Primjena tehnika rudarenja podataka u analitici učenja

Forjan, Dario

Undergraduate thesis / Završni rad

2020

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

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:211:160103

Rights / Prava: Attribution-NonCommercial-NoDerivs 3.0 Unported/Imenovanje-Nekomercijalno-Bez prerada 3.0

Download date / Datum preuzimanja: **2024-03-29**

SVEUČILIŠTE U ZAGREBU
FAKULTET ORGANIZACIJE I INFORMATIKE
VARAŽDIN

Repository / Repozitorij:

Faculty of Organization and Informatics - Digital Repository

zir.nsk.hr

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

SVEUČILIŠTE U ZAGREBU

FAKULTET ORGANIZACIJE I INFORMATIKE

V A R A Ž D I N

Dario Forjan

PRIMJENA TEHNIKA RUDARENJA PODATAKA U ANALITICI UČENJA

ZAVRŠNI RAD

Varaždin, 2020.

SVEUČILIŠTE U ZAGREBU

FAKULTET ORGANIZACIJE I INFORMATIKE

V A R A Ž D I N

Dario Forjan

Studij: Poslovni sustavi

PRIMJENA TEHNIKA RUDARENJA PODATAKA U ANALITICI UČENJA

ZAVRŠNI RAD

Mentorica:

Doc. dr. sc. Dijana Oreški

Varaždin, srpanj 2020

Dario Forjan

Izjava o izvornosti

Izjavljujem da je moj završni/diplomski rad izvorni rezultat mojeg rada te da se u izradi istoga nisam koristio drugim izvorima osim onima koji su u njemu navedeni. Za izradu rada su korištene etički prikladne i prihvatljive metode i tehnike rada.

Autor/Autorica potvrdio/potvrdila prihvaćanjem odredbi u sustavu FOI-radovi

_____

Sažetak

Tema ovog rada je proučavanje kako korištenje sustava za e-učenje utječe na prolaznost studenata na kolegiju. U radu će biti prikazano kako pojedine aktivnosti utječu na konačan ishod odnosno, konačnu ocjenu studenata. Na početku rada, nakon uvoda u temu, govorit ćemo o rudarenju podataka, te nekim njihovim zadaćama i fazana samog procesa rudarenja podataka. Nadalje definirana su neka prethodna istraživanja, koja su slična ovoj temi, kako bi se kasnije mogla napraviti analiza, tj. usporedba sa ovim radom. Nakon toga, definirani su podaci i opis problema kojeg je potrebno analizirati. Zatim su detaljno objašnjene sve metode tehnike rudarenja koje su korištene u analizi podataka sa sustava za e-učenje, i nakon toga po svakoj metodi je napravljena detaljna analiza podataka primjerenih za tu metodu. Na kraju rada je napravljen zaključak te popis literature, tablica i slika.

Ključne riječi: e-učenje, tehnike rudarenja podataka, učenje, klaster, stablo odlučivanja, neuronska mreža

1. Uvod 1
2. Rudarenje podataka 2
2.1 Definicija rudarenja podataka 2
2.2 Zadaće rudarenja podataka....3
2.2.1 Deskripcija....3
2.2.2 Klasifikacija....3
2.2.3 Estimacija 3
2.2.4 Predikcija....4
2.2.5 Klasteriranje....4
2.2.6 Asocijacija 4
3. Faze rudarenja podataka 5
3.1 Definicija poslovnog problema 7
3.2 Priprema podataka 7
3.2.1 Određivanje potrebnih podataka....7
3.2.2 Transformacija podataka 7
3.2.3 Uzorkovanje podataka....8
3.2.4 Vrednovanje podataka....8
3.3 Modeliranje 8
3.4 Implementacija 9
4. Pregled prethodnih istraživanja....10
5. Opis i priprema podataka....29
5.1 Opis problema 29
5.2 Opis podataka 29
6. Opis korištenih metoda rudarenja podataka....31
6.1 Klasteriranje 31
6.2 Neuronska mreža 32
6.3 Stablo odlučivanja....32
7. Klaster analiza 34
8. Stablo odlučivanja 37
9. Neuronska mreža....39
10. Opis i diskusija rezultata....41
11. Zaključak....42
12. Literatura....43
13. Popis tablica i slika....44
13.1 Popis slika....44
13.2 Popis tablica....45

1. Uvod

Elektroničko učenje predstavlja širok pojam koji podrazumijeva korištenje informacijskih tehnologija prilikom obrazovanja. E-učenje se javlja u današnje vrijeme sve više, pogotovo na fakultetima. Putem e-učenje, održava se sve više konferencija, razna online obrazovanja zaposlenika, različiti sustavi za učenje stranih jezika i slično. E- učenje se odvija na različitom mjestu od mjesta predavanja, kao što je na primjer studentski dom ili pak vlastita kuća. Također i zahtijeva posebne tehnike planiranja kolegija i načina komuniciranja.

Neke od prednosti e-učenja su:

- Smanjenje troškova jer se smanjuju izdaci za putovanje
- Dostupnost nastavnih materijala 24 sata na dan
- Pristup najnovijim informacijama
- Polaznici mogu izabrati lekcija koje odgovaraju njihovim mogućnostima i potrebama
- Razvoj ICT kompetencija

Što se tiče o samim podacima, njihov broj je sve veći i veći i on neprekidno raste, a samim time i važnost tih podataka. Podaci se mogu prikupljati na puno različitih načina, te različitih oblika. Naravno kako se prikupljaju, tako se i obrađuju na različite načine. Podaci mogu biti dostupni javno, tako da im može pristupiti bilo tko, a mogu biti i privatni, koji se koriste za posebne svrhe. Također, podaci mogu biti besplatni za preuzeti, a mogu i biti dostupni uz plaćanje određene naknade za njihovo korištenje. Dobra stvar rastu broja i količine podataka je to da imamo više dostupnih informacija koje mogu pomoći na poslu, kod učenja ili jednostavno ako se želimo više informirati o nečemu što nas zanima. Naravno, ne može uvijek sve bit dobro pa tako i onda se javljaju problemi sa velikom količinom podataka koju treba obraditi da bi lakše došli do željenih rezultata. Zbog toga razvile su se nove metode, tehnike i alati koji nam pomažu u obrađivanju i istraživanju tih podataka. Pa tako postoji jedna proces koji se zove rudarenje podataka, koja podrazumijeva istraživanje i analizu velikih količina podataka, kako bi dobili neke smislene rezultate i uzorke, pa tako i onda došli do određenih zaključaka, i dobivanja novih znanja. To znači da rudarenjem nad podacima koje smo prije prikupili, možemo dobiti informacije od kojih se mogu napraviti nova znanja. [7]

2. Rudarenje podataka

U svakodnevnom životu, organizacije posluju sa velikom količinom podataka iz kojih je dosta teško izdvojiti dobru odluku te ono što je bitno. Zbog toga, te organizacije su se počele baviti rudarenjem podataka, kako bi lakše pronašle bitne informacije u gomili podataka, pomoću metoda za rudarenje podataka. Rudarenje podataka nam predstavlja sortiranje, grupiranje, te organiziranje podataka kako bi lakše mogli dobiti kvalitetne informacije koje su nam potrebne.

2.1 Definicija rudarenja podataka

Druga definicija za rudarenje podataka je otkrivanje znanja u bazama podataka (eng. Knowledge discovery in databases – KDD) i to predstavlja proces otkrivanja uzoraka na većem skupu podataka koji mogu biti korisni poduzećima, odnosno mogu pridonijeti poduzećima da mogu donijeti kvalitetnije poslovne odluke koje mogu biti bitne za daljnje uspješno poslovanje poduzeća. Postoje dva načina na koja možemo shvatiti rudarenje podataka, a to su širi i uži način. Uži način nam predstavlja, odnosno podrazumijeva neku posebnu fazu obrade podataka, dok šire shvaćanje podrazumijeva proces otkrivanja znanja i informacija iz nama dostupnih podataka. Postoji više vrsta podataka koji se mogu obrađivati, pa tako podaci mogu biti tekstualni, nestrukturirani, organizirani u vremenske sesije ili podaci koji su organizirani u baze podataka i to je ujedno najčešći oblik za obrađivanje podataka. Još neki pojmovi koji se javljaju uz rudarenje podataka su i ekstrakcija znanja, analiza obrazaca, žetva informacija, te cijeđenje podataka. (Srića, 2018.; Garača i Jadrić, 2011.)

U današnje vrijeme, rudarenje podataka je postalo od velike značajnosti i koristi jer bez toga, velike organizacije ne bi mogle donijeti kvalitetne odluke prilikom poslovanja. Rudarenje podataka ima veliku primjernu i u drugim djelatnostima kao što su komercijalni sustavi da bi istražili na koji način kupci odlučuju što će kupiti, što ih najviše privlači i na temelju toga poboljšali i povećali prodaju robe. Rudarenje podataka ima i primjenu u vojsci, za otkrivanje neprijatelje, njihovih baza, donošenje odluka o napadu i slično. Za sve to nam je potrebna moderna tehnologija, nešto poput super računala koja su izrazito snažna i koja mogu obrađivati velike količine podataka. [7]

„Rudarenje podataka je istraživanje i analiza velikih količina podataka u nastojanju otkrivanja smislenih obrazaca i pravila.“ (Berry i Linoff, 2004.)“

2.2 Zadaće rudarenja podataka

Postoji šest zadaća rudarenja podataka i to su:

1. Deskripcija
2. Klasifikacija
3. Estimacija
4. Predikcija
5. Klasteriranje
6. Asocijacija

2.2.1 Deskripcija

Deskripcija predstavlja postupak jednostavnog opisivanja nekih činjenica, procesa ili predmeta u prirodi, te njihovih potvrđivanja odnosa i veza ali bez znanstvenih tumačenja ili objašnjavanja. Ta se metoda koristi u samoj početnoj fazi istraživanja, te ona može imati još veću vrijednost ako je to jednostavno opisivanje povezano na neki način s važnijim obilježjima drugih opisanih činjenica, predmeta ili procesa, te uzročnih veza i odnosa. [9]

2.2.2 Klasifikacija

To je najstarija i najjednostavnija znanstvena metoda. One je sistematska te potpuna podjela općeg pojma na posebne, odnosno u okviru samog opsega pojma. [9]

Ciljna varijabla je vrijednost koja se predviđa, i ona kod metode klasifikacije predstavlja određenu kategoriju. Na primjer, hoće li klijent vratiti kredit ili ne. Za klasifikaciju se najviše koriste stablo odlučivanja, regresija i neuronske mreže. [10]

2.2.3 Estimacija

Drugi naziv za estimaciju je regresija. Ona je vrlo slična kao i klasifikacija, samo što kod estimacije, ciljna varijabla nije kategorijska nego numerička. Modeli se izgrađuju uz pomoć kompletnih zapisa, tj. onih zapisa koji samo sadržavaju vrijednosti ciljne varijable uz prediktorske varijable. Onda se za nova promatranja procjenjuje dalje nova vrijednost ciljne varijable opet na temelju vrijednosti prediktorskih varijabli. Na primjer, to može biti procjena prosječne ocjene studenta na diplomskom studiju na temelju prosječne ocjene tog studenta na preddiplomskom studiju.[11]

2.2.4 Predikcija

Predikcija je slična kao i klasifikacija i estimacija, ali u kod predikcije, ciljna varijabla predstavlja zapravo buduću vrijednost. Metode ili tehnike koje se koriste kod estimacija i klasifikacije se također mogu koristiti za predikciju ali pod pravim okolnostima. Neke od tih poznatijih metoda i tehnika su jednostavna linearna regresija i korelacija, višestruka regresija, te neuronske mreže i stablo odlučivanja. Primjer predikcije može biti predviđanje postotka rasta smrtnih slučajeva u prometu u sljedećoj godini ako se poveća ograničenje brzine. [11]

2.2.5 Klasteriranje

Klasteriranje predstavlja grupiranje zapisa, opservacija ili slučajeva u klastere sličnih objekata. Klaster je skup zapisa koji su međusobno slični, ali su različiti od zapisa u drugim klasterima. Razlika između klasteriranja i klasifikacije je u tome što za klasteriranje ne postoji ciljna varijabla. Kod klasteriranja nema ciljne varijable jer ono ne služi za klasificiranje, procjenu ili predviđanje vrijednosti ciljne varijable. Algoritmi koji se koriste kod klasteriranja, pokušavaju segmentirati cijeli skup podataka u relativno homogene podgrupe ili klastere, ali tako da sličnost zapisa u klasteru bude maksimalna, a sličnost zapisa izvan tog klastera bude minimalna. Klasteriranje se vrlo često provodi kao početni korak u KDD procesu, pa se rezultirajući klasteri koriste onda kao dodatni ulazni atributi koji se koriste za daljnje tehnike rudarenja podataka. Primjer klasteriranja može biti smanjenje broja dimenzija skupa podataka sa stotinama dimenzija. [1]

