sat</sub> 13	286	251	270	283	283	303	309	309
NWTIS 2K <sub>sat</sub> 12	323	258	263	253	283	328	323	323
WDIP 2K <sub>sat</sub> 13	353	353	353	353	353	353	353	353
WDIP 2K <sub>sat</sub> 12	312	324	313	348	353	353	353	353

Tablica 88: Točnosti - kNN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 2K <sub>dan</sub> 13	81,72	83,87	86,02	84,95	87,10	86,02
FIB 2K <sub>dan</sub> 12	82,61	84,78	84,76	84,79	84,75	85,87
NWTIS 2K <sub>dan</sub> 13	93,33	93,36	93,35	93,32	93,30	93,32
NWTIS 2K <sub>dan</sub> 12	84,21	94,74	94,76	94,77	94,74	94,75
WDIP 2K <sub>dan</sub> 13	88,01	92,03	88,02	96,03	92,04	96,04
WDIP 2K <sub>dan</sub> 12	83,33	83,31	83,35	87,50	n/a	87,60

Tablica 89: Minimalan broj koeficijenata za maksimalnu točnost - kNN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 2K <sub>dan</sub> 13	20	13	17	16	5	78
FIB 2K <sub>dan</sub> 12	6	25	37	17	65	20
NWTIS 2K <sub>dan</sub> 13	4	4	4	4	20	4
NWTIS 2K <sub>dan</sub> 12	4	27	11	21	20	34
WDIP 2K <sub>dan</sub> 13	8	7	31	20	9	63
WDIP 2K <sub>dan</sub> 12	17	8	6	23		14



Tablica 90: Točnosti - kNN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - puni broj koeficijena

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 2K <sub>dan</sub> 13	79,57	82,80	82,80	78,49	83,87	83,87
FIB 2K <sub>dan</sub> 12	78,26	81,52	80,43	82,61	81,52	80,43
NWTIS 2K <sub>dan</sub> 13	93,33	100,00	100,00	100,00	84,21	93,33
NWTIS 2K <sub>dan</sub> 12	84,21	89,47	84,21	94,74	84,21	84,21
WDIP 2K <sub>dan</sub> 13	80,00	80,00	88,00	80,00	n/a	80,00
WDIP 2K <sub>dan</sub> 12	75,00	75,00	79,17	79,17	n/a	79,17

Tablica 91: Puni broj koeficijena - kNN – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 2K <sub>dan</sub> 13	28	35	48	60	71	88
FIB 2K <sub>dan</sub> 12	27	37	47	57	73	132
NWTIS 2K <sub>dan</sub> 13	28	38	48	58	73	145
NWTIS 2K <sub>dan</sub> 12	28	38	48	58	73	86
WDIP 2K <sub>dan</sub> 13	28	38	48	58	n/a	73
WDIP 2K <sub>dan</sub> 12	28	38	48	58	n/a	85

Tablica 92: Točnosti - kNN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 3K <sub>sat</sub> 13	67,74	69,89	70,97	72,04	72,04	73,12	72,04	87,02
FIB 3K <sub>sat</sub> 12	68,48	73,91	71,74	72,83	73,91	77,17	76,09	75,01
NWTIS 3K <sub>sat</sub> 13	93,33	93,34	93,3	93,35	93,33	93,25	93,36	89,47
NWTIS 3K <sub>sat</sub> 12	84,21	84,15	78,95	84,21	84,25	89,47	89,45	89,47
WDIP 3K <sub>sat</sub> 13	72,01	76,03	68,01	68,02	72,05	72,06	72,03	72,02
WDIP 3K <sub>sat</sub> 12	87,50	91,67	87,50	91,67	87,50	87,60	87,25	91,67

*Tablica 93: Minimalan broj koeficijenata za maksimalnu točnost - kNN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 3K <sub>sat</sub> 13	239	194	113	250	38	138	87	315
FIB 3K <sub>sat</sub> 12	173	207	363	13	169	41	73	186
NWTIS 3K <sub>sat</sub> 13	19	10	10	5	11	10	10	13
NWTIS 3K <sub>sat</sub> 12	208	158	83	92	28	14	273	9
WDIP 3K <sub>sat</sub> 13	64	58	117	21	10	91	153	200
WDIP 3K <sub>sat</sub> 12	206	81	112	194	91	103	208	197

*Tablica 94: Točnosti - kNN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 3K <sub>sat</sub> 13	56,99	61,29	62,37	58,06	62,37	61,29	68,82	63,44
FIB 3K <sub>sat</sub> 12	59,78	64,13	61,96	70,65	65,22	68,48	69,57	67,39
NWTIS 3K <sub>sat</sub> 13	86,67	86,67	93,33	86,67	86,67	80,00	86,67	86,67
NWTIS 3K <sub>sat</sub> 12	63,16	68,42	68,42	63,16	68,42	73,68	78,95	78,95
WDIP 3K <sub>sat</sub> 13	60,00	56,00	48,00	48,00	56,00	60,00	56,00	52,00
WDIP 3K <sub>sat</sub> 12	79,17	79,17	75,00	75,00	75,00	79,17	75,00	75,00

*Tablica 95: Puni broj koeficijenata - kNN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 3K <sub>sat</sub> 13	353	353	353	353	353	353	353	353
FIB 3K <sub>sat</sub> 12	403	403	403	403	403	403	403	403
NWTIS 3K <sub>sat</sub> 13	283	248	268	268	268	268	268	268
NWTIS 3K <sub>sat</sub> 12	323	258	303	263	278	328	324	324
WDIP 3K <sub>sat</sub> 13	353	353	353	353	353	353	353	353
WDIP 3K <sub>sat</sub> 12	308	323	313	348	353	353	353	353

Tablica 96: Točnosti - kNN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 3K <sub>dan</sub> 13	68,82	66,67	69,89	69,87	72,04	73,12
FIB 3K <sub>dan</sub> 12	82,61	70,65	73,91	76,09	72,83	85,87
NWTIS 3K <sub>dan</sub> 13	93,33	93,35	93,31	93,33	93,34	93,36
NWTIS 3K <sub>dan</sub> 12	84,21	78,95	84,21	84,25	84,21	78,95
WDIP 3K <sub>dan</sub> 13	68,01	64,02	68,04	68,05	64,03	68,05
WDIP 3K <sub>dan</sub> 12	79,17	79,15	87,50	79,16	n/a	83,33

Tablica 97: Minimalan broj koeficijenata za maksimalnu točnost - kNN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 3K <sub>dan</sub> 13	28	4	11	28	68	81
FIB 3K <sub>dan</sub> 12	6	5	19	20	4	20
NWTIS 3K <sub>dan</sub> 13	11	10	13	13	10	10
NWTIS 3K <sub>dan</sub> 12	14	15	31	10	21	7
WDIP 3K <sub>dan</sub> 13	23	19	40	7	7	71
WDIP 3K <sub>dan</sub> 12	4	5	48	4	n/a	15

Tablica 98: Točnosti - kNN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - puni broj koeficijenata

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 3K <sub>dan</sub> 13	68,82	63,44	64,52	65,59	68,82	66,67
FIB 3K <sub>dan</sub> 12	78,26	66,3	69,57	70,65	69,57	80,43
NWTIS 3K <sub>dan</sub> 13	86,67	80,00	86,67	80,00	86,67	80,00
NWTIS 3K <sub>dan</sub> 12	68,42	73,68	73,68	84,21	73,68	73,68
WDIP 3K <sub>dan</sub> 13	68,00	52,00	52,00	52,00	52,00	60,00
WDIP 3K <sub>dan</sub> 12	75,00	70,83	87,50	75,00	n/a	70,83

*Tablica 99: Puni broj koeficijenata - kNN – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova*

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 3K <sub>dan</sub> 13	28	38	48	58	73	88
FIB 3K <sub>dan</sub> 12	27	37	47	57	73	132
NWTIS 3K <sub>dan</sub> 13	28	38	48	58	73	147
NWTIS 3K <sub>dan</sub> 12	28	38	48	58	73	86
WDIP 3K <sub>dan</sub> 13	28	38	48	58	58	143
WDIP 3K <sub>dan</sub> 12	28	38	48	58		85

*Tablica 100: Točnosti - kNN – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	87,55	87,65	86,42	86,42	88,89
FMD 2K <sub>sat</sub> 12	90,79	90,81	90,85	90,75	90,79	92,11

*Tablica 101: Minimalan broj koeficijenata za maksimalnu točnost - kNN – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	31	33	14	14	39
FMD 2K <sub>sat</sub> 12	7	8	7	7	7	7

*Tablica 102: Točnosti - kNN – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata*

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	82,72	82,72	83,95	85,19	81,48
FMD 2K <sub>sat</sub> 12	88,16	86,84	89,47	90,79	89,47	88,16

*Tablica 103: Puni broj koeficijenata - kNN – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova*

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	38	48	58	78	103
FMD 2K <sub>sat</sub> 12	28	38	48	58	74	88