2.2.6 Asocijacija

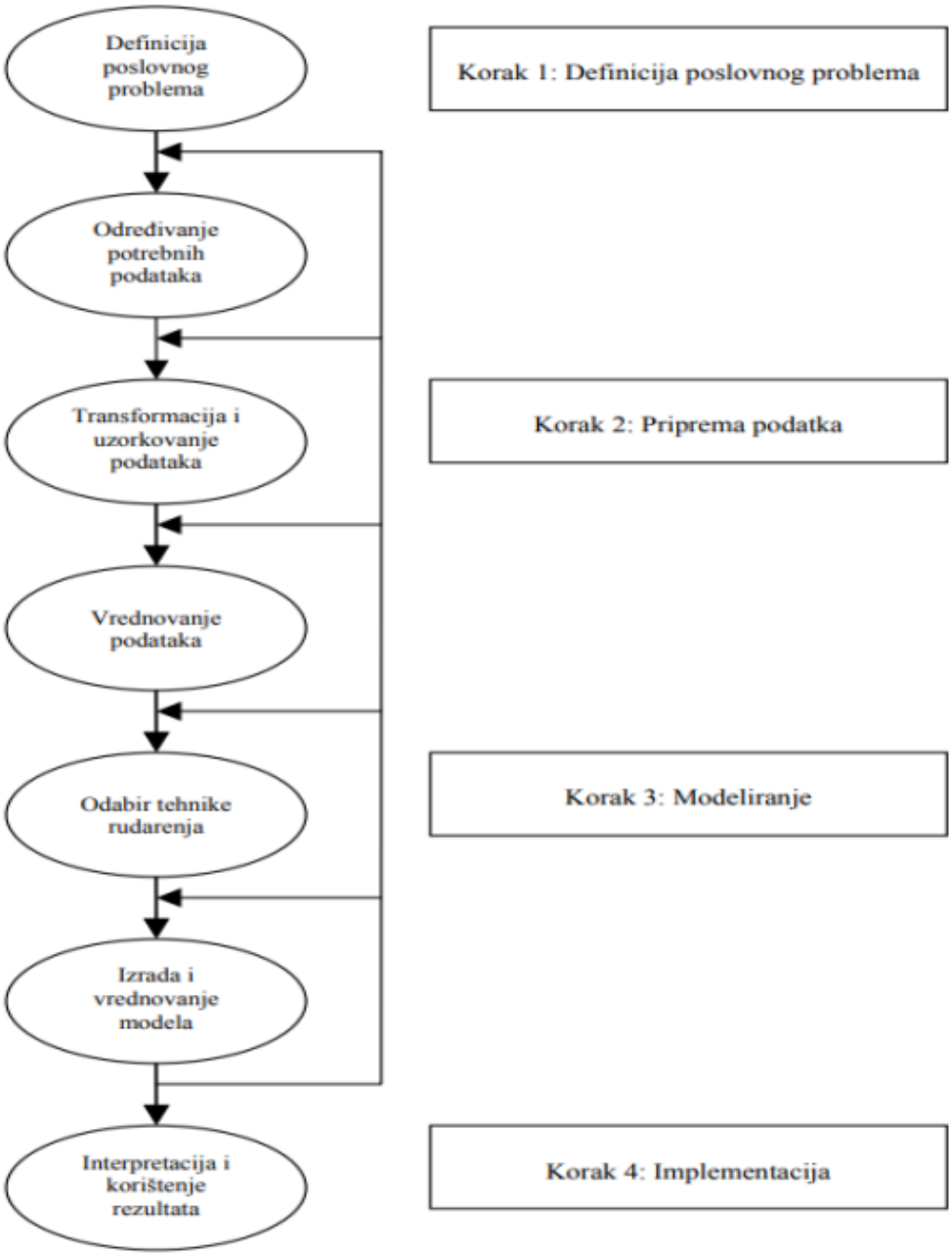
Glavni cilj ove zadaće rudarenja podataka je otkriti koji su atributi međusobno povezani, tj. otkriti pravila koja kvantificiraju odnos između dva ili više atributa. Vrlo je popularna i često korištena u poslovnom svijetu gdje je poznata pod nazivom analiza afiniteta ili analiza potrošačke košare. Pravila asocijacije su „Ako *uvjet*, tada *posljedica*“. Recimo, neka trgovina može otkriti da od 500 kupaca koji kupuju subotom ujutro, 100 ih kupi mlijeko, i od tih 100, 20 ih je kupilo žvakače gume. Prema pravilu asocijacije „Ako kupuje mlijeko, kupuje žvakače“ sa značajem od 100/500 = 20% te pouzdanost od 20/100 = 20%. Još neki primjer asocijacije bi mogao biti istraživanje postotka odnosno proporcije djece čiji roditelji im čitaju, a i sami su dobri čitači. [7]

3. Faze rudarenja podataka

Bitno je napomenuti da je rudarenje podataka više umjetnost nego znanost. Potrebno je još istaknuti da nema, tj. ne postoji nekakav posebni recept ili „kuharica“ po kojoj bi mogli uspješno rudariti podatke da bi dobili nekakve korisne informacije. No, za to postoje određeni koraci koji mogu pomoći da bi se vjerojatnost uspjeha povećala (Baragoin et.al., 2001.). Ti koraci su prikazani na slici 1.

U prvom koraku moramo definirati poslovni problem koji želimo analizirati. Sljedeći korak je priprema podataka, koja se sastoji potrebnih podataka, transformacije i uzorkovanja i na kraju vrednovanja podataka. Treći korak je modeliranje koji se sastoji od odabira metode rudarenja te izrade i vrednovanja modela. Zadnji korak je implementacija koja obuhvaća interpretaciju i korištenje rezultata. [7]

Sam taj proces rudarenje podataka je iterativan. To znači da se u bilo kom trenutku možemo vratiti na neki od prošlih koraka ako na primjer kod odabira tehnike za rudarenje podataka nismo dobro odabrali podatke koje ćemo koristiti, pa se onda možemo vratiti na korak za odabir potrebnih podataka. Ovakav primjera vraćanja na neki prethodni korak je više pravilo nego nekakav izuzetak jer kod rudarenja podatka je jako važno dobro odabrati i pripremiti podatke te definirati problem, a to je vrlo teško od prve napraviti, pa onda moramo u većini slučajeva ponovo ponavljati neke korake. To je s jedne strane dobro jer se onda tokom tog procesa rudarenja podataka povećava naše znanje o problemu i podacima koje analiziramo, te je takva ponovljena verzija u većini slučajeva bolja od originala. U nastavku ćemo detaljnije opisati te korake. [10]



Slika 1: Faze rudarenja podataka [7]

3.1 Definicija poslovnog problema

U prvom koraku je potrebno definirati poslovni problem, najbolje u obliku pitanja na koje ćemo moći odgovoriti na kraju, tj. po završetku procesa. Da bi uspješno proveli ovaj korak, trebali bi prvo provest analizu područja gdje je već rudarenje podataka korišteno. Zatim moramo odrediti koje osobe ćemo pridružiti za sudjelovanje na projektu rudarenja podacima. Najčešće je to manja skupina ljudi koja se sastoji od specijalista za rudarenje podataka koji dobro poznaje metode otkrivanja znanja, pa informatičar koji ima iskustva na području za rad s bazama podataka i skladištima podataka i još jedan stručnjak iz organizacije koji ima iskustva sa potencijalnim primjerima rada u poslovanju. Naravno potrebna je i jedna osoba, menadžer, koja će voditi ovaj tim, ali ne da direktno radi sa timom nego da im pomaže i rješavanju eventualnih problema. [7]

3.2 Priprema podataka

Ova faza se sastoji od nekoliko koraka. To su određivanje potrebnih podataka, transformaciju, uzorkovanje, te vrednovanje podataka. Prema nekim autorima, ova faza je vremenski najduža te ona oduzima od 60% do 90% ukupnog vremena samog rudarenja podataka. Podaci koji se koriste kod rudarenja podatka mogu biti pohranjeni u različitim oblicima, a to su relacijske baze podataka ili skladišta podataka. To mogu biti recimo neki operacionalni sustavi kao što su bankomati, web serveri i slično. Ljudi koji rade na tom projektu rudarenja podataka, moraju zajedno odrediti koji podaci će im biti potrebni za izradu modela i koji su nepotrebni. [10]

3.2.1 Određivanje potrebnih podataka

Određivanje potrebnih podataka je prvi korak pripreme podataka koji će se koristiti za izradu modela, i u njemu određujemo koje varijable ćemo izbaciti iz analize, i koja varijabla će nam biti ciljna tj. zavisna. Na primjer, kod analize kredita, ciljna varijabla nam pokazuje je li klijent vratio kredit ili nije. Na kraju, kao konačan rezultat određivanja potrebnih podataka, dobijemo popis varijabli koje ćemo koristiti kod izrade modela. [10]

3.2.2 Transformacija podataka

Transformacija podataka je drugi korak pripreme podataka. Ona se odnosi na pretvorbu podatka u oblik koji će nam biti pogodan za rudarenje podataka. Podaci za

rudarenje moraju biti u tablicama, tako da se u stupcima nalaze varijable, a u recima opažanja. U ovom koraku se podrazumijevaju operacije s podacima poput agregacije, grupiranja, selekcije, filtriranja i spajanja. [7]

3.2.3 Uzorkovanje podataka

Sljedeći korak je uzorkovanje podataka. Ovdje nam je potrebno postaviti pitanje: Koliko je podatak dovoljno? Naravno, na to pitanje možemo teško odgovoriti jer broj potrebnih podataka ovisi o algoritmu. Baze podataka sadrže velike količine podataka, glavni zadatak uzorkovanja je smanjiti tu veliku količinu podataka tako da bude prikladna za daljnju izradu modela. Kod izrade stabla odlučivanja, potrebno je oko dvije do tri tisuće podataka, dok za neuronske mreže je potrebno puno više podataka. Podaci za uzorak se najčešće odabiru slučajnim odabirom i kada se odabere uzorak, potrebno ga je podijeliti na nekoliko dijelova, na dio podataka za izradu modela te dio za testiranje. Ovakvim pristupom rudarenju podataka se provjerava njegova efikasnost nad podacima koji nisu izabrani za njegovu izradu. [7]

3.2.4 Vrednovanje podataka

Posljednji korak kod pripreme podataka je vrednovanje podataka. Potrebno je analizirati postojanje netipičnih vrijednosti te prljavih podataka. Netipične vrijednosti postoje gotovo u svakoj bazi podataka kao što su nepostojeće, netočne ili nejasne vrijednosti i zbog toga je potrebno odlučiti što napraviti sa tim netipičnim vrijednostima. Razlog tome je prelazak iz jedna baze u drugu, ali može se javit kao posljedica pogrešnog unosa u računalo. [7]

3.3 Modeliranje

U fazi modeliranja se odabire metoda rudarenja podataka te se mora izraditi i vrednovati model. U samom procesu postoji nekoliko metoda, a to su: statistika baze i skladišta podataka, umjetna inteligencija te vizualizacija. Metode rudarenja podatka se mogu podijeliti u tri kategorije: otkrivanje, klasifikacija te predviđanje.[10]

Metode za klasifikaciju varijabli se koriste za predviđanje kategorije, te se najčešće koriste stablo odlučivanja, logit regresija ili neutralne mreže. Metode otkrivanja se koriste tamo gdje se traži pravilnost u podacima o kojima nemamo prethodnog znanja o njihovom obliku. Postoje mnoge metode otkrivanja pravilnosti u podacima, a najpoznatije su metode segmentiranja (engl. Cluster analysisi) i asocijativna pravila. Metode za predviđanje vrijednosti varijable se koriste kod predviđanja vrijednosti ali numeričkih oblika. Metode koje

se najčešće koriste su neuronske mreže, linearna regresija kao i metode vremenskih serija. [10]

Ovisno o cilju analize, ovisi sam odabir metode. Na primjer, ako imamo za cilj otkrivanje novih segmenata na tržištu, onda će se koristiti neka od metoda otkrivanja. Kod rudarenja podataka, najviše korištene metode su metode klasifikacije, dok metode koje se najmanje koriste su metode predviđanja vrijednosti. [10]

U fazi modeliranja se najčešće koristi više metoda zbog same provjere. Kada se usporede rezultati na konkretnom uzorku podataka, onda se odabire metoda rudarenja podatka. Naravno, bez pravog softvera za rudarenje podataka, ne bi mogli primjenjivati ove metode. Na internetu ima dosta besplatnih softvera koji nam mogu poslužiti kod procesa rudarenja podataka. Softveri mogu sadržavati više metoda ili biti specijalizirani samo za jednu metodu. [10]

Kada primijenimo neku od tih metoda, onda možemo vrednovati rezultate. Naravno, ne smijemo zaboraviti da podaci koji su izabrani u uzorku, su podijeljeni u dva dijela . dio za izradu modela i dio podataka za testiranje samog modela. [10]

3.4 Implementacija

Implementacija je posljednji korak procesa rudarenja podataka. Vrlo je važno da sami rezultati modela budu u jednostavnom i razumljivom obliku kao što su grafikoni ili tablice. To je najčešće zbog toga jer korisnik koji čita rezultate nije nekakav stručnjak nego je obična osoba koja će moći pročitati i razumjeti rezultate modela. Kao što je rečeno, sam taj proces rudarenja podataka je iterativan, te se je moguće vratiti na bilo koji prethodni korak, ako smo negdje pogriješili ili nešto nije u redu. Normalna je stvar da će se pojaviti neke greške u samim koracima procesa rudarenja podatka, i to nam ne bi smjelo predstavljati problem nego ćemo time se vratiti nazad i još bolje razumjeti naše podatke, pa će i sami rezultati kasnije biti kvalitetniji. [7]

4. Pregled prethodnih istraživanja

Filozofski fakultet u Zagrebu je proveo istraživanje putem ankete u svibnju 2007. godine s ciljem mjerenja kvalitete integriranog sustava za e-učenje (Omega). Upitnik se sastojao od 106 varijabli koju su se odnosile na važne aspekte za analiziranje stavova prema e-učenju, te su napravili rezultate prema uzorku od 148 studenata Filozofskog fakulteta Sveučilišta u Zagrebu. Uzorak se sastojao od 74% studentica i 26% studenata što odgovara njihovoj spolnoj strukturi na razini fakulteta. Najzastupljeniji su bili studenti Odsjeka za anglistiku (34%), Odsjeka za sociologiju (29%) i Odsjeka za informacijske znanosti (20%). Te podatke su obradili u programu za statističku obradu podataka *Statistical Package for Social Sciences* verzija 13 (SPSS, Chicago IL). Rezultati su pokazali da većina studenta podupire ideje potrebe povećane integracije Omege u tradicionalnu nastavu, te da Omega poboljšava kvalitetu nastave uz korištenje e-učenja. Velik broj studenata se je još složio da je Omega koristan dodatak nastavi na fakultetu te da olakšava distribuciju nastavnih materijala. [2]

Još jedno istraživanje je provedeno na Agronomskom fakultetu gdje je promatrano korištenje sustava za e-učenje na uspješnost studenata Agronomskog fakulteta. Njihov cilj je bio dokazati da korištenje sustava za e-učenje omogućuje poboljšanje kvalitete obrazovanja, odnosno utječe na povećanje uspješnosti obrazovanja. Korištena je statističko komparativna analiza na reprezentativnom uzorku uz korištenje anketnog upitnika kao instrumenta istraživanja. Anketirali su se nastavnici koji koriste sustave za e-učenje u nastavi i ocjenjivanju, te su se tako tražili rezultati o prosječnoj prolaznosti studenata prije korištenja sustava za e-učenje i nakon. Analizu su proveli na dvije skupine uzoraka: za razdoblje od 2007. do 2009. godine i od 2010. do 2012. godine. Ti uzorci predstavljaju podatke o prolaznosti studenata za razdoblje od 3 godine, kad nije korišten sustav za e-učenje, te za razdoblje od 3 godine kada je sustav za e-učenje. Prvi uzorak prije korištenja sustava za e-učenje je veličine 341. Veličine uzoraka po godinama su: 2007. godine: 84; 2008. godine: 131; 2009. godine: 126. Drugi uzorak nakon korištenje sustava za e-učenje je veličine 396. Veličine uzorka po godinama su: 2010. godine: 143; 2011. godine:136; 2012. godine: 117. [2] Rezultati su sljedeći, prikazani u tablicama:

Ocjena	Broj studenata	Struktura
2	68	19,9 %
3	136	39,9 %
4	105	30,8 %
5	32	9,4 5%
UKUPNO	341	100 %

Tablica 1: Ocjene studenata prije uvođenja i primjene sustava za e-učenje [2]

Iz tablice 1 je vidljivo da je najčešća ocjena prije korištenja sustava za e-učenje bila ocjena 3.

Ocjena	Broj studenata	Struktura
2	20	5,1 %
3	131	33,1 %
4	206	52,0 %
5	39	9,8 %
UKUPNO	396	100 %

Tablica 2: Ocjene studenata nakon uvođenja i primjene sustava za e-učenje [2]

Iz ovih tablica se može zaključiti da je više od polovice studenata (52%) ostvarilo ocjenu 4, isto tako je znatno smanjen broj studenata sa ocjenom 2 (s 19,9% na 5,1%)

Iz ovog istraživanja se može zaključiti da je poboljšanje ocjena statistički značajno. Uvođenje sustava za e-učenje u nastavu i ocjenjivanje je imalo pozitivan utjecaj na kvalitetu visokoobrazovnog procesa, odnosno na ocjene. [3]

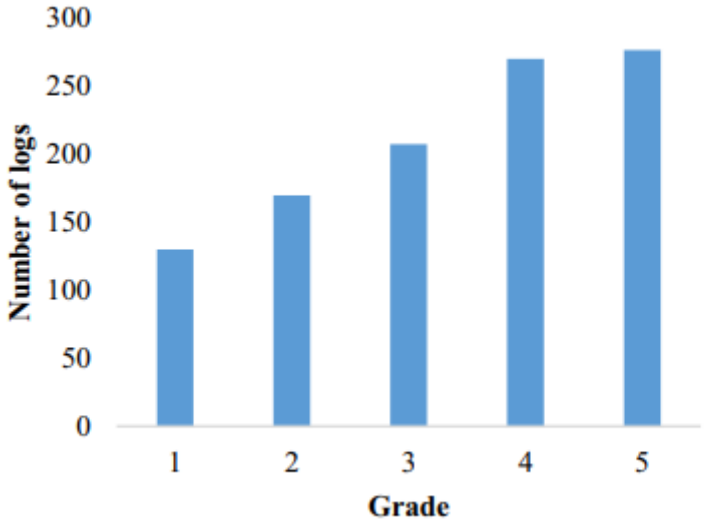
Sljedeće istraživanje je bazirano na praćenju ponašanja studenata i njihovog uspjeha baziranog prema logiranju na Moodle. LMS (eng. Learning Managment Systems – LMS) u današnje vrijeme, osim što služe kao podrška za lice u lice učenje, podržavaju online predavanja te razne online studijske programe. Online učenje je u većini slučajeva vezano za područja visokog školstva. Studenti mogu diplomirati bez da fizički prisustvuju nastavi na fakultetu. Najpoznatiji LMS u hrvatskoj, tj. na Sveučilištu u Zagrebu je Moodle (koji je dostupan kao Merlin ostalim visokom ustanovama u hrvatskoj). Ovisno o broju i intenzitetu online aktivnosti koje podržava Moodle, tečajevi na Sveučilištu u Zagrebu mogu biti

podijeljeni u 3 kategorije. Najveća kategorija, razina 3, zahtijeva da svi materijali potrebni za učenje budu dostupni online. [14]

Trendovi u hrvatskoj su doprinijeli implementaciji jednog koncepta koji se zove analitika učenja (eng. Learning Analytics - LA). Kada student pristupi LMS sa svojim računom, onda se stvara digitalni profil koji je spremljen u LMS log datoteke. Glavna ideja LA je analizirati neobrađene podatke iz log datoteka i generirati nova znanja o ponašanju studenata. Npr. ako su studenti koji su pristupili materijalima za učenje dan prije kolokvija, pali, onda je potrebno napraviti neke aktivnosti koje će utjecati na to da da pristupe materijalima za učenje ranije. [14]

U ovom istraživanju postavljena su dva pitanja: P1: U kojoj su mjeri pojedine varijable izvedene iz log podataka, pouzdan prediktor akademskog uspjeha? P2: Koja je razina sličnosti u korištenju LMS-a između spolova?. Istraživanje je provedeno za predmet koji se zove „Business Decision Making“. Sveukupno je registrirano 73 studenata na Moodle, kroz 180-minutna predavanja i seminare koji su se održavali svaku srijedu. [14]

Slika 2 prikazuje graf omjera ocjena i broja logiranja studenata na predmet. Možemo zaključiti da studenti koji su se više puta ulogirali imaju bolje ocjene, 4 i 5.

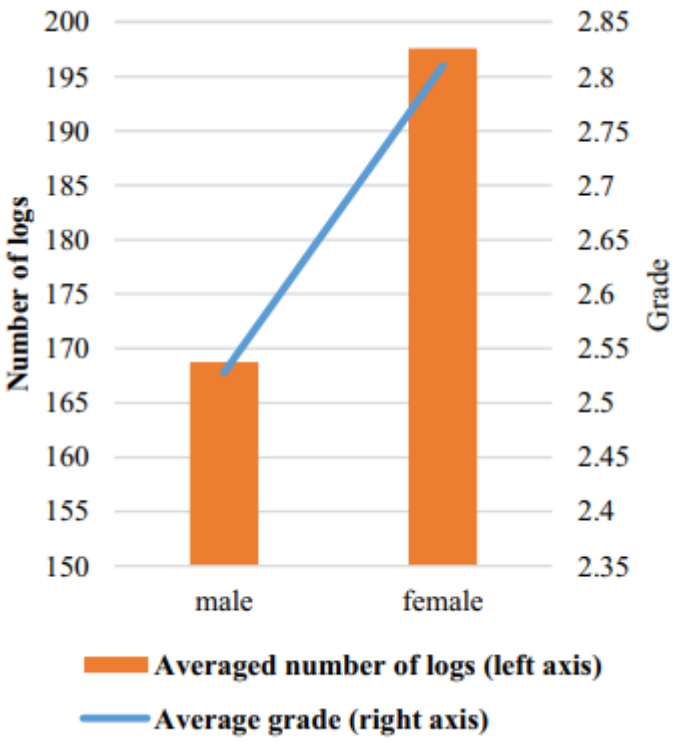


Slika 2: Frekvencija broja logiranja prema ocjenama [14]

Varijabla	*Ocjena*	*Korištenje datoteka*	*Korištenje foruma*	*Korištenje linkova*	*Upolad-anje zadataka*
Ocjena	-	0.35*	0.18	-0.01	0.08
Korištenje datoteka	0.35*	-	0.41	0.05	0.19
Korištenje foruma	0.18	0.41*	-	0.13	0.38*
Korištenje linkova	-0.01	0.05	0.13	-	0.09
Upolad-anje zadataka	0.08	0.19	0.38*	0.09	-

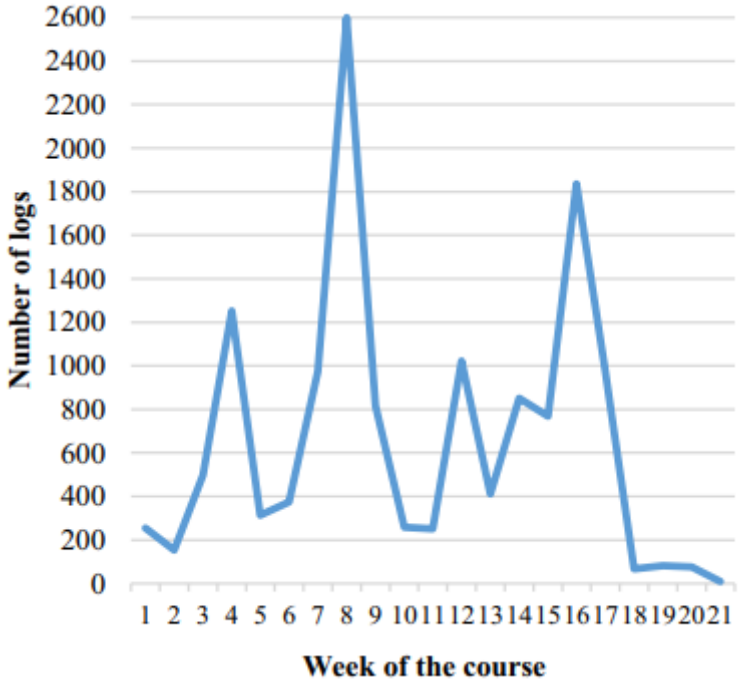
Tablica 3: Korelacija između varijabli (*p<0.05) [14]

U tablici 3 možemo vidjeti korelaciju između pojedinih varijabli kolegija, mjerenog prema ocjenama i trudu u korištenju datoteka, foruma, i korištenju dodatnih linkova te upload-u pojedinih zadataka. Velika korelacija je između ocjena i korištenju datoteka, ona je pozitivna što znači da studenti sa većom frekvencijom korištenja datoteka imaju veće ocjene. Naravno postoji i nedostatak korištenja ostalih aktivnosti koji utječu na ocjenu. [14]



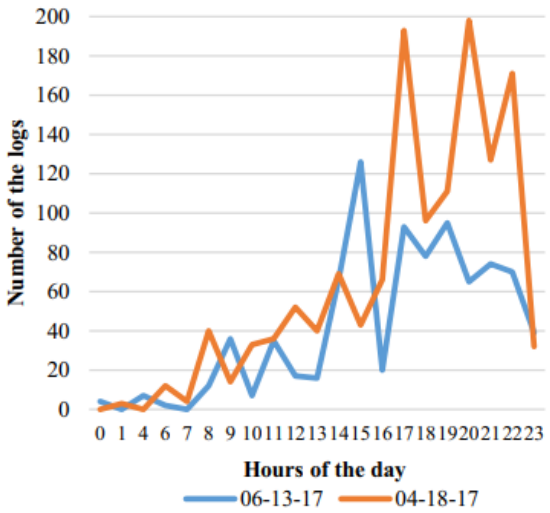
Slika 3: Analiza po spolu [14]

Na slici 3, imamo prikazan graf koji nam govori da studentice imaju veći broj logiranja na Moodle nego studenti. Razlika je također vidljiva u prosječnim ocjenama. Studentice imaju veću prosječnu ocjenu nego studenti. [14]



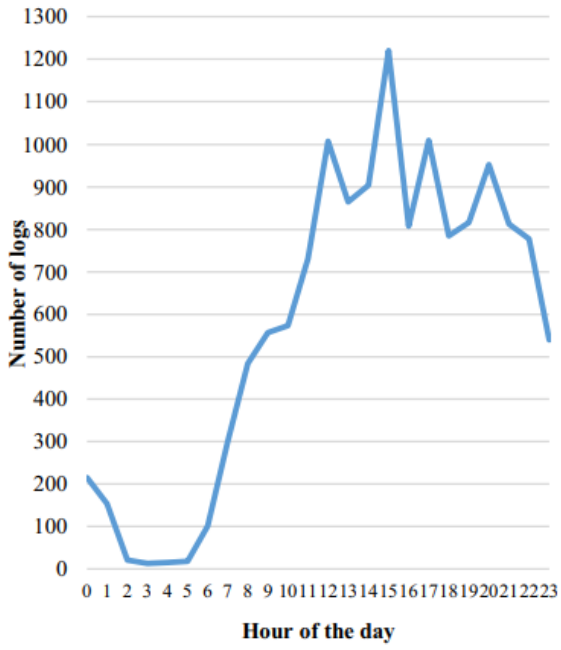
Slika 4: Omjer logiranja po tjednima [14]