*Tablica 104: Točnosti - NB – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	77,39	76,41	76,34	77,45	75,27	77,42	77,42	72,04
FIB 2K <sub>sat</sub> 12	78,26	77,21	78,26	81,52	78,26	79,35	79,35	n/a
NWTIS 2K <sub>sat</sub> 13	99,01	98,27	99,06	98,44	99,33	97,87	99,05	97,44
NWTIS 2K <sub>sat</sub> 12	89,47	94,58	94,68	94,74	89,47	94,74	87,50	89,47
WDIP 2K <sub>sat</sub> 13	88,33	88,24	92,05	92,15	80,32	76,12	80,25	88,15
WDIP 2K <sub>sat</sub> 12	91,27	91,67	91,57	91,77	91,72	91,67	87,5	91,67

*Tablica 105: Minimalan broj koeficijenata za maksimalnu točnost - NB – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	7	8	8	8	9	314	312	166
FIB 2K <sub>sat</sub> 12	6	4	4	7	4	5	5	n/a
NWTIS 2K <sub>sat</sub> 13	13	7	4	6	4	102	5	257
NWTIS 2K <sub>sat</sub> 12	57	59	19	187	29	72	11	8
WDIP 2K <sub>sat</sub> 13	12	8	89	22	82	65	261	10
WDIP 2K <sub>sat</sub> 12	210	17	64	35	128	22	11	65

*Tablica 106: Točnosti - NB – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	64,52	74,19	72,04	64,52	68,82	68,82	68,82	66,67
FIB 2K <sub>sat</sub> 12	72,83	72,83	69,57	72,83	70,65	73,91	70,65	n/a
NWTIS 2K <sub>sat</sub> 13	86,67	86,67	93,33	86,67	86,67	86,67	93,33	86,67
NWTIS 2K <sub>sat</sub> 12	78,95	78,95	78,95	78,95	73,68	94,74	75,00	78,95
WDIP 2K <sub>sat</sub> 13	76,00	76,00	84,00	84,00	76,00	73,47	80,00	68,00
WDIP 2K <sub>sat</sub> 12	79,17	79,17	79,17	83,33	79,17	79,17	75,00	79,17

*Tablica 107: Puni broj koeficijenata - NB – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 2K <sub>sat</sub> 13	353	353	353	353	353	353	353	353
FIB 2K <sub>sat</sub> 12	403	403	403	403	403	403	403	n/a
NWTIS 2K <sub>sat</sub> 13	278	248	268	283	278	303	311	311
NWTIS 2K <sub>sat</sub> 12	323	258	563	263	283	328	368	368
WDIP 2K <sub>sat</sub> 13	373	442	403	403	403	403	403	403
WDIP 2K <sub>sat</sub> 12	308	313	313	343	363	373	368	363

*Tablica 108: Točnosti - NB – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 2K <sub>dan</sub> 13	77,42	79,57	78,38	78,45	78,49	73,12
FIB 2K <sub>dan</sub> 12	80,43	79,35	76,09	78,26	78,36	78,26
NWTIS 2K <sub>dan</sub> 13	93,45	93,43	93,26	93,56	93,33	93,42
NWTIS 2K <sub>dan</sub> 12	89,55	94,63	94,35	89,47	94,72	94,74
WDIP 2K <sub>dan</sub> 13	88,34	96,15	92,00	92,25	92,18	92,15
WDIP 2K <sub>dan</sub> 12	87,55	87,45	87,50	83,33	87,50	n/a

*Tablica 109: Minimalan broj koeficijenata za maksimalnu točnost - NB – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 2K <sub>dan</sub> 13	6	4	6	6	6	5
FIB 2K <sub>dan</sub> 12	5	6	5	4	4	4
NWTIS 2K <sub>dan</sub> 13	6	5	4	4	4	5
NWTIS 2K <sub>dan</sub> 12	10	17	21	10	60	7
WDIP 2K <sub>dan</sub> 13	13	9	5	56	23	81
WDIP 2K <sub>dan</sub> 12	8	23	38	4	39	n/a

*Tablica 110: Točnosti - NB – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - puni broj koeficijenata imanja*

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 2K <sub>dan</sub> 13	65,59	67,74	69,89	64,52	62,37	64,52
FIB 2K <sub>dan</sub> 12	61,96	63,04	57,61	59,78	63,04	65,22
NWTIS 2K <sub>dan</sub> 13	100,00	93,33	86,55	86,67	80,00	80,00
NWTIS 2K <sub>dan</sub> 12	78,95	89,47	78,95	73,68	78,95	78,95
WDIP 2K <sub>dan</sub> 13	80,00	84,00	84,00	84,00	84,00	84,00
WDIP 2K <sub>dan</sub> 12	79,17	79,17	83,33	83,33	83,33	n/a

*Tablica 111: Puni broj koeficijenata - NB – jednosemestralni predmeti – 2 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova*

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 2K <sub>dan</sub> 13	28	38	48	58	72	88
FIB 2K <sub>dan</sub> 12	28	38	48	58	73	133
NWTIS 2K <sub>dan</sub> 13	28	38	48	58	71	145
NWTIS 2K <sub>dan</sub> 12	28	38	48	58	73	86
WDIP 2K <sub>dan</sub> 13	28	38	48	58	73	143
WDIP 2K <sub>dan</sub> 12	28	38	48	58	85	n/a

*Tablica 112: Točnosti - NB – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 3K <sub>sat</sub> 13	62,37	60,22	60,22	61,29	60,22	61,25	61,15	62,37
FIB 3K <sub>sat</sub> 12	65,22	61,96	60,87	60,95	61,96	64,13	60,85	61,75
NWTIS 3K <sub>sat</sub> 13	86,67	86,57	93,23	86,77	86,57	86,47	93,33	80,25
NWTIS 3K <sub>sat</sub> 12	84,21	78,95	78,90	89,47	89,37	78,85	89,37	78,55
WDIP 3K <sub>sat</sub> 13	72,53	72,35	76,05	76,15	80,25	76,25	80,95	76,12
WDIP 3K <sub>sat</sub> 12	83,23	83,45	83,45	83,35	83,33	79,15	79,17	79,17

*Tablica 113: Minimalan broj koeficijenata za maksimalnu točnost - NB – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 3K <sub>sat</sub> 13	4	6	7	13	328	345	5	7
FIB 3K <sub>sat</sub> 12	8	388	381	338	7	6	6	4
NWTIS 3K <sub>sat</sub> 13	5	7	50	19	77	48	193	12
NWTIS 3K <sub>sat</sub> 12	181	13	30	16	83	10	43	13
WDIP 3K <sub>sat</sub> 13	73	45	282	255	315	193	305	13
WDIP 3K <sub>sat</sub> 12	9	14	16	27	4	9	11	20

*Tablica 114: Točnosti - NB – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata*

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 3K <sub>sat</sub> 13	48,39	52,69	53,76	47,31	51,61	50,54	52,69	49,46
FIB 3K <sub>sat</sub> 12	52,17	55,43	53,26	52,17	50,00	56,52	51,09	48,91
NWTIS 3K <sub>sat</sub> 13	73,33	80,00	73,33	80,00	66,67	73,33	86,67	73,33
NWTIS 3K <sub>sat</sub> 12	63,16	57,89	57,89	52,63	63,16	73,68	78,95	68,42
WDIP 3K <sub>sat</sub> 13	60,00	56,00	60,00	60,00	60,00	52,00	60,00	60,00
WDIP 3K <sub>sat</sub> 12	62,50	58,33	62,50	66,67	62,50	66,67	58,33	58,33



Tablica 115: Puni broj koeficijenata - NB – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova

	1000 sati	1400 sati	1800 sati	2000 sati	2150 sati	2500 sati	2800 sati	maks. sati
FIB 3K <sub>sat</sub> 13	353	353	349	353	353	353	353	353
FIB 3K <sub>sat</sub> 12	403	403	403	403	403	403	403	403
NWTIS 3K <sub>sat</sub> 13	283	250	268	283	278	303	311	311
NWTIS 3K <sub>sat</sub> 12	323	258	563	263	281	328	323	323
WDIP 3K <sub>sat</sub> 13	353	403	403	403	403	403	403	403
WDIP 3K <sub>sat</sub> 12	303	323	313	343	363	363	363	363

Tablica 116: Točnosti - NB – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 3K <sub>dan</sub> 13	60,22	58,03	58,06	61,29	58,06	59,14
FIB 3K <sub>dan</sub> 12	56,52	61,96	64,13	61,96	63,04	61,96
NWTIS 3K <sub>dan</sub> 13	86,54	86,45	86,59	86,55	86,67	80,00
NWTIS 3K <sub>dan</sub> 12	84,34	84,34	84,21	78,95	84,22	84,21
WDIP 3K <sub>dan</sub> 13	72,56	68,15	68,36	68,22	68,06	72,25
WDIP 3K <sub>dan</sub> 12	87,52	87,50	87,50	83,25	83,33	n/a