Na slici 4 je prikazan graf koji prikazuje učestalost broja logiranja na Moodle prema tjednima predavanja. Možemo vidjeti da se studenti najviše prijavljuju u 8 i 16 tjednu. A u tim tjednima su zapravo održani kolokviji pa je zbog toga i onda najviše logiranja zabilježenih na Moodle. [14]



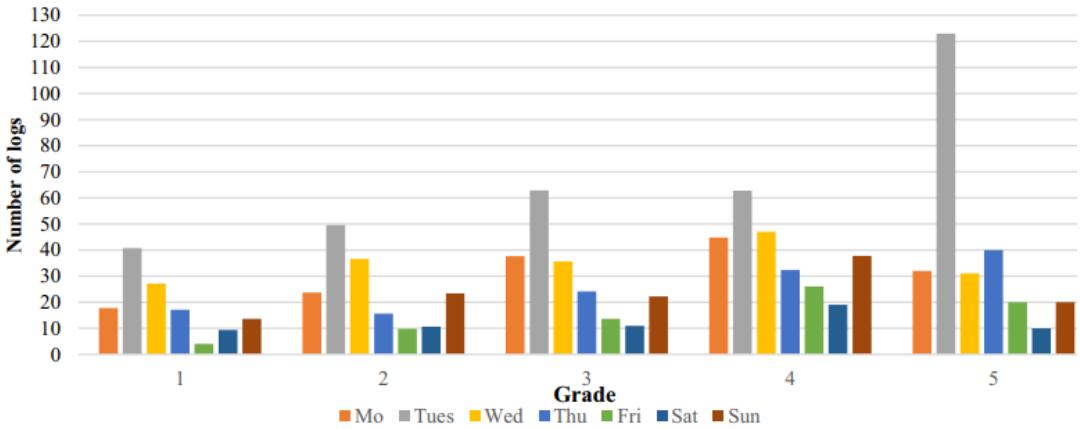
Slika 5: Omjer broja logiranja prema danima prije kolokvija [14]

Na slici 5 možemo vidjeti graf koji prikazuje omjer broja logiranja prema danima prije kolokvija. Studenti su se najviše puta ulogirali između 17:00 i 20:00 sati na dan prije prvog kolokvija, te između 15:00 i 16:00 sati na dan prije drugog testa. Tijekom tih sati, studenti su skidali materijale za učenje i počeli učiti. Vidljivo je još da je aktivnost na predmetu više primjetna u kasnim satima za prvi kolkovij. [14]



Slika 6: Omjer logiranja prema satima u danu [14]

Vremenska analiza sa slike 6 nam pokazuje da su se studeni najviše puta ulogirali poslije 11:00 sati. Taj broj se prema večeri smanjivao, a poslije ponoći nije bilo aktivnosti, međutim to ne znači da studenti nisu bili aktivni na predmetu, nego su možda radili aktivnosti bez prisutsva na Moodle-u. [14]



Slika 7: Omjer broja logiranja prema danu u tjednu sa postignutom ocjenom [14]

Na slici 7 je prikazan graf analize aktivnosti studenata prema postignutim ocjenama i danima u tjednu. Na iznenađenje, studenti sa većom ocjenom su većinu vremena bili aktivni na Moodle-u dan prije predavanja, seminara i kolokvija. [14]

Iz ovog istraživanja možemo zaključiti da su studentice aktivnije i uspješnije na samom kolegiju. Studenti su najaktivniji u tjednima kolokvija, tj. najviše na dan prije samo testa. Vidljivo je da su studenti najaktivniji u zadnjim „trenutcima“, prije samo roka pojedinih aktivnosti, no to ne možemo reći sa sigurnošću jer jer istraživanje provedeno na samo jednom predmetu, te nad studentima koji studiraju informatiku. [14]

Sljedeće istraživanje je vezano uz usporedbu strojnog učenja na pristup LMS podacima. LMS smo već ranije spomenuli pa znamo što je to dok strojno učenje predstavlja različite pristupe klasifikaciji i regresijskim modelima. Ideja ovog istraživanja je istražiti sposobnosti strojnog učenja na pristup obrazovnim podacima. [15]

Koriste se dva seta podataka sa Sveučilišta u Zagrebu, Fakulteta Organizacije i Informatike. Jedan je sa Moodle-a koji se sastoji od log datoteka pristupa materijalima kolegija, dok se drugi set podataka sastoji od ocjena koji su studenti postigli na predmetu. [15]

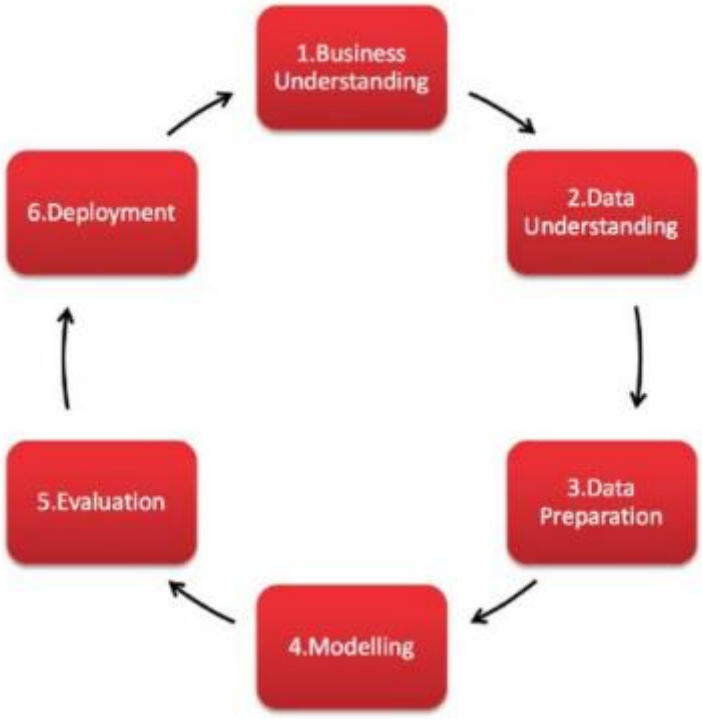
Naziv varijable	Opis
Prolaz	Varijabla koja nam govori ako je student položio kolegij.
Dodatni kviz	Kviz za dodatne bodove.
Blic 1	Prva nenajavljena kratka provjera.
Blic 2	Druga nenajavljena kratka provjera.
Blic 3	Treća nenajavljena kratka provjera.
Blic 4	Četvrta nenajavljena kratka provjera.
Kolokvij iz excela	Kolokvij iz MS excela na labosima
Kolokvij iz access-a	Kolokvij iz MS accessa na labosima.
Ponavljanje kolokvija teorija	Ponavljanje kolokvija iz cijelog semestra.
Dolasci na labose	Dolasci na laboratorijske vježbe.
Dolasci na predavanja	Dolasci na predavanja.
Samoprocjena 1	Samoprovjera znanja 1
Samoprocjena 2	Samoprovjera znanja 2
Samoprocjena 3	Samoprovjera znanja 3
Samoprocjena 4	Samoprovjera znanja 4
Samoprocjena 5	Samoprovjera znanja 5
Samoprocjena 6	Samoprovjera znanja 6
Samoprocjena 7	Samoprovjera znanja 7
Samoprocjena 8	Samoprovjera znanja 8
Samoprocjena 9	Samoprovjera znanja 9
Samoprocjena 10	Samoprovjera znanja 10
Samoprocjena 11	Samoprovjera znanja 11
Samoprocjena 12	Samoprovjera znanja 12
Mapa	Predstavlja mapu za određene datoteke.

Datoteka	Predstavlja datoteke o dodatnim materijalima vezanim za kolegij.
Forum	Forum za međusobnu komunikaciju sa drugim studentima i profesorima.
Izvještaj o studentu	Predstavlja izvještaj o pojedinom studentu koji prikazuje njegov napredak na kolegiju.
Lekcija	Predstavlja korištenje materijala za učenje pojedinih lekcija.
Odabir grupe	Odabir laboratorijske grupe.
Postavljanje datoteke	Postavljanje datoteke na Moodle, npr. Razne zadaće.
Poveznica	Poveznice na druge materijale vezane za kolegij.
Pregledni izvještaj o studentu	Predstavlja pregledni izvještaj o pojedinom studentu.
Stranica	Predstavlja razne stranice kao što su literatura, vodiči i slično.
Sustav	Predstavlja pristup kolegiju.
Test	Predstavlja izvještaj o testu.
Zadaća	Predstavlja pristup dijelu predviđenom za zadaće.

Tablica 4: Opis varijabli [15]

U ovom istraživanju se koristi međuindustrijski proces za rudarenje podatka (CRISP DM) standard. On se sastoji od šest faza prikazanih na slici 8. Prva faza se fokusira na zadatke istraživanja i definiciju problema rudarenja podataka. Glavni zadatak je istražiti performanse algoritme strojnog učenja na podatke LMS-a. Faza razumijevanja podataka se sastoji od: tipova varijabla, njihovih karakteristika te distribucija. Priprema podataka se odnosi na 70% do 90% procesa rudarenja podataka. Smanjenje značajki je najvažniji dio pripreme podataka. U ovom istraživanju je korištena faktorska analiza da se napravi smanjenje značajki podataka. Modeliranje se sastoji od razvoja i procjene modela. U ovom koraku se odabire metoda modeliranja. Uzete su četiri metode: neuronske mreže, stablo odlučivanja, k-najbliži susjedi, te Naive Bayes klasifikator. Evaluacija nam govori koliko dobro se model ponaša nad testiranim podacima. Za evaluaciju su korištene dvije metode mjerenja

performansi: root-mean-square, te vrijeme izvršavanja algoritama strojnog učenja. Posljednja faza nam govori kako treba koristiti rezultate modeliranja. [15]



Slika 8: CRISP DM metodologija [15]

Za ovo istraživanje je prvo napravljena analiza faktora. Analizom faktora se napravila pretvorba ulaznih značajki i nove značajke (faktore) koji nisu u korelaciji. Prva četriri ekstrahirana faktora predstavljaju 61.08% totalne varijance. Faktorskom analizom se ciljalo na poboljšanje preformansi algoritama strojnog učenja, isključujući značajke čija prisutnost ometa točnost prediktivnih modela. Rezultati faktorske analize su poslužili kao ulazni podati za sva četiri algoritma strojnog učenja. Od 35 varijabli, 4 su isključene faktorskom analizom: „Kolokvij iz excel-a“, „Kolokvij iz access-a“, „Ponavljanje kolokvija teorije“ i „Dodatne ocjene“ otkad su one dio izlane varijable, „Prolaz“. Budući da faktorska analiza objašnjava 61% ukupne varijance sa 4 faktora, onda su izvućena 4 faktora: „Samoprocjena znanja 1“, „Ostali dodatni materijali“, „Prisutnost i blic testovi“, te „Samoprocjena znanja 2“. Rezultati faktorske analize su korišteni kao ulazi u prediktivno modeliranje sa četiri različita pristupa strojnog učenja. [15]

Model	Preciznost
PCA+NN	74,15%
PCA + DT	71.23%
PCA + KNN	67.61%
PCA + BC	64.36%

Tablica 5: Usporedba preciznosti [15]

Pomoću rezultata mjerenja performansi iz tablice 5 možemo donositi različite zaključke o različitim pristupima strojnog učenja na ovaj set podataka. Vidljivo je da neuronska mreža ima najveću preciznost. No i dalje se postavlja pitanje: mogu li se rezultati generalizirati? Da bi se dobio odgovor na to pitanje provela se statistička analiza. Njezina svrha je pronaći razinu na kojoj preciznost predstavlja ponašanje alogritama strojnog učenja. Testirana su dva pristupa na jednoj domeni i korištena su dva podudarna t-test uzorka. U ovom djelu istraživanja, cilj je bio testirati ako su razlike u sredstvima značajne. Pretpostavka je da razlika između sredstava nula. Pretpostavke za provođenje t-testa su bile zadovoljene. [15]

Hipoteza	Model	t-test
H0: PCA+NN = PCA+DT	PCA+NN	P=0,04
	PCA+DT	
H0: PCA+NN = PCA+KNN	PCA+NN	P=0,01
	PCA+KNN	
H0: PCA+NN = PCA+BC	PCA+NN	P=0,01
	PCA+BC	
H0: PCA+DT = PCA+KNN	PCA+DT	P=0,02
	PCA+KNN	
H0: PCA+DT = PCA+BC	PCA+DT	P=0,02
	PCA+BC	
H0: PCA+KNN = PCA+BC	PCA+KNN	P=0,04
	PCA+BC	

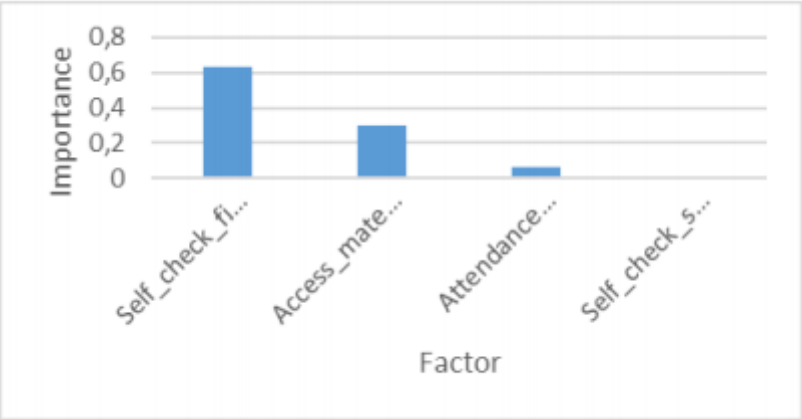
Tablica 6: Testiranja statističke značajnosti razlika između točnosti[15]

U tablici 6 možemo vidjeti rezultate t-testa. Rezultati na pokazuju velike razlike između NN sa DT, te k-NN i BC. Prema rezultatima, pristup zasnovan na pogreškama bio je najbolji prediktivni pristup modeliranju na LMS podacima. Naive Bayes klasifikator je dao najlošije rezultate. Većina varijavli jer numeričkog tipa pa zbog tog neuronske mreže su dale najbolje rezultate jer nije bilo potrebno raditi nikakve transformacije. Stablo odlučivanja je dobro odradilo posao i sa kategorijskim i numeričkim varijablama, dok Bayesov klasifikator jer bio najbolji u radu sa katerogirjskim varijablama. Ovi rezultati su pokazali da pristup zasnovan na pogreškama strojnog učenja je nedvojbeno pobjedio model zasnovan na osnovnim informacijama, model temeljen na sličnostima te model temeljen na vjerojatnostima prediktivnog modeliranja. [15]

Hipoteza	Model	t-test
H0: PCA+DT = PCA+KNN	PCA+DT	P=0,01
	PCA+KNN	
H0: PCA+DT = PCA+BC	PCA+DT	P=0,02
	PCA+BC	
H0: PCA+DT = PCA+NN	PCA+DT	P=0,04
	PCA+NN	
H0: PCA+KNN = PCA+BC	PCA+KNN	P=0,02
	PCA+BC	
H0: PCA+ KNN = PCA+NN	PCA+ KNN	P=0,01
	PCA+NN	
H0: PCA+BC= PCA+NN	PCA+BC	P=0,02
	PCA+NN	

Tablica 7: Testiranja statističke značajnosti razlika između vremena izvršavanja [15]

Vrijeme izvršavanja predstavlja ukupno vrijeme potrebno da se napravi prediktivni model koji se sastoji od vremena smanjenja začajki i razvoja prediktivnih modela zasnovanih na strojnom učenju. Rezultati pokazuju da stablo odlučivanja nadmašuje ostale pristupe prema brzini. Drugi po brzini je k-najbliži susjed, dok su neuronske mreže bile najsporiji pristup prediktivnog modeliranja. [15]

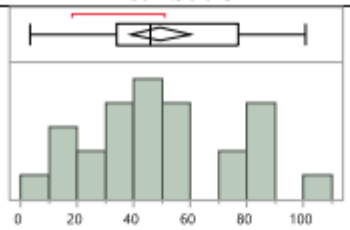
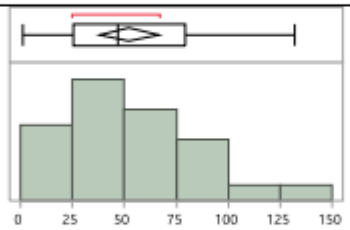
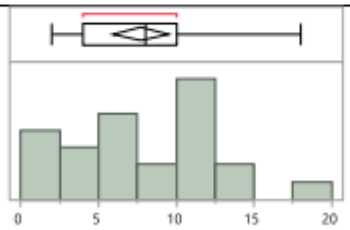
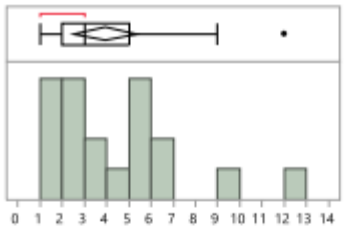
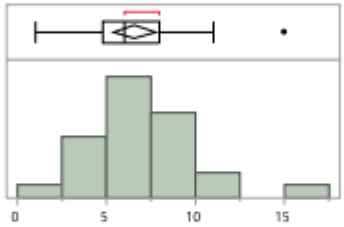
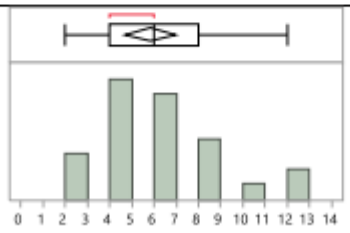


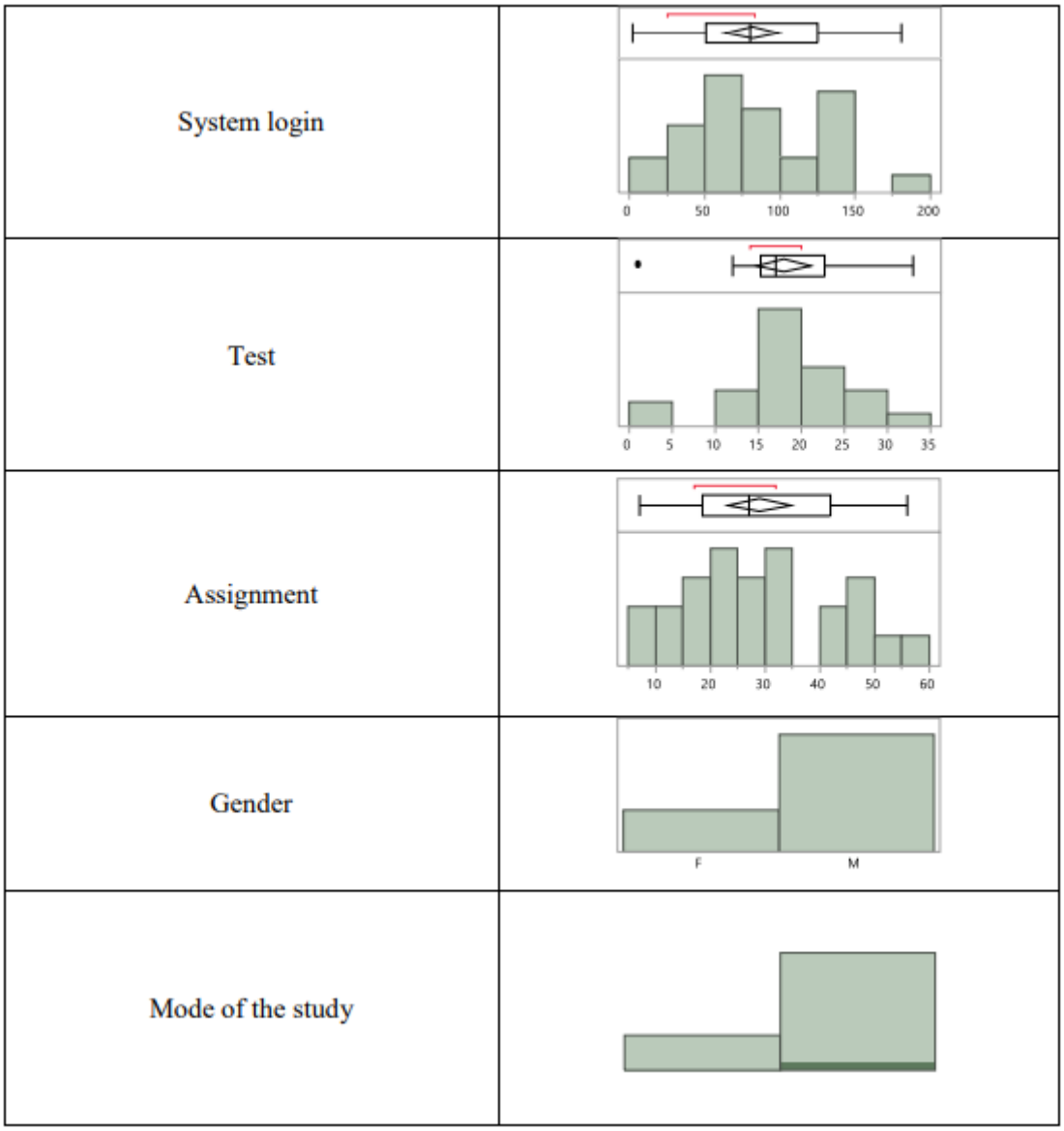
Slika 9: Analiza osjetljivosti [15]

Faktor „Pristup podacima“ (Access_materials) se sastoji od 8 od ukupno 11 varijabli koje se fokusiraju na različite pristupe studenata na LMS. Faktor „Prisutnost i blic testovi“ (Attendance and blitz tests) se sastoje od varijabli prisutnosti te tri od četiri nenajavljena testa.[15]

Sa tim faktorima se vide jasne veze između varijabli, gdje je glavno pravilo grupiranja sličnost. Ovo istraživanje se fokusiralo na faze modeliranja rudarenja podataka te istraživanja koja od četiri pristupa strojnog učenja su više efikasnija za klasifikaciju. Istraživanje je pokazalo da su neuronske mreže najbolje klasificirale studente nego ostali pristupi. Ovo istraživanjem se demonstriralo kako algoritmi strojnog učenja mogu biti korišteni da se identificiraju najbitniji atributi u LMS podacima. Studentski uspjesi mogu biti poboljšani sa korištenje metoda rudarenja podataka. [15]

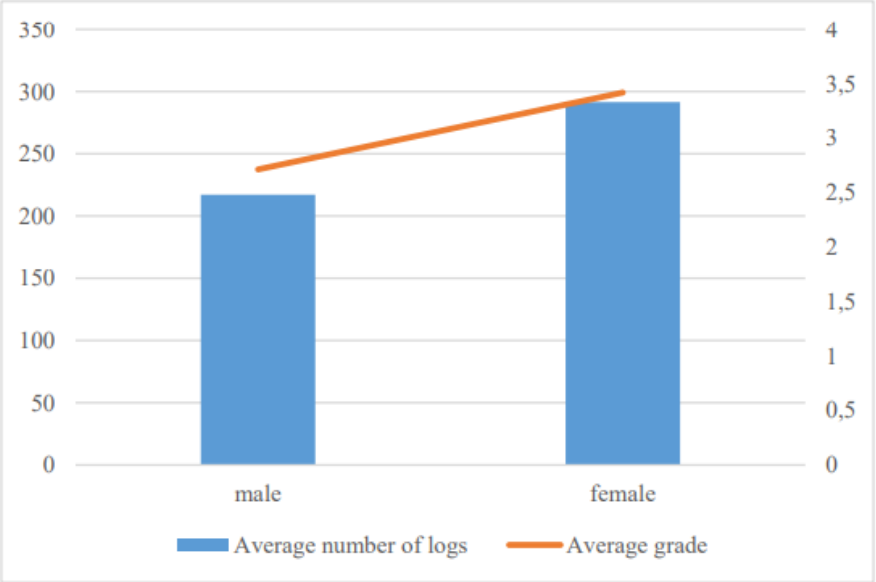
Posljednje istraživanje je vezano uz analizu ICT studenata u korištenju LMS-a te njihovog konačnog uspjeha. 28 studenata je bilo uključeno u ovo istraživanje, iz kolegija Otkrivanje znanja sa Sveučilišta u Zagrebu, Fakulteta Organizacije i Informatike. Studenti su imali tri sata predavanja svaki ponedjeljak i tri sata laboratorijskih vježbi održanih u dvije grupe: jedna svaki utorak, a druga svaku srijedu. Prikupljeni podaci su analizirani i sastavljene su sljedeće informacije: (i) opisna statistika i distribucija varijabli koje ukazuju na upotrebu određenih resursa, (ii) vizualizacija obrazaca korištenja, (iii) izračuni korelacija. [16]

Variable	Distribution
Files view	
Forum view	
Student report	
Folder view	
Choice	
File upload	

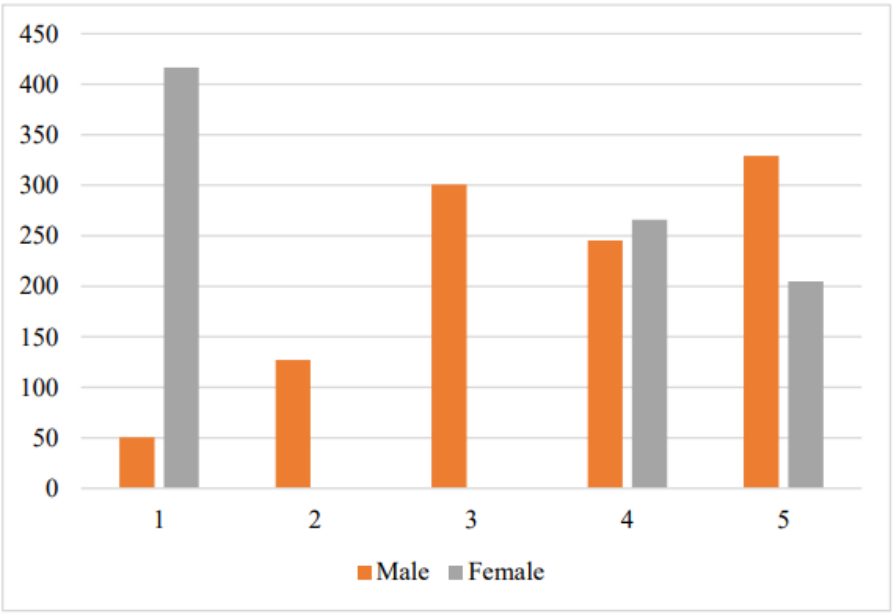


Slika 10: Opis podataka [16]

Varijabla „Files view“ je povezana sa pristupom svake datoteke u sustavu. Bilo je studenata koji su pristupili materijalima za učenje samo nekoliko puta, dok je s druge strane bilo i studenata koji su pristupili više od 100 puta. Uzimajući u obzir „Forum view“, većina studenata se ulogirala 25-50 puta na sustav. Na ovom predmetu je bilo više muških (21), nego ženskih (7) studenata. Bila su još 22 redovna studenata, te 6 vanredovnih. Na slici 10 možemo vidjeti prosječan broj logiranja po semestru. Studentice imaju bolju prosječnu ocjenu nego studenti muškog spola. [16]