Tablica 117: Minimalan broj koeficijenata za maksimalnu točnost - NB – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 3K <sub>dan</sub> 13	4	5	14	4	4	7
FIB 3K <sub>dan</sub> 12	4	5	6	4	4	4
NWTIS 3K <sub>dan</sub> 13	4	22	6	6	8	7
NWTIS 3K <sub>dan</sub> 12	9	27	7	14	4	29
WDIP 3K <sub>dan</sub> 13	25	6	37	13	13	10
WDIP 3K <sub>dan</sub> 12	18	5	45	8	9	n/a

Tablica 118: Točnosti - NB – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova - puni broj koeficijenata

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 3K <sub>dan</sub> 13	54,84	53,76	52,69	54,84	52,69	51,61
FIB 3K <sub>dan</sub> 12	50,00	56,52	53,26	52,17	51,09	54,35
NWTIS 3K <sub>dan</sub> 13	80,00	66,67	53,33	73,33	60,00	60,00
NWTIS 3K <sub>dan</sub> 12	68,42	68,42	63,16	68,42	68,42	63,16
WDIP 3K <sub>dan</sub> 13	64,00	60,00	60,00	60,00	60,00	60,00
WDIP 3K <sub>dan</sub> 12	83,33	70,83	70,83	70,83	58,33	n/a

Tablica 119: Puni broj koeficijenata - NB – jednosemestralni predmeti – 3 kategorije - vrijeme kumulacije 1 dan– za različite dužine vremenskih nizova

	60 dana	80 dana	100 dana	120 dana	150 dana	maks. dana
FIB 3K <sub>dan</sub> 13	28	38	48	58	72	88
FIB 3K <sub>dan</sub> 12	28	38	48	58	73	133
NWTIS 3K <sub>dan</sub> 13	28	38	48	58	73	147
NWTIS 3K <sub>dan</sub> 12	28	38	48	58	73	86
WDIP 3K <sub>dan</sub> 13	28	38	48	58	58	143
WDIP 3K <sub>dan</sub> 12	28	38	48	58	86	n/a

Tablica 120: Točnosti - NB – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	85,19	82,72	83,95	85,19	82,72
FMD 2K <sub>sat</sub> 12	52,63	81,58	89,47	86,84	85,53	86,84

Tablica 121: Minimalan broj koeficijenata za maksimalnu točnost - NB – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - primijenjena tehnika sažimanja

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	7	5	4	4	6
FMD 2K <sub>sat</sub> 12	13	6	4	6	9	5

*Tablica 122: Točnosti - NB – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova - puni broj koeficijenata*

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	70,37	56,79	59,26	58,02	72,84
FMD 2K <sub>sat</sub> 12	46,05	50,00	52,63	59,21	61,84	77,63

*Tablica 123: Puni broj koeficijenata - NB – jednotjedni predmet – 2 kategorije - vrijeme kumulacije 1 sat– za različite dužine vremenskih nizova*

	60 sati	80 sati	100 sati	120 sati	150 sati	maks. sati
FMD 2K <sub>sat</sub> 13	n/a	38	48	58	78	188
FMD 2K <sub>sat</sub> 12	28	38	48	58	73	188

*Tablica 124: Usporedba točnosti predikcije za različite dužine vremenskih nizova (primijenjena tehnika sažimanja) po algoritmima i grupama predmeta – Friedmanov test, sumarni rezultati u tablici 19*

**matrica: ANN 3Kat sat**

-----  
-----  
FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:  
TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS

-----  
Number of observation: 48  
Number of blocks: 6  
Number of treatments: 8  
Number of replicates: 1

-----  
A (sum of squares of ranks): 1199.0000  
C (correction factor): 972.0000

-----  
Chi-square approximation (more conservative)  
Friedman test statistic T1 (uncorrected): 4.2709  
chi-square=T1 df=7 - p-value (2 tailed): 0.7481

-----  
F-statistic approximation (less conservative)  
Friedman test statistic T2 (corrected): 0.5660  
F=T2 df-num=7 df-denom=35 - p-value (2 tailed): 0.7782

-----  
The 8 treatments have identical effects

**matrica: ANN 2 Kat sat**

-----  
-----  
FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:  
TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS

-----  
Number of observation: 48  
Number of blocks: 6  
Number of treatments: 8  
Number of replicates: 1

-----  
A (sum of squares of ranks): 1199.5000

C (correction factor): 972.0000

---

Chi-square approximation (more conservative)  
Friedman test statistic T1 (uncorrected): 15.4462  
chi-square=T1 df=7 - p-value (2 tailed): 0.0307

---

F-statistic approximation (less conservative)  
Friedman test statistic T2 (corrected): 2.9085  
F=T2 df-num=7 df-denom=35 - p-value (2 tailed): 0.0166

---

The 8 treatments have not identical effects

### POST-HOC MULTIPLE COMPARISONS

---

Critical value: 14.2563

Absolute difference among mean ranks

0	0	0	0	0	0	0	0	0
0.5000	0	0	0	0	0	0	0	0
18.0000	18.5000	0	0	0	0	0	0	0
18.0000	18.5000	0	0	0	0	0	0	0
5.5000	6.0000	12.5000	12.5000	0	0	0	0	0
5.5000	5.0000	23.5000	23.5000	11.0000	0	0	0	0
7.5000	8.0000	10.5000	10.5000	2.0000	13.0000	0	0	0
5.0000	5.5000	13.0000	13.0000	0.5000	10.5000	2.5000	0	0

Absolute difference > Critical Value

0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	1	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

---

**matrica: ANN 3Kat dan**

---

FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:

## TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS

---

Number of observation: 36

Number of blocks: 6

Number of treatments: 6

Number of replicates: 1

---

A (sum of squares of ranks): 537.5000

C (correction factor): 441.0000

---

Chi-square approximation (more conservative)

Friedman test statistic T1 (uncorrected): 4.3782

chi-square=T1 df=5 - p-value (2 tailed): 0.4963

---

F-statistic approximation (less conservative)

Friedman test statistic T2 (corrected): 0.8544

F=T2 df-num=5 df-denom=25 - p-value (2 tailed): 0.5249

---

The 6 treatments have identical effects

---

**matrica: ANN 2Kat dan**

---

## FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS: TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS

---

Number of observation: 36

Number of blocks: 6

Number of treatments: 6

Number of replicates: 1

---

A (sum of squares of ranks): 527.5000

C (correction factor): 441.0000

---

Chi-square approximation (more conservative)

Friedman test statistic T1 (uncorrected): 5.9538

chi-square=T1 df=5 - p-value (2 tailed): 0.3107

---

F-statistic approximation (less conservative)

Friedman test statistic T2 (corrected): 1.2380

F=T2 df-num=5 df-denom=25 - p-value (2 tailed): 0.3212

-----  
The 6 treatments have identical effects  
-----

**matrica: ANN 2Kat sat FMD**  
-----  
-----

FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:  
TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS  
-----

Number of observation: 12

Number of blocks: 2

Number of treatments: 6

Number of replicates: 1  
-----

A (sum of squares of ranks): 178.0000

C (correction factor): 147.0000  
-----

Chi-square approximation (more conservative)

Friedman test statistic T1 (uncorrected): 5.1613

chi-square=T1 df=5 - p-value (2 tailed): 0.3965  
-----

F-statistic approximation (less conservative)

Friedman test statistic T2 (corrected): 1.0667

F=T2 df-num=5 df-denom=5 - p-value (2 tailed): 0.4726  
-----

The 6 treatments have identical effects  
-----

**matrica: SVM 2Kat sat**  
-----  
-----

FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:  
TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS  
-----

Number of observation: 48

Number of blocks: 6

Number of treatments: 8

Number of replicates: 1  
-----

A (sum of squares of ranks): 1217.5000

C (correction factor): 972.0000

---

Chi-square approximation (more conservative)  
Friedman test statistic T1 (uncorrected): 10.7780  
chi-square=T1 df=7 - p-value (2 tailed): 0.1486

---

F-statistic approximation (less conservative)  
Friedman test statistic T2 (corrected): 1.7260  
F=T2 df-num=7 df-denom=35 - p-value (2 tailed): 0.1349

---

The 8 treatments have identical effects

---

**matrica: SVM 2 Kat dan**

---

FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:  
TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS

---

Number of observation: 36  
Number of blocks: 6  
Number of treatments: 6  
Number of replicates: 1

---

A (sum of squares of ranks): 543.5000  
C (correction factor): 441.0000

---

Chi-square approximation (more conservative)  
Friedman test statistic T1 (uncorrected): 12.0488  
chi-square=T1 df=5 - p-value (2 tailed): 0.0341

---

F-statistic approximation (less conservative)  
Friedman test statistic T2 (corrected): 3.3560  
F=T2 df-num=5 df-denom=25 - p-value (2 tailed): 0.0187