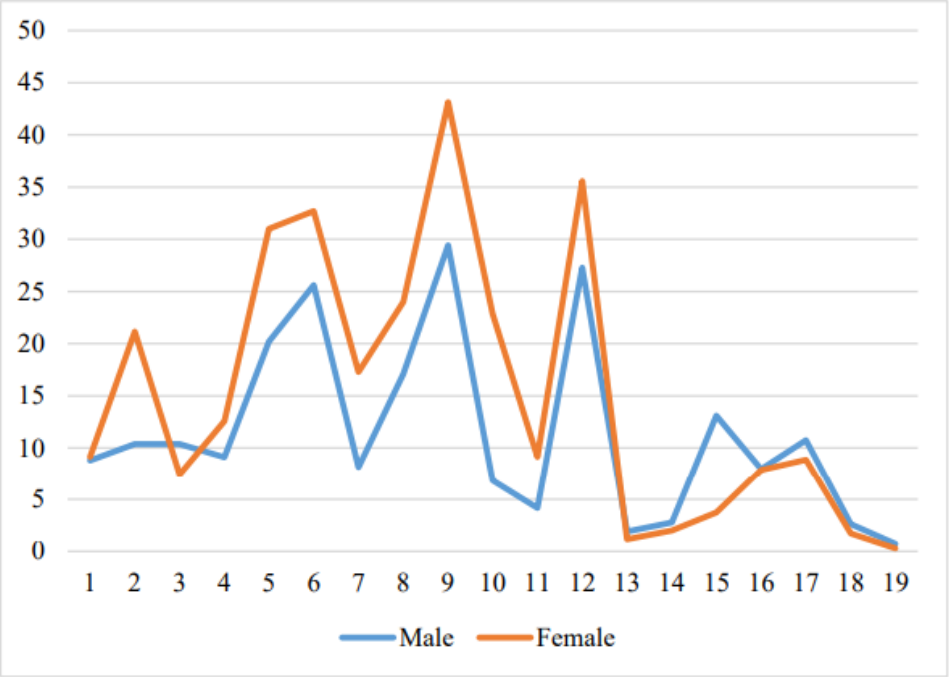


Slika 11: Prosječan broj logiranja po spolu [16]

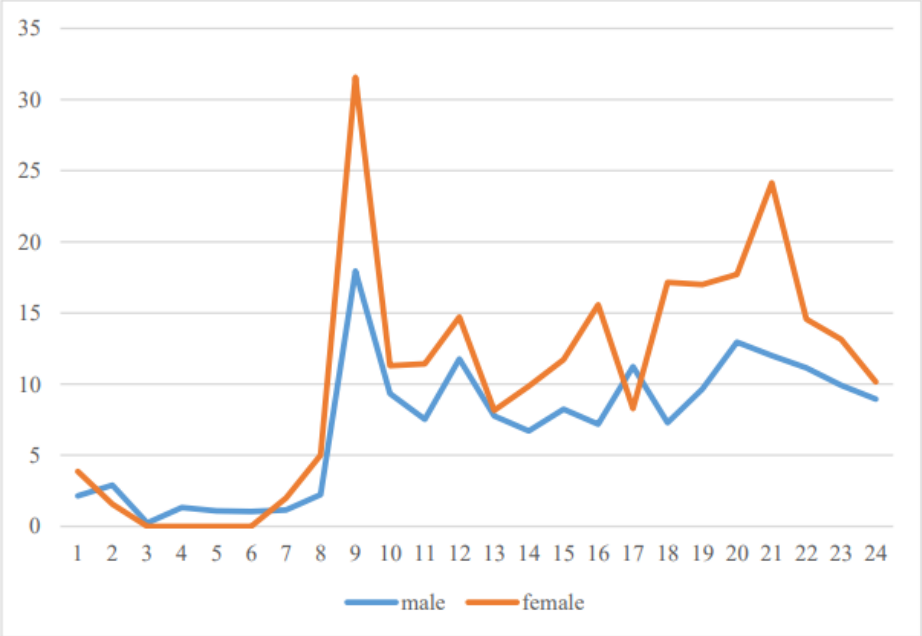


Slika 12: Prosječan broj logiranja prema ocjeni [16]

Na slici 11 možemo vidjeti graf prosječnih logiranja prema ocjeni i spolu. Zanimljiva činjenica je ta da najveći prosječni broj logiranja imaju studentice koje padnu kolegij. Detaljnom analizom je utvrđeno da studentice koje imaju najmanju ocjenu, ali najveći broj logiranja u sustavu su zapravno izvanredovne studentice. Izvaredovnih studenti u većini slučajeva ne prisustvuju nastavi fizički. To nas dovodi do zaključka da je prisustvovanje nastavi važan faktor za polaganje kolegija. Na slici 13 možemo vidjeti broj logiranja od 1. do 19. tjedna nastave. 8. i 12. tjedan su tjedni sa najvećim brojem logiranja. U 8. tjednu je bio rok za predaju zadataka dok je u 12. tjednu bio kolokvij. [16]



Slika 13: Omjer logiranja prema tjednima [16]



Slika 14: Omjer logiranja prema satima u danu [16]

Sa slike 14 je vidljivo da se najviše aktivnosti zbivalo tijekom jutarnjih sati, između 8 i 9, a to je zbog toga jer su se tada izvodile laboratorijske vježbe. Nadalje aktivnost je padal do večernjih sati, možda zbog toga da studenti provjere neke novosti na forumima. [16]

Varijabla	Varijabla	Korelacija
Forum	Prikaz datoteke	0,988804
Izvještaj studenata	Prikaz datoteke	0,990833
Izvještaj studenata	Forum	0,989881
Prikaz mape	Prikaz datoteke	0,984207
Prikaz mape	Forum	0,979175
Prikaz mape	Izvještaj studenata	0,979979
Odabir	Prikaz datoteke	0,993976
Odabir	Forum	0,989721
Odabir	Izvještaj studenata	0,995237
Odabir	Prikaz mape	0,980252
Upload datoteka	Prikaz datoteke	0,995952
Upload datoteka	Forum	0,986588
Upload datoteka	Izvještaj studenata	0,993106
Upload datoteka	Prikaz mape	0,979779
Upload datoteka	Odabir	0,994497
Izvještaj	Prikaz datoteke	0,975506
Izvještaj	Forum	0,971357
Izvještaj	Izvještaj studenata	0,979057
Izvještaj	Prikaz mape	0,997928
Izvještaj	Odabir	0,980492
Izvještaj	Upload datoteka	0,972551
Test	Prikaz datoteke	0,993891
Test	Forum	0,992324
Test	Izvještaj studenata	0,993271

Test	Prikaz mape	0,981866
Test	Odabir	0,994953
Test	Upload datoteka	0,991452
Test	Izvještaj	0,977592

Tablica 8: Analiza korelacije [16]

Rezultati pokazuju izrazitu korelaciju između testova, pregleda datoteka i mapa. Međutim, postoji i nedovoljna povezanost između ocjena i ostalih aktivnosti na kolegiju. Demografske karakteristike studenata (spol, model studiranja) nisu povezane sa ocjenama.[16]

Kod ovog istraživanja, koristio se pristup analitike učenja za analizu aktivnosti IT studenata na LMS da bi se analizirala njihova uspješnost. Studentice su više aktivnije i uključenije u aktivnosti LMS-a. [16]

5. Opis i priprema podataka

5.1 Opis problema

U ovom radu ćemo proučavati različite aktivnosti sa sustava Moodle koje utječu na prolaznost studenta. Zanima nas ako studenti koji provode više vremena na sustavu Moodle, koji rješavaju samoprocjene i druge zadatke i zadaće, imaju bolju prolaznost, odnosno, bolju ocjenu u konačnosti. To ćemo pomoću metoda rudarenja podataka pokušati analizirati.

5.2 Opis podataka

Podaci koje ćemo koristiti su sa sustava za e-učenje (Moodle) i oni su svi numerički osim varijable „Prolaz“. Ti podaci su vezani za prolaznost studenata iz predmeta Informatika 1. Skup podataka se sastoji od 36 atributa i 309 instanci.

Osnovni podaci		
Redni broj	**Naziv varijable**	**Opis**
1.	Prolaz	Varijabla koja nam govori ako je student položio kolegij.
2.	Dodatni kviz	Kviz za dodatne bodove.
3.	Blic 1	Prva nenajavljena kratka provjera.
4.	Blic 2	Druga nenajavljena kratka provjera.
5.	Blic 3	Treća nenajavljena kratka provjera.
6.	Blic 4	Četvrta nenajavljena kratka provjera.
7.	Kolokvij iz excela	Kolokvij iz MS excela na labosima
8.	Kolokvij iz access-a	Kolokvij iz MS accessa na labosima.
9.	Ponavljanje kolokvija teorija	Ponavljanje kolokvija iz cijelog semestra.
10.	Dolasci na labose	Dolasci na laboratorijske vježbe.
11.	Dolasci na predavanja	Dolasci na predavanja.
12.	Samoprocjena 1	Samoprovjera znanja 1
13.	Samoprocjena 2	Samoprovjera znanja 2

14.	Samoprocjena 3	Samoprovjera znanja 3
15.	Samoprocjena 4	Samoprovjera znanja 4
16.	Samoprocjena 5	Samoprovjera znanja 5
17.	Samoprocjena 6	Samoprovjera znanja 6
18.	Samoprocjena 7	Samoprovjera znanja 7
19.	Samoprocjena 8	Samoprovjera znanja 8
20.	Samoprocjena 9	Samoprovjera znanja 9
21.	Samoprocjena 10	Samoprovjera znanja 10
22.	Samoprocjena 11	Samoprovjera znanja 11
23.	Samoprocjena 12	Samoprovjera znanja 12
24.	Mapa	Predstavlja mapu za određene datoteke.
25.	Datoteka	Predstavlja datoteke o dodatnim materijalima vezanim za kolegij.
26.	Forum	Forum za međusobnu komunikaciju sa drugim studentima i profesorima.
27.	Izvještaj o studentu	Predstavlja izvještaj o pojedinom studentu koji prikazuje njegov napredak na kolegiju.
28.	Lekcija	Predstavlja korištenje materijala za učenje pojedinih lekcija.
29.	Odabir grupe	Odabir laboratorijske grupe.
30.	Postavljanje datoteke	Postavljanje datoteke na Moodle, npr. Razne zadaće.
31.	Poveznica	Poveznice na druge materijale vezane za kolegij.
32.	Pregledni izvještaj o studentu	Predstavlja pregledni izvještaj o pojedinom studentu.
33.	Stranica	Predstavlja razne stranice kao što su literatura, vodiči i slično.
34.	Sustav	Predstavlja pristup kolegiju.
35.	Test	Predstavlja izvještaj o testu.
36.	Zadaća	Predstavlja pristup dijelu predviđenom za zadaće.

Tablica 9: Osnovni podaci o varijablama [izrada autora]

6. Opis korištenih metoda rudarenja podataka

Sredinom devedesetih godina prošlog stoljeća došlo je do promjena u području rudarenja podataka, koje objedinjava skup metoda i postupaka, čiji je cilj otkrivanje određenih zakonitosti u velikim količinama podataka, no tek u zadnjih deset godina, komercijalno rudarenje dobiva strateški značaj u poslovnim organizacijama. (Garača i Jadrić, 2011.; Klepac, 2006.)

Za konačne rezultate ne postoji univerzalna metoda, nego ovisno o prirodi problema, dostupnosti podataka i sklonostima izvođača, mi odabiremo metodu koja nam zapravo najviše odgovara za naš problem. Kod rudarenja podataka, velik dio se oslanja na samu izradu kvalitetnog modela, a taj model je zapravo algoritam, tj. skup pravila koja povezuju ulaze sa ciljnom varijablom. (Garača i Jadrić, 2011.; Klepac, 2006.)

Sve metode imaju isti cilj, a to je da se njima prikažu kretanja podatka prema kojima onda možemo donijeti jednostavne i kvalitetne zaključke. Kada provodimo metode rudarenja podataka, preporučljivo je da se koristi više od jedne metode, ako je to moguće, da bi na kraju mogli uspoređivati rezultate dobivene različitim metodama, te da se mogu komentirati zaključci, ovisno o metodi koja nam najviše odgovara.

Metode koje sam odabrao za rudarenje podataka su: klasteriranje, neuronska mreža i stablo odlučivanja, a program u kojem ću raditi se zove JMP. Sada ćemo opisati svaku od navedenih metoda.