---

The 6 treatments have not identical effects

POST-HOC MULTIPLE COMPARISONS

---

Critical value: 11.1748  
Absolute difference among mean ranks



0	0	0	0	0	0
14.0000	0	0	0	0	0
18.5000	4.5000	0	0	0	0
16.5000	2.5000	2.0000	0	0	0
17.5000	3.5000	1.0000	1.0000	0	0
17.5000	3.5000	1.0000	1.0000	0	0

Absolute difference > Critical Value

0	0	0	0	0	0
1	0	0	0	0	0
1	0	0	0	0	0
1	0	0	0	0	0
1	0	0	0	0	0
1	0	0	0	0	0

-----  
**matrica: SVM 2Kat sat FMD**  
 -----

-----  
 FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:  
 TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS  
 -----

Number of observation: 12  
 Number of blocks: 2  
 Number of treatments: 6  
 Number of replicates: 1

-----  
 A (sum of squares of ranks): 179.5000  
 C (correction factor): 147.0000  
 -----

Chi-square approximation (more conservative)  
 Friedman test statistic T1 (uncorrected): 4.2308  
 chi-square=T1 df=5 - p-value (2 tailed): 0.5167  
 -----

F-statistic approximation (less conservative)  
 Friedman test statistic T2 (corrected): 0.7333  
 F=T2 df-num=5 df-denom=5 - p-value (2 tailed): 0.6290  
 -----

The 6 treatments have identical effects  
 -----

matrica: kNN 3Kat sat

-----  
-----  
FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:  
TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS  
-----

Number of observation: 48  
Number of blocks: 6  
Number of treatments: 8  
Number of replicates: 1  
-----

A (sum of squares of ranks): 1216.0000  
C (correction factor): 972.0000  
-----

Chi-square approximation (more conservative)  
Friedman test statistic T1 (uncorrected): 13.8279  
chi-square=T1 df=7 - p-value (2 tailed): 0.0543  
-----

F-statistic approximation (less conservative)  
Friedman test statistic T2 (corrected): 2.4542  
F=T2 df-num=7 df-denom=35 - p-value (2 tailed): 0.0370  
-----

The 8 treatments have not identical effects

POST-HOC MULTIPLE COMPARISONS  
-----

Critical value: 15.2074

Absolute difference among mean ranks

0	0	0	0	0	0	0	0	0
13.5000	0	0	0	0	0	0	0	0
3.0000	16.5000	0	0	0	0	0	0	0
11.5000	2.0000	14.5000	0	0	0	0	0	0
12.0000	1.5000	15.0000	0.5000	0	0	0	0	0
20.5000	7.0000	23.5000	9.0000	8.5000	0	0	0	0
16.0000	2.5000	19.0000	4.5000	4.0000	4.5000	0	0	0
17.5000	4.0000	20.5000	6.0000	5.5000	3.0000	1.5000	0	0

Absolute difference > Critical Value

0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0
1	0	1	0	0	0	0	0
1	0	1	0	0	0	0	0

-----  
**matrica: kNN 2Kat sat**  
-----  
-----

FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:  
TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS  
-----

Number of observation: 48  
Number of blocks: 6  
Number of treatments: 8  
Number of replicates: 1  
-----

A (sum of squares of ranks): 1211.5000  
C (correction factor): 972.0000  
-----

Chi-square approximation (more conservative)  
Friedman test statistic T1 (uncorrected): 7.2046  
chi-square=T1 df=7 - p-value (2 tailed): 0.4079  
-----

F-statistic approximation (less conservative)  
Friedman test statistic T2 (corrected): 1.0353  
F=T2 df-num=7 df-denom=35 - p-value (2 tailed): 0.4247  
-----

The 8 treatments have identical effects  
-----  
-----

**matrica: kNN 3kat dan**  
-----  
-----

FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:  
TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS  
-----

Number of observation: 36

Number of blocks: 6  
Number of treatments: 6  
Number of replicates: 1

---

A (sum of squares of ranks): 542.5000  
C (correction factor): 441.0000

---

Chi-square approximation (more conservative)  
Friedman test statistic T1 (uncorrected): 9.7783  
chi-square=T1 df=5 - p-value (2 tailed): 0.0818

---

F-statistic approximation (less conservative)  
Friedman test statistic T2 (corrected): 2.4178  
F=T2 df-num=5 df-denom=25 - p-value (2 tailed): 0.0642

---

The 6 treatments have identical effects

---

-----  
**matrica: kNN 2Kat dan**

---

**FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:  
TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS**

---

Number of observation: 36  
Number of blocks: 6  
Number of treatments: 6  
Number of replicates: 1

---

A (sum of squares of ranks): 544.5000  
C (correction factor): 441.0000

---

Chi-square approximation (more conservative)  
Friedman test statistic T1 (uncorrected): 10.9179  
chi-square=T1 df=5 - p-value (2 tailed): 0.0530

---

F-statistic approximation (less conservative)  
Friedman test statistic T2 (corrected): 2.8608  
F=T2 df-num=5 df-denom=25 - p-value (2 tailed): 0.0355

-----  
The 6 treatments have not identical effects

POST-HOC MULTIPLE COMPARISONS

-----  
Critical value: 11.5775

Absolute difference among mean ranks

0	0	0	0	0	0
8.5000	0	0	0	0	0
12.5000	4.0000	0	0	0	0
16.5000	8.0000	4.0000	0	0	0
9.5000	1.0000	3.0000	7.0000	0	0
19.0000	10.5000	6.5000	2.5000	9.5000	0

Absolute difference > Critical Value

0	0	0	0	0	0
0	0	0	0	0	0
1	0	0	0	0	0
1	0	0	0	0	0
0	0	0	0	0	0
1	0	0	0	0	0

-----  
**matrica: kNN 2Kat sat FMD**

-----  
FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:  
TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS

-----  
Number of observation: 12

Number of blocks: 2

Number of treatments: 6

Number of replicates: 1

-----  
A (sum of squares of ranks): 181.0000

C (correction factor): 147.0000

-----  
Chi-square approximation (more conservative)

Friedman test statistic T1 (uncorrected): 9.7794

chi-square=T1 df=5 - p-value (2 tailed): 0.0817

-----  
 F-statistic approximation (less conservative)  
 Friedman test statistic T2 (corrected): 44.3333  
 F=T2 df-num=5 df-denom=5 - p-value (2 tailed): 0.0004  
 -----

The 6 treatments have not identical effects

POST-HOC MULTIPLE COMPARISONS

-----  
 Critical value: 1.9912

Absolute difference among mean ranks

0	0	0	0	0	0
2.5000	0	0	0	0	0
4.5000	2.0000	0	0	0	0
3.0000	5.5000	7.5000	0	0	0
1.5000	4.0000	6.0000	1.5000	0	0
6.5000	4.0000	2.0000	9.5000	8.0000	0

Absolute difference > Critical Value

0	0	0	0	0	0
1	0	0	0	0	0
1	1	0	0	0	0
1	1	1	0	0	0
0	1	1	0	0	0
1	1	1	1	1	0

-----  
**matrica: NB 3Kat sat**  
 -----  
 -----

FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:  
 TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS

-----  
 Number of observation: 48  
 Number of blocks: 6  
 Number of treatments: 8  
 Number of replicates: 1  
 -----

A (sum of squares of ranks): 1219.0000  
 C (correction factor): 972.0000  
 -----

Chi-square approximation (more conservative)

Friedman test statistic T1 (uncorrected): 4.7611  
chi-square=T1 df=7 - p-value (2 tailed): 0.6891

-----  
F-statistic approximation (less conservative)  
Friedman test statistic T2 (corrected): 0.6393  
F=T2 df-num=7 df-denom=35 - p-value (2 tailed): 0.7204  
-----

The 8 treatments have identical effects

-----  
**matrica: NB 2Kat sat**  
-----

FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:  
TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS

-----  
Number of observation: 48  
Number of blocks: 6  
Number of treatments: 8  
Number of replicates: 1  
-----

A (sum of squares of ranks): 1174.5000  
C (correction factor): 972.0000  
-----

Chi-square approximation (more conservative)  
Friedman test statistic T1 (uncorrected): 13.6370  
chi-square=T1 df=7 - p-value (2 tailed): 0.0580  
-----

F-statistic approximation (less conservative)  
Friedman test statistic T2 (corrected): 2.4040  
F=T2 df-num=7 df-denom=35 - p-value (2 tailed): 0.0405  
-----

The 8 treatments have not identical effects

POST-HOC MULTIPLE COMPARISONS

-----  
Critical value: 13.9008

Absolute difference among mean ranks

0	0	0	0	0	0	0	0
1.0000	0	0	0	0	0	0	0
3.0000	2.0000	0	0	0	0	0	0
20.5000	19.5000	17.5000	0	0	0	0	0
1.0000	2.0000	4.0000	21.5000	0	0	0	0
7.5000	6.5000	4.5000	13.0000	8.5000	0	0	0
2.0000	3.0000	5.0000	22.5000	1.0000	9.5000	0	0
1.0000	2.0000	4.0000	21.5000	0	8.5000	1.0000	0

Absolute difference > Critical Value

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0
0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0
0	0	0	1	0	0	0	0

-----  
**matrica: NB 3Kt dan**  
-----

-----  
FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:  
TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS  
-----

Number of observation: 36  
Number of blocks: 6  
Number of treatments: 6  
Number of replicates: 1

-----  
A (sum of squares of ranks): 542.0000  
C (correction factor): 441.0000  
-----

Chi-square approximation (more conservative)  
Friedman test statistic T1 (uncorrected): 3.3168  
chi-square=T1 df=5 - p-value (2 tailed): 0.6513  
-----

F-statistic approximation (less conservative)  
Friedman test statistic T2 (corrected): 0.6215  
F=T2 df-num=5 df-denom=25 - p-value (2 tailed): 0.6847  
-----

The 6 treatments have identical effects  
-----

**matrica: NB 2Kat dan**  
-----

-----  
FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:  
TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS  
-----

Number of observation: 36  
Number of blocks: 6  
Number of treatments: 6  
Number of replicates: 1



-----  
A (sum of squares of ranks): 545.0000  
C (correction factor): 441.0000  
-----

Chi-square approximation (more conservative)  
Friedman test statistic T1 (uncorrected): 4.6394  
chi-square=T1 df=5 - p-value (2 tailed): 0.4614  
-----

F-statistic approximation (less conservative)  
Friedman test statistic T2 (corrected): 0.9147  
F=T2 df-num=5 df-denom=25 - p-value (2 tailed): 0.4876  
-----

The 6 treatments have identical effects  
-----

**matrica: NB 2Kat sat FMD**  
-----

-----  
**FRIEDMAN TEST FOR IDENTICAL TREATMENT EFFECTS:  
TWO-WAY BALANCED, COMPLETE BLOCK DESIGNS**  
-----

Number of observation: 12  
Number of blocks: 2  
Number of treatments: 6  
Number of replicates: 1  
-----

A (sum of squares of ranks): 180.0000  
C (correction factor): 147.0000  
-----

Chi-square approximation (more conservative)  
Friedman test statistic T1 (uncorrected): 1.6667  
chi-square=T1 df=5 - p-value (2 tailed): 0.8931  
-----

F-statistic approximation (less conservative)  
Friedman test statistic T2 (corrected): 0.2000  
F=T2 df-num=5 df-denom=5 - p-value (2 tailed): 0.9490  
-----

The 6 treatments have identical effects  
-----

### **Kod 1:**

SQL upit koji podatke iz tablice studentskih logova za određeni predmet, dobivene iz LMS sustava (u kojem je vrijeme pristupa zabilježeno točno u sekundi kada se pristup dogodio), prebacuje u tablicu u kojoj je vrijeme pristupa zabilježeno u okviru vremena akumulacije T. U prikazanom slučaju to je 1 sat.

```
INSERT INTO FIBprposatu (id,time,user_jmbag) SELECT id,  
TRUNCATE((UNIX_TIMESTAMP(time)-1349042400)/3600,0), user_jmbag  
from FIBpr
```

gdje je 1349042400 vrijeme početka kolegija (30-09-2012 u ponoć za akademsku godinu 2012./2013.).

### **Kod 2:**

SQL upit kojim se iz tablice pristupa po satu grupiraju podaci po satu (danu) i studentu. Stvara se vremenski niz za svakog pojedinog studenta.

```
SELECT user_jmbag,time,count(user_jmbag) FROM `FIBprposatu`  
group by time,user_jmbag
```

### **Kod 3:**

Matlab kod kojim se podaci dobiveni SQL upitom (kod 2) preraspodjeljuju u Matlab matricu u kojoj su u prvoj koloni sati pristupa, a u ostalim kolonama studentski pristupi u tom satu. Ulazni podaci se nalaze u Matlab metrici FIBprpodanu sa studentskim pristupima grupiranim po satu i studentu te matrica s rezultatima ispita, a izlazni podaci su matrica u kojoj prva kolona sadrži podatke o vremenu (promatranim satima), a ostale kolone broj pristupa za svakog studenta u odgovarajućem satu.

```
l1=length(FIBprpodanu(:,1)); %dužina matrice s klikovima  
l2=length(uspjehFIBt(1,:)); %broj studenata  
for i=1:l2 po svim studentima  
    for j=1:269
```

```

%i maksimalnom broju vremenskih jedinica
    pr(j+1,i+1)=0;
    for k=1:11
        if(uspjehFIBt(2,i)==FIBprpodanu(k,1) &&
j==FIBprpodanu(k,2))
% smješta podatke o klikovima za svakog studenta
%u odovarajuće dane vodeći računa o studentu
        pr(j+1,i+1)=FIBprpodanu(k,3);
        end;
    end;
end;
end
for j=1:269
    pr(j+1,1)=j;
%upisuje dane u prvu kolonu
end;

```

#### Kod 4:

Matlab kod za računanje FT vremenskog niza svakog studenta, kod bira prvih N najintenzivnijih koeficijenata te odgovarajuće frekvencije, a zatim sortira po frekvencijama. Potom računa zastupljenost frekvencija i stvara konačnu matricu FT koeficijenata za svakog studenta sortirajući po zastupljenosti frekvencija. Ulazni podaci su: matrica vremenskih nizova svakog studenta i broj sati ili dana u kojem je vremenski niz generiran. Izlazni podatak je matrica najintenzivnijih i najzastupljenijih FT koeficijenata.

```

%dužina vremenskog niza za koji se računa FT /dani/
N=178;
Npol=floor(N/2);
bc=20;
fs=1;
f=-fs/2:fs/(N-1):fs/2;
svi(:,1)=f;
%unesi broj studenata od druge kolone do kraja
for i=2:1:65
%upisi naslov matrice s podacima
x=prnw24(1:N,i);
%računa FFT

```

```

z=abs(fftshift(fft(x))).^2;
svi(:,i)=z;
end
c=Npol+1;
for i=c:N
    b=i-Npol;
    svil(b,:)=svi(i,:);
end
%bira prvih N najvećih
for j=2:65
    s3=sortrows(svil,j);
    prvihN(1:bc,2*j-3)=s3(Npol-bc+1:Npol,1);
    prvihN(1:bc,2*j-2)=s3(Npol-bc+1:Npol,j)/Npol;
end
    prvihN=flipud(prvihN);
    [m,n] = size(prvihN);
    broj=0;
for k=1:2:n
    for i=1:m
        %izdvaja frekvencije koje odgovaraju najintenzivnijim N koef.
        broj=broj+1;
        fr(broj,1)=prvihN(i,k);
    end
end
%sortira po frekvencijama i izbacuje duplikate
frsort=sort(fr);
frfin=unique(frsort);
l1=length(frfin);
for k=1:l1
    p=frfin(k);
    broj=1;
    %puni konačnu tablicu FT koeficijenata
    for j=1:2:n
        broj=broj+1;
        kr(k,1)=p;
        kr(k,broj)=0;
        for i=1:m
            if (prvihN(i,j)==p)
                kr(k,broj)=prvihN(i,j+1);
            end
        end
    end
end
end
end
%racuna broj FT koef. različitih od nule za odgovarajuću frekv.
for k=1:l1
    brn=sum(kr(k,1:64)==0);

```

```

    kr(k,n/2+2)=n/2-brn;
end
%sortira po zastupljenosti /zadnjoj koloni/
krsort=sortrows(kr,n/2+2);
krsorta=flipud(krsort);

```

### Kod 5:

Matlab kod koji matricu najintenzivnijih i najzastupljenijih FT koeficijenata prebacuje u Excel datoteku spremnu za Rapidminer. Iz matrice studentskih uspjeha stvara kolonu s kategorijama. Ulazni podaci su matrica s FT koeficijentima i matrica sa studentskim uspjesima, a izlazni podatak je Excel tablica pripremljena za unos u Rapidminer.

```

A='krsorta';%ime excel datoteke
du=length(uspjehnw12(3,☺));
filename=strcat('D:\zarapidexcel\','a','.xlsx');
fns=eval(a);
[m,n] = size(fns);
frekve(1)=cellstr(' ');
for i=2:m+1
    b=int2str(i-1);
    frekve(i)=strcat('f',cellstr(b));
end
fns(:,n) = [];
%upisuje podatke u excel datoteku
xlswrite(filename,transpose(fns),1,'B1');
xlswrite(filename,frekve,1,'A1');
for i=1:du
    cc=int2str(i+1);
    xrange=strcat('A',cc);
    if(uspjehnw12(3,i)==3) xlswrite(filename,'m',1,xrange);end;
    if(uspjehnw12(3,i)<3) xlswrite(filename,'b',1,xrange); end;
    if(uspjehnw12(3,i)>3) xlswrite(filename,'g',1,xrange); end;
end;