6.1 Klasteriranje

Klasteriranje je grupiranje ili sjedinjavanje objekata sličnih osobina. Cilj ove metode rudarenja podataka je pronaći sličnost unutar populacije koristeći zadani skup atributa. Postoje 2 načina klasteriranja:

1. K-means klasteriranje
2. Hijerarhijsko klasteriranje

U K-means klasteriranju prvo se odredi K centara klastera (npr. slučajnim odabirom). U svakom klasteru se nalazi određeni broj instanci, za njih se određuje srednja vrijednost svih varijabli koja određuje centar klastera – veličinu koja reprezentira čitav klaster. Drugi korak je podijeliti svaku stavku k njoj najbližoj vrijednosti centra klastera. U trećem koraku se svaki centar klastera premjesti prema sredini stavki koje su mu dodijeljene. U zadnjem koraku se

ponove drugi i treći korak tako dugo dok se ne dođe do konvergencije (kriterij konvergencije: promjena u raspodjeli klastera je manja od praga)

U hijerarhijskom klasteriranju kreira se hijerarhijska dekompozicija skupa podataka korištenjem određenog kriterija. Nedostatak ove metode je nemogućnost ponavljanja klasifikacije nakon podjele na klastere. Postoje dva načina: aglomerativno i divizijsko [4]

6.2 Neuronska mreža

Neuronska mreža pripada u inteligentne metode rudarenja podataka. Cilj ove metode je pronaći skrivene veze među podacima. Neuronske mreže su programi ili hardverski sklopovi koji, najčešće iterativnim postupkom iz prošlih podataka nastoje pronaći vezu između ulaznih i izlaznih varijabli modela, kako bi se za nove ulazne varijable dobila vrijednost izlaza. Umjetni je neuron jedinica za obradu podataka koja prima ponderirane ulazne vrijednosti od drugih varijabli prema nekoj formuli transformira primljenu vrijednost, te šalje izlaz drugim varijablama. Učenje se odvija promjenom vrijednosti težina među varijablama (težine wji su ponderi kojima se množe ulazne vrijednosti u neki neuron). S obzirom na broj slojeva, tip učenja, tip veze između neurona, veza između ulaznih i izlaznih podataka, ulazne i prijenosne funkcije, namjenu, razlikuju se brojni algoritmi neuronskih mreža. Zbog njegove opće namjene (budući da je pogodan i za probleme predviđanja i klasifikacije) te učestale upotrebe u istraživanjima, za modeliranje je korišten algoritam višeslojni perceptron. Višeslojni perceptron pripada u nadgledane algoritme s vezom unaprijed u kojoj se slojevi mreže povezuju na način da signali putuju samo u jednom smjeru, od ulaza prema izlazima mreže. Najpoznatiji i najčešće korišteni algoritam primijenjen za učenje i treniranje višeslojnih perceptrona mreža je tzv. Mreža „širenje unatrag“. Standardni algoritam mreže „širenje unatrag“ uključuje optimizaciju greške koristeći deterministički algoritam gradijentnog opadanja. Nedostatak ovog algoritma je problem čestog pronalaženja lokalnog umjesto globalnog minimuma greške. Strukturu mreže čine ulazni sloj, izlazni sloj i najmanje jedan skriveni sloj s vezom unaprijed. [5]

6.3 Stablo odlučivanja

Ovom metodom dobiva se grafički prikaz modela utjecaja ulaznih varijabli na izlaznu, koja je izražena u obliku klasa ili kategorija. Svaki čvor u grafičkom stablu predstavlja jednu ulaznu varijablu, na čijim su rubovima označena „djeca-čvorovi“ za svaku moguću vrijednost neke ulazne varijable. Svaki list u stablu predstavlja vrijednost ciljne (izlazne) varijable ako su dane vrijednosti ulaznih varijabli predstavljene putom od korijena stabla do tog lista. Stablo

se dobiva „učenjem“ na podacima, na način da se vrši grananje izvornog skupa podataka u pod skupove na temelju testiranja vrijednosti varijabli. Proces se ponavlja na svakom izvedenom podskupu na rekurzivni način. Rekurzija je završena kada podskup određenog čvora ima sve iste vrijednosti izlazne varijable ili kada daljnje grananje više ne pridonosi poboljšanju rezultata. Prednosti stabla odlučivanja su: jednostavni za razumijevanje i objašnjavanje, traže jednostavnu pripremu podataka, mogu raditi s numeričkim i kategorijskim vrijednostima, brzi su, koriste model bijele kutije. [5]

7. Klaster analiza

Na odabranom skupu podataka proveo sam klaster analizu. Isprobano je nekoliko klaster analiza s različitim brojem klastera. Odabrao sam klaster analizu sa 2 klastera jer ako sam povećavao ili smanjivao klastere tada su se pojavljivali klasteri sa samo jednom ili dvije instance što baš i nije imalo prevelikog smisla. To možemo vidjeti na slici 15.

Cluster Comparison

Method	NCluster	CCC	Best
K Means Cluster	2	0,35912	Optimal CCC
K Means Cluster	3	-3,4269	
K Means Cluster	4	-5,6452	
K Means Cluster	1	0	

Slika 15: Usporedba broja klastera [izrada autora]

Pronađeni optimalni broj klastera je 2 što se može vidjeti na slici 15 (Optimal CCC). Taj podatak klastera dobivamo i usporedbom standardnih devijacija. Daljnjim povećanjem broja klastera povećava se iznos CCC, a samim time i standardna devijacija (standardna devijacija mora biti što bliže 0).

Cluster Summary

Cluster	Count	Step	Criterion
1	222	9	0
2	87		

Slika 16: Raspodjela instanci [izrada autora]

Na slici 16 možemo vidjeti da prvi klaster čini 222 instanci, dok drugi klaster čini 87 instanci. Prvi skup gdje je većina polaznika položila ispitivani kolegiji su bolji u gotovo svim kategorijama. Neke od tih kategorija u kojima su bolji (velike razlike) jesu: „Sustav“(odnosno vrijeme provedenu na moodle sustavu), puno su više koristili materijale za učenje („Lekcija“) te su puno više koristili i rješavali samoprocjene dok u drugoj skupini te su aktivnosti gotovo zanemarene. Drugu skupinu odnosno klaster čine polaznici koji u većini nisu uspjeli položiti kolegij. Jedina kategorija u kojima su oni bolji tj. aktivnost u kojima su više sudjelovali od

prvog klastera je: „Ponavljanje kolokvija teorija“

Cluster Means

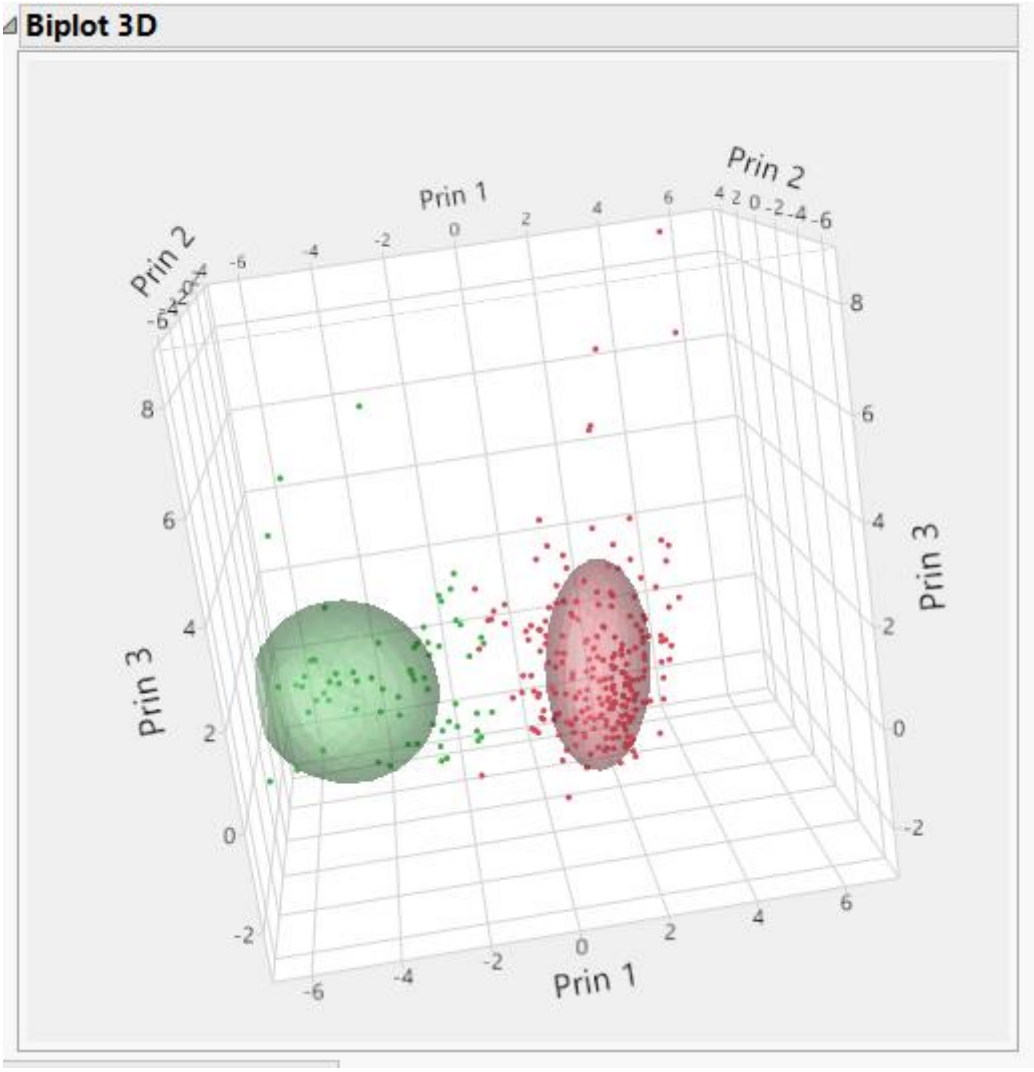
Cluster	Prolaz	Dodatni kviz	Blic 1	Blic 2	Blic 3	Blic 4	Kolokvij excel	Kolokvij access	Ponavljanje kolokvija teorija	Dolasci na labose	Dolasci na predavanja	Samoprocjena 1
1	0,6036036	2,08981982	3,75986486	4,84364865	3,735	3,36887387	3,73536036	6,19144144	1,03603604	13,7586486	7,63513514	14,9925676
2	0,12643678	0,62781609	2,44954023	3,02505747	2,0983908	1,80137931	2,20689655	3,44827586	1,78735632	10,5631034	5,63218391	11,5489655

Samoprocjena 2	Samoprocjena 3	Samoprocjena 4	Samoprocjena 5	Samoprocjena 6	Samoprocjena 7	Samoprocjena 8	Samoprocjena 9	Samoprocjena 10	Samoprocjena 11	Samoprocjena 12	Datoteka
14,055991	18,5957207	14,8072973	10,2272973	19,1693243	3,34666667	15,708018	8,88774775	17,6268018	10,0825676	10,4183784	34,5495495
9,48114943	10,3405747	8,88942529	4,65747126	3,11241379	0,34862069	1,76436782	0,50114943	0,72609195	0,23091954	0,13597701	22,7931034

Forum	Lekcija	Postavljanje datoteke	Poveznica	Stranica	Sustav	Zadaća
45,6306306	1577,33784	11,5855856	2,65315315	2,67567568	432,990991	44,9054054
30,0114943	718,988506	8,8045977	1,87356322	2,6091954	232,816092	32,2413793

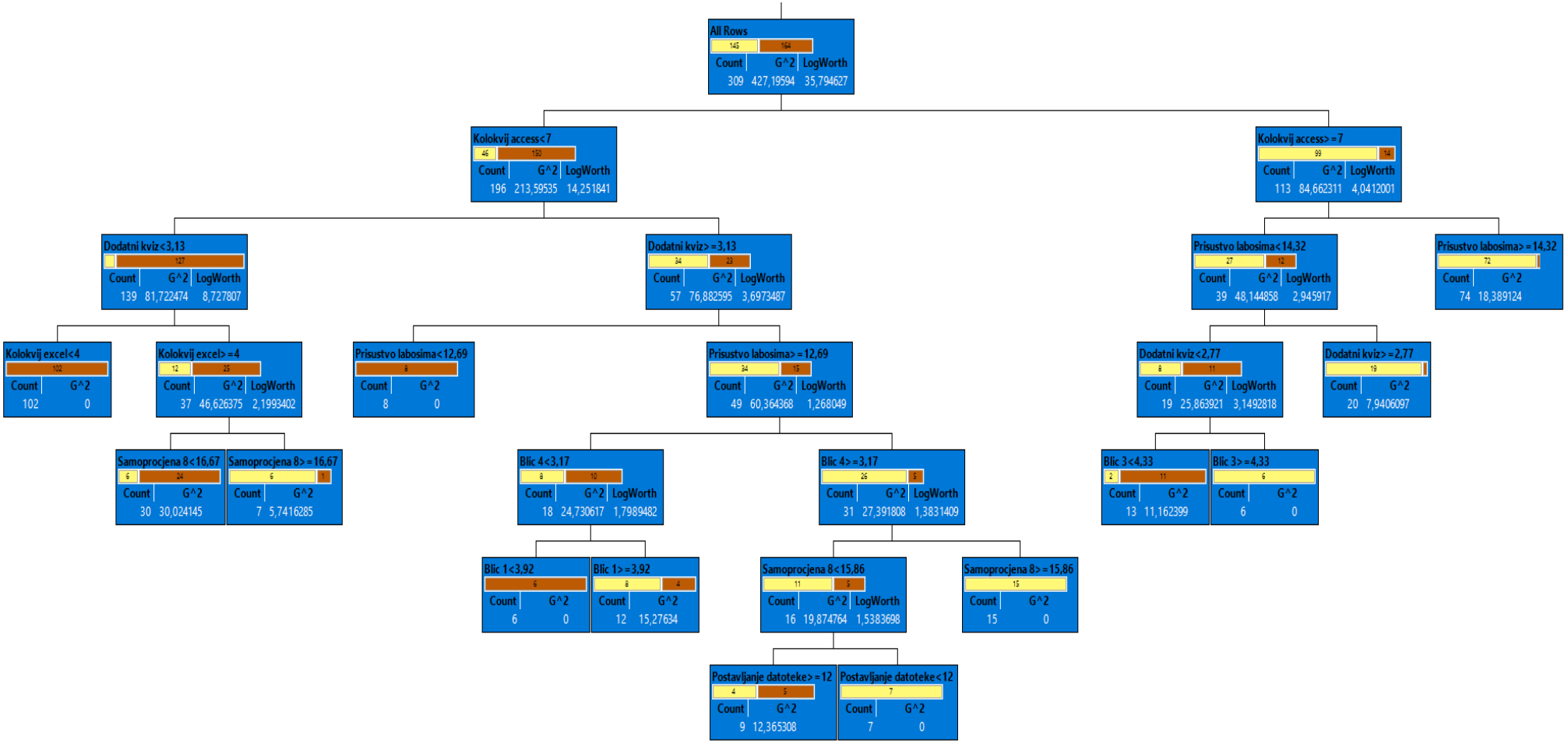
Slika 17: Sredine klastera [izrada autora]

Na slici 18 možemo vidjeti 3D prikaz klastera. Vidi se međusobni položaj i udaljenost tih dvaju klastera, te se isto tako vidi i raspored instanci unutar svakog klastera, te raspored instanci koje su u okolini klastera.