```

### Kod 6:

Program napisan u Lazarusu napravljen za sortiranje i pregled rezultata modeliranja napravljenih u Rapidmineru. Program je napravljen u RAD alatu Lazarus (Free Pascal). Ulazni podaci su \*.res datoteke koje generira Rapidminer i u kojima se nalaze rezultati modeliranja, a izlazni podaci su

točnosti predikcije za određeni broj upotrebljenih FT koeficijenata. Točnosti su sortirane po veličini.

```
unit kin1;
{$mode objfpc}{$H+}
interface
uses
  Classes, SysUtils, FileUtil, Forms, Controls, Graphics,
  Dialogs, StdCtrls,
  Grids, ComCtrls, StrUtils, graf, fpSpreadsheet,
  fpspreadsheetctrls,
  fpsallformats, xlsbiff5, fpstypes;
type
  { TForm1 }
  TForm1 = class(TForm)
    Button1: TButton;
    Button10: TButton;
    Button11: TButton;
    Button12: TButton;
    Button2: TButton;
    Button3: TButton;
    Button4: TButton;
    Button5: TButton;
    Button6: TButton;
    Button7: TButton;
    Button8: TButton;
    Button9: TButton;
    Edit1: TEdit;
    Edit2: TEdit;
    Edit3: TEdit;
    Edit4: TEdit;
    Edit5: TEdit;
    brojkc: TEdit;
    Edit6: TEdit;
    Label1: TLabel;
    Label3: TLabel;
    OpenFileDialog1: TOpenDialog;
    PageControl1: TPageControl;
    SaveDialog1: TSaveDialog;
    StringGrid1: TStringGrid;
    StringGrid2: TStringGrid;
    StringGrid3: TStringGrid;
    StringGrid4: TStringGrid;
    StringGrid5: TStringGrid;
    StringGrid6: TStringGrid;
    StringGrid7: TStringGrid;
```

```

sWorkbookSource1: TsWorkbookSource;
TabSheet1: TTabSheet;
TabSheet2: TTabSheet;
TabSheet3: TTabSheet;
TabSheet4: TTabSheet;
TabSheet5: TTabSheet;
TabSheet6: TTabSheet;
TabSheet7: TTabSheet;
procedure brojkcChange(Sender: TObject);
procedure Button10Click(Sender: TObject);
procedure Button11Click(Sender: TObject);
procedure Button12Click(Sender: TObject);
procedure Button1Click(Sender: TObject);
procedure Button2Click(Sender: TObject);
procedure Button3Click(Sender: TObject);
procedure Button4Click(Sender: TObject);
procedure Button5Click(Sender: TObject);
procedure Button6Click(Sender: TObject);
procedure Button7Click(Sender: TObject);
procedure Button8Click(Sender: TObject);
procedure Button9Click(Sender: TObject);
procedure FormActivate(Sender: TObject);
procedure FormResize(Sender: TObject);
procedure PageControl1Change(Sender: TObject);
procedure StringGrid1SelectCell(Sender: TObject; aCol, aRow:
Integer;
    var CanSelect: Boolean);
procedure TabControl1Change(Sender: TObject);
    private
        mbf:integer;
    { private declarations }
    naslov:string;
public
    { public declarations }
end;
var
    Form1: TForm1;
implementation
{$R *.lfm}
{ TForm1 }
function zadnji(i:integer; var a: tstringgrid):integer;
var j:integer;
begin
    j:=0;
    repeat
        j:=j+1;
    until a.Cells[i,j]='';

```

```

    zadnji:=j-1;
end;
function zadured(var a: tstringgrid):integer;
var j:integer;
begin
    j:=0;
    repeat
        j:=j+1;
    until a.Cells[j,1]='';
    zadured:=j-1;
end;
procedure swap( var a, b:String );
var
    temp : String;
begin
    temp := a;
    a := b;
    b := temp;
end;
procedure BubbleSort( var a: tstringgrid );
var
    n,newn, i,k:integer;
    tempx,tempy:string;
begin
    with a do
        begin
            for k:=0 to form1.mbf-1 do
                begin
                    n := zadnji( 2+k*2,a );
                    repeat
                        newn := 0;
                        for i := 2 to n do
                            begin
                                if cells[2+k*2,i-1] > cells[2+k*2,i] then
                                    begin
                                        tempy:=cells[2+k*2,i-1];
                                        tempx:=cells[1+k*2,i-1];
                                        Cells[2+k*2,i-1]:=Cells[2+k*2,i];
                                        Cells[2+k*2,i]:=tempy;
                                        Cells[1+k*2,i-1]:=Cells[1+k*2,i];
                                        Cells[1+k*2,i]:=tempx;
                                        newn := i ;
                                    end;
                            end;
                        end ;
                    n := newn;
                until n = 0;
            end;
        end;
end;

```



```

end;
end;
procedure TForm1.Button1Click(Sender: TObject);
var ppx,filename,sa,sbr,py,px,pz,fn,fn1 : string;
    punostr:TStringList;
    brftk,k1,bkc,bkc1,i,j,br,br1,lna,preskoci,f:integer;
    newtab:ttabsheet;
    newgrid:TStringgrid;
begin
    preskoci:=strtoint(edit3.text);
    if OpenFileDialog1.Execute then
    begin
        filename := OpenFileDialog1.FileName;
        br:=OpenDialog1.Files.Count;
        mbf:=br;
        lna:=0;
        edit6.Text:=OpenDialog1.Files.Strings[0];
        naslov:=OpenDialog1.Files.Strings[0];
        bkc1:=0;
        repeat
            bkc1:=bkc1+1;
            fn:=ExtractWord(bkc1,edit6.Text,['\']);
        until fn='';
        brojkc.Text:=Inttostr(bkc1-1);
        bkc:=bkc1-1;
        Edit6.text:=StringReplace(
ExtractWord(bkc,edit6.Text,['\'],'.res','.txt',[rfReplaceAll]);
        for i:=0 to br-1 do
            begin
                lna:=0;
                bkc:=Strtoint(brojkc.Text);
                punostr:=TstringList.Create;

                sa:= OpenFileDialog1.Files.Strings[i];
                punostr.LoadFromFile(sa);

                fn:=ExtractWord(bkc,sa,['\']);
                fn:=StringReplace(fn,Edit4.text,'',[rfReplaceAll]);
                fn:=StringReplace(fn,Edit5.text,'',[rfReplaceAll]);
                br1:=punostr.Count;

                StringGrid1.ColCount:=1120;
                StringGrid1.RowCount:=1000;
                StringGrid3.ColCount:=1120;
                StringGrid3.RowCount:=1000;
                StringGrid4.ColCount:=1120;
                StringGrid4.RowCount:=1000;
            end;
        end;
    end;
end;

```

```

StringGrid1.Cells[2+2*i,0]:=fn;
f:=0;
  for j:=0 to br1-1 do
    begin
      px:=ExtractWord(1,punostr.Strings[j],[' ']);
      py:=ExtractWord(2,punostr.Strings[j],[' ']);
      pz:=ExtractWord(3,punostr.Strings[j],[' ']);
      if (pz='PerformanceVector:') then
        begin
          end;
          if (px='accuracy:') then
            begin
StringGrid1.Cells[2+i*2,lna+1]:=StringReplace(copy(py,1,5),' ','.'
.[rfReplaceAll]);
            end;
            if (copy(px,1,1)='f') then
              begin
                ppx:=ExtractWord(1,px,['f']);

ppx:=StringReplace(ppx,':','',[rfReplaceAll]);
                StringGrid1.Cells[1+2*i,lna+2]:= ppx;
                f:=f+1;
            end;
            if (copy(px,1,1)='') then lna:=lna+1;
            if (pz='AttributeWeights:') then
              begin
                lna:=lna-1;
                f:=0;
              end;

            end;
            punostr.Clear;
            end;
            StringGrid1.RowCount:=1000;
            punostr.Destroy;
          end;
        {for i:=1 to Stringgrid1.RowCount-1 do
          begin
            brftk:=0;
            for j:=4 to Stringgrid1.ColCount-1 do if
stringgrid1.Cells[j,i]<>' ' then brftk:=brftk+1;
            stringgrid1.cells[4,i]:=Inttostr(brftk);
            end;}
          end;
        procedure TForm1.brojkcChange(Sender: TObject);
        begin

```

```

end;
procedure TForm1.Button10Click(Sender: TObject);
begin
  StringGrid1.Clean;
  StringGrid1.RowCount:=5;
  StringGrid1.ColCount:=5;
  StringGrid2.Clean;
  StringGrid2.RowCount:=5;
  StringGrid2.ColCount:=5;
end;
procedure TForm1.Button11Click(Sender: TObject);
var n,i,z,vis,j,l:integer;
begin
  for i := 0 to StringGrid1.RowCount - 1 do
    StringGrid3.Rows[i].Assign(StringGrid1.Rows[i]);
  for i := 0 to StringGrid2.RowCount - 1 do
    StringGrid4.Rows[i].Assign(StringGrid2.Rows[i]);
    mbf:=round(zadured(stringgrid1)/2);
  for l:=0 to mbf-1 do
    begin
      n:=zadnji(2+l*2,stringgrid1);
      j:=1;
      for i:=n-3 to n do
        begin
          Stringgrid7.cells[1+l,j]:=
stringgrid1.cells[2+l*2,i];
          Stringgrid7.cells[2+mbf+l,j]:=
stringgrid1.cells[1+l*2,i];
          j:=j+1;
        end;
      end;
      BubbleSort(stringgrid3);
      BubbleSort(stringgrid4);
      for i:=1 to 23 do
        begin
          StringGrid5.Cells[i,1]:=StringGrid3.Cells[i,0];
          StringGrid6.Cells[i,1]:=StringGrid3.Cells[i,0];
        end;
      {vis:=0;
      for i:=0 to 23 do
        begin
          z:=zadnji(1+2*i,stringgrid1);
          if z>vis then vis:=z;
        end;
      for j:=1 to 23 do
        for i:=vis-7 to vis do
          begin

```

```

        StringGrid5.Cells[j,i-vis-7+2]:=StringGrid3.Cells[j,i-
vis-7+1];
        StringGrid6.Cells[j,i-vis-7+2]:=StringGrid3.Cells[j,i-
vis-7+2];
        end;}
    end;
procedure TForm1.Button12Click(Sender: TObject);
var filename,fn,sa,px,py:string;
    vis,duz,br,lna,bkcl,bkc,i,br1,j,k,br2:integer;
    punostr,list1:TStringList;
begin
    StringGrid1.ColCount:=1120;
    StringGrid1.RowCount:=2000;
    StringGrid2.ColCount:=1120;
    StringGrid2.RowCount:=2000;
    StringGrid3.ColCount:=1120;
    StringGrid3.RowCount:=2000;
    StringGrid4.ColCount:=1120;
    StringGrid4.RowCount:=2000;
    StringGrid5.ColCount:=1120;
    StringGrid5.RowCount:=2000;
    StringGrid6.ColCount:=1120;
    StringGrid6.RowCount:=2000;
    StringGrid7.ColCount:=1120;
    StringGrid7.RowCount:=2000;
    if OpenFileDialog1.Execute then
    begin
        filename := OpenFileDialog1.FileName;
        br:=OpenDialog1.Files.Count;

        edit6.Text:=OpenDialog1.Files.Strings[0];
        naslov:=OpenDialog1.Files.Strings[0];
        bkcl:=0;
        repeat
            bkcl:=bkcl+1;
            fn:=ExtractWord(bkcl,edit6.Text,['\']);
        until fn='';
        brojkc.Text:=Inttostr(bkcl-1);
        bkc:=bkcl-1;
        Edit6.text:=StringReplace(
ExtractWord(bkc,edit6.Text,['\'],'.res','.txt',[rfReplaceAll]);
        for i:=0 to br-1 do
            begin
                punostr:=TstringList.Create;
                list1:=Tstringlist.Create;
                list1.Delimiter := ',';
                sa:= OpenFileDialog1.Files.Strings[i];

```

```

punostr.LoadFromFile(sa);
br1:=punostr.Count;
lna:=0;
  for j:=0 to br1-1 do
    begin
      px:=punostr.strings[j];
      list1.DelimitedText:=px;
      br2:=list1.Count;
      mbf:=round(br2/2);

      try
        if i=0 then
          begin
            for k:=0 to br2-1 do
              begin

StringGrid1.Cells[k+1,lna]:=StringReplace(List1[k],',','.',[rfRe
placeAll]);

                end;
              end;
            if i=1 then
              begin
                for k:=0 to br2-1 do
                  begin

StringGrid2.Cells[k+1,lna]:=StringReplace(List1[k],',','.',[rfRe
placeAll]);

                    end;
                  end;
                lna:=lna+1;
              except
            end;
          end;
        end;
      end;
    end;
  end;
end;
procedure TForm1.Button2Click(Sender: TObject);
begin
  close;
end;
procedure TForm1.Button3Click(Sender: TObject);
const
  Str_Worksheet1 = 'Original';
  Str_Worksheet2 = 'Korekcija';
  Str_Worksheet3 = 'sortOrig';
  Str_Worksheet4 = 'sortkorr';
var MyWorkbook: TsWorkbook;

```

```

MyWorksheet: TsWorksheet;
duz,vis,nx,ny,i,j,z:integer;
dname1,dname2,glava,dname,linija:string;
f:TextFile;
begin
duz:=opendialog1.Files.Count;
if duz<3 then duz:=mbf;
if SaveDialog1.Execute then
begin
dname:= ExtractFilePath(SaveDialog1.FileName)+'rezultat';
createdir(dname);

vis:=0;
for i:=0 to duz-1 do
begin
z:=zadnji(1+2*i,stringgrid1);
if z>vis then vis:=z;
end;
dname1:=dname+'/'+'edit6.text+'orig.csv';
AssignFile(f,dname1);
{$I-}
Rewrite(f);
{$I-}
with StringGrid1 do
begin
for j:=0 to vis do
begin
linija:='';
for i:=0 to duz-1 do
begin

linija:=linija+Cells[1+i*2,j]+' '+Cells[2+i*2,j]+' ';
end;
Writeln(f,linija);
end;
end;
closeFile(f);
vis:=0;
for i:=0 to duz-1 do
begin
z:=zadnji(1+2*i,stringgrid2);
if z>vis then vis:=z;
end;
dname1:=dname+'/'+'edit6.text+'korekcija.csv';
AssignFile(f,dname1);
{$I-}

```

```

Rewrite(f);
{$I-}
with StringGrid2 do
begin
  for j:=0 to vis do
  begin
    linija:='';
    for i:=0 to duz-1 do
    begin

linija:=linija+Cells[1+i*2,j]+' '+Cells[2+i*2,j]+' ';

      end;
      Writeln(f,linija);
    end;
  end;
  closeFile(f);
  vis:=0;
  for i:=0 to duz-1 do
  begin
    z:=zadnji(1+2*i,stringgrid1);
    if z>vis then vis:=z;
  end;
  dname1:=dname+'/'+'edit6.text'+ 'SortOrig.csv';
  AssignFile(f,dname1);
  {$I-}
  Rewrite(f);
  {$I-}
  with StringGrid3 do
  begin
    for j:=0 to vis+6 do
    begin
      linija:='';
      for i:=0 to duz+6 do
      begin

linija:=linija+Cells[1+i*2,j]+' '+Cells[2+i*2,j]+' ';

      end;
      Writeln(f,linija);
    end;
  end;
  vis:=0;
  for i:=0 to duz-1 do
  begin
    z:=zadnji(1+2*i,stringgrid1);
    if z>vis then vis:=z;
  end;

```

```

        end;
dname1:=dname+'/'+'edit6.text+'SortKorekc.csv';
AssignFile(f,dname1);
{$I-}
Rewrite(f);
{$I-}
with StringGrid4 do
begin
    for j:=0 to vis+6 do
        begin
            linija:='';
            for i:=0 to duz+6 do
                begin
linija:=linija+Cells[1+i*2,j]+' '+Cells[2+i*2,j]+' ';

                    end;
                    Writeln(f,linija);
                end;
            end;
dname1:=dname+'/'+'edit6.text+'SveSkup.xls';
for i:=0 to duz-1 do
begin
    z:=zadnji(1+2*i,stringgrid1);
    if z>vis then vis:=z;
end;
MyWorkbook := TsWorkbook.Create;
MyWorksheet:=
MyWorkbook.AddWorksheet(UTF8ToAnsi(Str_Worksheet1));
MyWorksheet.WriteVertAlignment(0, 0, vaCenter);
for i:=0 to duz-1 do
begin
    for j:=1 to vis do
        begin
            MyWorksheet.WriteUTF8Text(0, 2+i*2,
Stringgrid1.cells[2+i*2,0]);
            MyWorksheet.WriteUTF8Text(j, 2+i*2,
Stringgrid1.cells[2+i*2,j]);
            MyWorksheet.WriteUTF8Text(j, 1+i*2,
Stringgrid1.cells[1+i*2,j]);
        end;
    end;
    for i:=0 to duz-1 do
        begin
            z:=zadnji(1+2*i,stringgrid1);
            if z>vis then vis:=z;
        end;

```



```

MyWorksheet :=
MyWorkbook.AddWorksheet(UTF8ToAnsi(Str_Worksheet2));
MyWorksheet.WriteVerticalAlignment(0, 0, vaCenter);
for i:=0 to duz-1 do
begin
for j:=1 to vis do
begin
MyWorksheet.WriteUTF8Text(0, 2+i*2,
Stringgrid2.cells[2+i*2,0]);
MyWorksheet.WriteUTF8Text(j, 2+i*2,
Stringgrid2.cells[2+i*2,j]);
MyWorksheet.WriteUTF8Text(j, 1+i*2,
Stringgrid2.cells[1+i*2,j]);
end;
end;
for i:=0 to duz-1 do
begin
z:=zadnji(1+2*i,stringgrid1);
if z>vis then vis:=z;
end;
MyWorksheet :=
MyWorkbook.AddWorksheet(UTF8ToAnsi(Str_Worksheet3));
MyWorksheet.WriteVerticalAlignment(0, 0, vaCenter);
for i:=0 to duz-1 do
begin
for j:=1 to vis do
begin
MyWorksheet.WriteUTF8Text(0, 2+i*2,
Stringgrid3.cells[2+i*2,0]);
MyWorksheet.WriteUTF8Text(j, 2+i*2,
Stringgrid3.cells[2+i*2,j]);
MyWorksheet.WriteUTF8Text(j, 1+i*2,
Stringgrid3.cells[1+i*2,j]);
end;
end;
for i:=0 to duz-1 do
begin
z:=zadnji(1+2*i,stringgrid1);
if z>vis then vis:=z;
end;
MyWorksheet :=
MyWorkbook.AddWorksheet(UTF8ToAnsi(Str_Worksheet4));
MyWorksheet.WriteVerticalAlignment(0, 0, vaCenter);
for i:=0 to duz-1 do
begin
for j:=1 to vis do
begin

```

```

        MyWorksheet.WriteUTF8Text(0,                2+i*2,
Stringgrid4.cells[2+i*2,0]);
        MyWorksheet.WriteUTF8Text(j,                2+i*2,
Stringgrid4.cells[2+i*2,j]);
        MyWorksheet.WriteUTF8Text(j,                1+i*2,
Stringgrid4.cells[1+i*2,j]);
        end;
    end;
    MyWorkbook.WriteToFile(dname1, sfExcel5, true);
end;
procedure TForm1.Button4Click(Sender: TObject);
begin
    Form2.ShowModal;
end;
procedure TForm1.Button5Click(Sender: TObject);
begin
    StringGrid1.Clean;
    StringGrid7.Clean;
    StringGrid1.RowCount:=5;
    StringGrid1.ColCount:=5;
    StringGrid2.Clean;
    StringGrid2.RowCount:=5;
    StringGrid2.ColCount:=5;
    StringGrid3.Clean;
    StringGrid3.RowCount:=5;
    StringGrid3.ColCount:=5;
    StringGrid4.Clean;
    StringGrid4.RowCount:=5;
    StringGrid4.ColCount:=5;
end;
procedure TForm1.Button6Click(Sender: TObject);
begin
    stringgrid1.RowCount:=strtoint(edit2.Text);
end;
procedure TForm1.Button7Click(Sender: TObject);
var
    bkcl:integer;
begin
    bkcl:=strtoint(brojkc.text);
    Edit6.text:=StringReplace(
ExtractWord(bkcl,Edit6.text,['\']),'.res','.txt',[rfReplaceAll])
;end;
procedure TForm1.Button8Click(Sender: TObject);
begin
    edit6.text:=naslov;
end;

```

```

procedure TForm1.Button9Click(Sender: TObject);
var nx,ny,i,j:integer;
    linija:string;
    f:TextFile;
begin
    SaveDialog1.FileName:=edit6.Text;
    if SaveDialog1.Execute then
        begin
            AssignFile(f,SaveDialog1.FileName);
            {$I-}
            Rewrite(f);
            {$I+}
            nx:= StringGrid2.ColCount;
            ny:= StringGrid2.RowCount;
            with StringGrid2 do
                begin
                    for j:=1 to ny-1 do
                        begin
                            linija:='';
                            for i:=1 to nx-1 do
                                begin
                                    linija:=linija+Cells[i,j]+' ';
                                end;
                                Writeln(f,linija);
                            end;
                            closeFile(f);
                        end;
                    ShowMessage(linija);
                end;
            end;
end;

procedure TForm1.FormActivate(Sender: TObject);
begin
    edit3.Text:='0';
end;

procedure TForm1.FormResize(Sender: TObject);
var sirf,visf:integer;
begin
    sirf:= Form1.Width;
    visf:= Form1.Height;
    StringGrid1.Width:=sirf-100;
    StringGrid1.Height:=visf-150;
end;

procedure TForm1.PageControl1Change(Sender: TObject);
begin
end;

```

```
procedure TForm1.StringGrid1SelectCell(Sender: TObject; aCol,
aRow: Integer;
var CanSelect: Boolean);
begin
    edit1.Text:=inttostr(aCol);
    edit2.Text:=inttostr(aRow);
end;

procedure TForm1.TabControl1Change(Sender: TObject);
begin
end;

end.
```

## ŽIVOTOPIS

Jasna Gamulin (djevojački Čop) rođena je 29. rujna 1964. u Zagrebu. Maturirala je 1983. godine u Pedagoškoj školi u Zagrebu. Ekonomski fakultet u Zagrebu je upisala 1984., a diplomirala je 1989. na smjeru Međunarodna suradnja, radom iz ekonomske analize. Tijekom studija bila je na stručnoj praksi u trajanju od 45 dana u banci Credito Italiano u Napulju, Italija. Također, tijekom studija je dobila stipendiju Plive za nadarene studente. Po završetku studija zapošljava se u Plivi u odjelu Nabave gdje je tijekom vremena napredovala do radnog mjesta koordinatorice u Nabavi. Odjeljivanjem Cedevite d.o.o. u zasebno poduzeće, nastavlja raditi na mjestu voditeljice u Nabavi. Nakon gotovo 20 godina rada na poslovima u nabavi, 2008. godine upisuje doktorski studij Informacijske znanosti na Fakultetu organizacije i informatike te prelazi na radno mjesto voditeljice ureda za studente Studija na engleskom jeziku na Medicinskom fakultetu Sveučilišta u Zagrebu, Centar za međunarodnu suradnju. Tijekom dugogodišnjeg radnog staža, mogu se uočiti dvije izrazite sklonosti: sklonost prema rudarenju podataka i sklonost prema radu u međunarodnom okruženju. Iako to možda na prvi pogled nije vidljivo, pronalaženje znanja u podacima, a također i rad s ljudima iz mnogobrojnih zemalja, poveznice su svih poslova koje je obavljala: od diplomskog rada iz ekonomske analize temeljenog na podacima iz poslovanja Plive, preko izrade nabavnih izvještaja koje je godinama izrađivala uz poslove uvoza, pa do rada u uredu za studente Studija medicine na engleskom i izrade znanstvenih radova iz područja web-formativnog ispitivanja studenata i predviđanja uspjeha studenata pomoću dubinske analize podataka.

Udata je i ima četvero djece. U slobodnom vremenu volontira na raznim poslovima u župi Svete Obitelji u Zagrebu, gdje s mužem Ozrenom koordinira rad obiteljske zajednice.

**Popis radova:**

1. Gamulin, J., O. Gamulin, D. Kermek (2016) Using Fourier coefficients in time series analysis for student performance prediction in blended learning environments, *Expert Systems*, 33-2, 176-188, DOI: 10.1111/exsy.12142.
2. Gamulin, J., O. Gamulin, D. Kermek (2015) The application of formative e-assessment data in final exam results modeling using neural networks, in *Proceedings of the 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2015/CE)*, Opatija, Croatia, IEEE Conference Publications, 726 - 730. DOI: 10.1109/MIPRO.2015.7160367
3. Gamulin, J., O. Gamulin, D. Kermek (2014) Comparing classification models in the final exam performance prediction, in *Proceedings of the 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2014/CE)*, Opatija, Croatia, IEEE Conference Publications, 663 - 668. DOI: 10.1109/MIPRO.2014.6859650
4. Gamulin, J., O. Gamulin, D. Kermek (2013) Data mining in hybrid learning: possibility to predict the final exam result, in *Proceedings of the 36th International Convention, Conference Computers in Education (MIPRO 2013/CE)*, Opatija, Croatia, IEEE Conference Publications, 591 - 596.
5. Gamulin, J., O. Gamulin (2012) The application of web-based formative quizzes in laboratory teaching in higher education environment, in *Proceedings of the 35th International Convention (MIPRO 2012/CE)*, Opatija, Croatia, IEEE Conference Publications, 1383 - 1388.
6. Gamulin, J., O. Gamulin (2011) Enhancing laboratory teaching in higher education environment using web-based formative colloquiums, in *Proceedings of the 34th International Convention (MIPRO 2011/CE)*, Opatija, Croatia, IEEE Conference Publications, 1189 - 1194.

7. Musulin, J., J. Gamulin, I.H., Crnojevac (2011) Knowledge management in Tourism: the importance of tacit knowledge and the problem of its elicitation and sharing, in *Proceedings of the 34th International Convention (MIPRO 2011/CTS)*, Opatija, Croatia, IEEE Conference Publications, 981 - 987.
8. Gamulin, J. (2010) The Model of Raw and Packaging Materials Cost Estimation Using Neural Networks, in *Proceedings of the 21st Central European Conference on Information and Intelligent Systems*, Fakultet organizacije i informatike, Varaždin, Croatia, 377-383.
9. Gamulin, J., J. Gugić, O. Gamulin (2010) Improving classroom teaching in higher education environment using web-based formative assessment, in *Proceedings of the 33rd International Convention (MIPRO 2010/CE)*, Opatija, Croatia, IEEE Conference Publications, 1001 - 1006.