Slika 18: 3D prikaz klastera [izrada autora]

8. Stablo odlučivanja

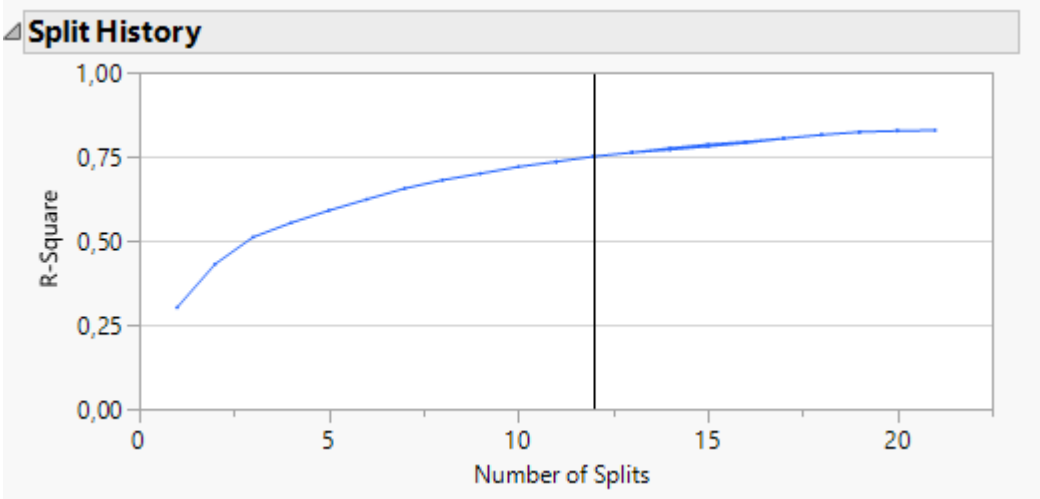


Slika 19: Stablo odlučivanja [izrada autora]

Na slici 19 možemo vidjeti stablo odlučivanja. Odabrao sam 12 grananja jer je pouzdanost kod toga grananja 75 % (dok je vjerojatnost pogreške relativno mala), a daljnjim povećanjem pouzdanost se jako malo povećavala dok se samo stablo dosta povećalo. Međutim, ako smanjimo broj grananja, tada se pouzdanost dosta smanjila. Iz stabla vidimo da je najvažniji element najbliže korijenu stabla. U mom slučaju to je „Kolokvij iz laboratorijskih vježbi“, te ga slijede „Dodatni kviz“ i „Dolasci na labose“. Za interpretaciju stabla dati ćemo jedan primjer interpretacije grane skroz desno na stablu. Pa tako možemo vidjeti da se nakon korijena studenti dijele u dvije skupine i to u one koji su imali više ili jednako od 7 (113) i manje od 7 boda (196) iz „Kolokvij access“. Na grani desno, od tih 113 studenta, njih 14 ih nije položilo kolegij dok je njih 99 položilo. Zatim se dijele prema „Dolasci na labose“ gdje je onih koji imaju više ili manje od 14.32 dolazaka 74, a onih koji su imali manje od 14.32 dolazaka je 39. Od tih 74 studenta , dvoje studenta nisu položili kolegij, dok je njih 72 položilo kolegij.

RSquare	N	Number of Splits
0,751	309	12

Slika 20: Prikaz pouzdanosti [izrada autora]



Slika 21: Graf pouzdanosti [izrada autora]

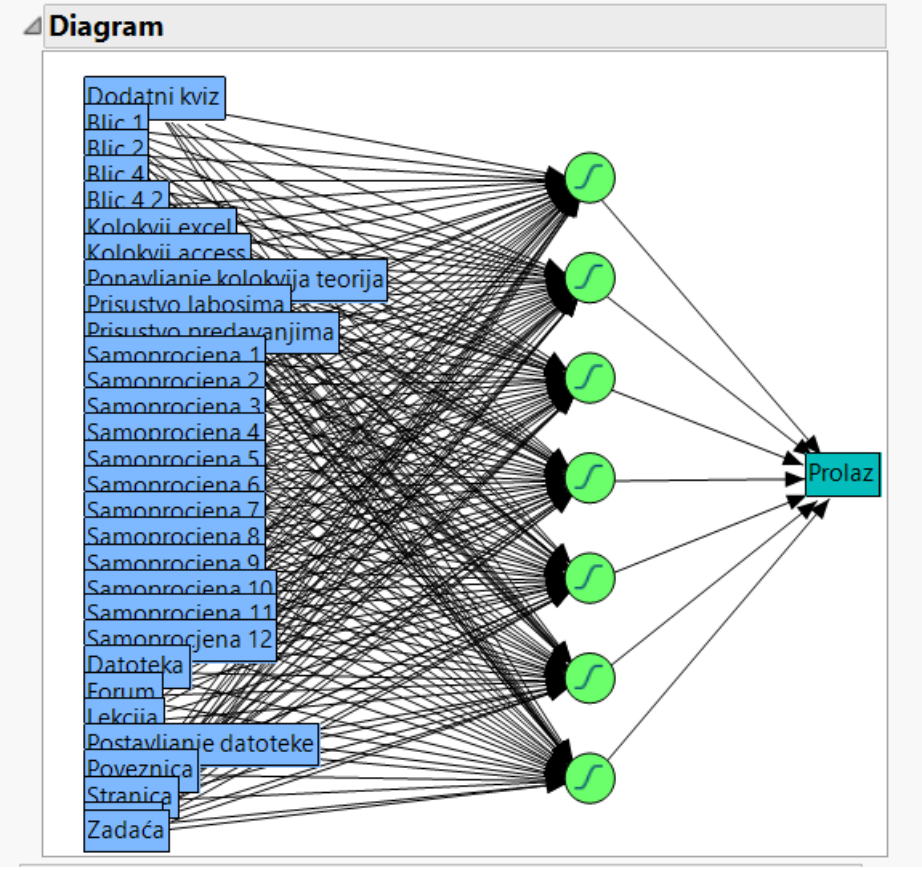
Na slici 21 možemo vidjeti graf pouzdanosti. Vidimo da je maksimalan broj grananja 21, te da je za prvo i drugo grananja mala pouzdanost te nakon toga raste dok ne dolazi do sporog rasta oko 12, 13 grananja.

9. Neuronska mreža

Za analizu skupa podataka sam također koristio metodu neuronske mreže. Za neuronsku mrežu sam se odlučio zbog toga jer imamo numeričke atribute. Neuronsku mrežu sam radio u programu JMP. Broj skrivenih neurona sam odredio po formuli:

(broj ulaza + broj izlaza) ^ 0.5 + (1 do 10)

Prema toj formuli sam izračunao potencijalnih deset brojeva skrivenih neurona. Kada sam napravio neuronsku mrežu za svih deset rezultata, ustanovio sam da je broj skrivenih neurona 7, jer je u tom slučaju točnost je najveća.



Slika 22: Neuronska mreža s 7 neurona [izrada autora]

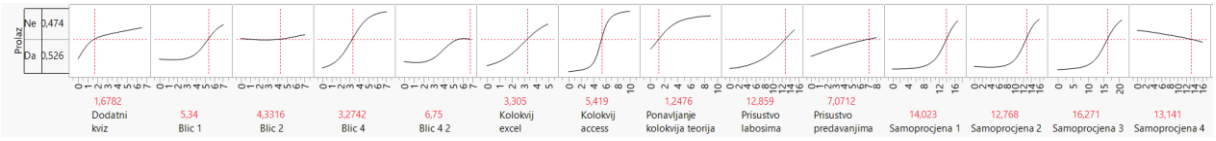
Kod neuronske mreže bitna je točnost te pouzdanost modela. Na slici 22 možemo vidjeti tablicu u kojoj vidimo da model pogrešno predviđa u 33% slučajeva. Taj podatak nam govori atribut RMSE. Pouzdanost modela se vidi iz atributa entropy Rsquare. On nam govori da je pouzdanost modela 49%.

Prolaz

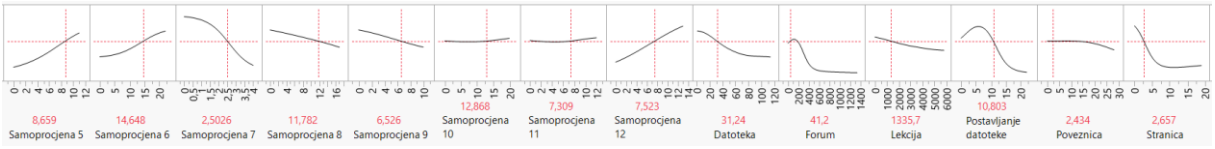
Measures	Value
Generalized RSquare	0,6547105
Entropy RSquare	0,4875757
RMSE	0,3277634
Mean Abs Dev	0,201693
Misclassification Rate	0,1634615
-LogLikelihood	36,850549
Sum Freq	104

Slika 23: Rezultati neuronske mreže [izrada autora]

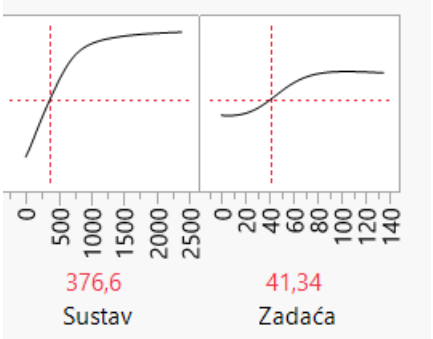
U nastavku ću još prikazati kako promjene vrijednosti ulaznih varijabli utječu na izlaznu varijablu.



Slika 24: Prediction profiler [izrada autora]



Slika 25: Prediction profiler [izrada autora]



Slika 26: Prediction profiler [izrada autora]

Na ovim profilerima možemo za svaku ulaznu varijablu posebno mijenjati vrijednosti i vidjeti koja varijabla više utječe, a koja manje na prolaz kolegija. Tako npr. možemo vidjeti kako prisustvo laboratorijskim vježbama utječe na prolaz. Ako je vrijednost varijable prisustvo labosima 4,61 šansa da nećemo proći kolegij je visokih 87%, a ako bi vrijednost te varijable bila 12 onda se šansa da ne prođemo kolegij drastično smanjila na 54%.

10. Opis i diskusija rezultata

Cilj ovog projekta bio je vidjeti kako pojedine aktivnosti sa sustava za e-učenje (Moodle), te koje aktivnosti utječu na uspjeh (prolaznost) studenta. Prije samog rudarenja podacima, izbacili smo neke varijable, kao što su „Map)“, „Odabir grupe za labose“, „Izvještaj o studentu“, „Pregledni izvještaj o studentu“ i „Provjera iz Accessa“ jer nisu imale toliki utjecaj na končanu prolaznost studenta.

Kod klaster analize, uzeli smo dva klastera jer ako smo uzeli više, imali smo klastere sa velikim odstupanjem, a i imamo samo dvije mogućnosti izlaza, prolaz ili pad. Zanimljiva činjenica kod te metode je da se u prvom klasteru nalazi većina studenata koja je položila kolegij te to potvrđuje većina varijabli, dok u drugom klasteru, gdje su studenti koji u većini nisu položili. Oni su jedino bolji u kategoriji „Ponovno pisanje gradiva teorije“.

Druga metoda rudarenja podataka je bila stablo odlučivanja. Tom metodom smo napravili 12 grananja kod kojih pouzdanost iznosi 75%. Ako bi napravili još više grananja, pouzdanost se bi povećavala ali u zanemarivo malom postotku. Varijabla koja najviše utječe na konačan ishod je „Kolokvij iz Accessa“. Iz toga možemo zaključiti da taj kolokvij najviše utječe na konačnu prolaznost studenta.

Sljedeća metoda koju smo koristili je neuronska mreža. Prema formuli, izračunali smo deset mogućih brojeva neurona, ali smo onda daljnjim testiranjem došli do konačnog broja od sedam neurona. Pouzdanost neuronske mreže iznosi 49% što je znatno manje nego kod stabla odlučivanja, dok u 33% slučajeva model pogrešno predviđa. Iz toga možemo zaključiti da stablo odlučivanja više odgovara za naš skup podataka. Kod te metode možemo vidjeti kako svaka pojedina varijabla utječe na prolaznost studenta.

11. Zaključak

U današnje vrijeme, količina podataka neprestano raste. Zbog toga javlja se problem prilikom donošenja odluka bitnih kod poslovanja velikih poduzeća. Kako bi stali na kraj tim mukama, poduzeća su se počela baviti rudarenjem podataka da bi pronašle bitne informacije u toj ogromnoj količini podataka. Ukratko možemo reći da je rudarenje podataka zapravo sortiranje, grupiranje i organiziranje podataka sa ciljem dobivanja kvalitetnih informacija. Postoji 6 zadaća rudarenja podataka, a to su : deskripcija, klasifikacija, estimacija, predikcija, klasteriranje i asocijacija. Nadalje postoje 4 koraka po kojima možemo provodit proces rudarenja podataka kako bi smo mogli dobiti kvalitetne rezultate kojima možemo rukovati. Proces rudarenja podataka je iterativan, tj. možemo se vratiti na prethodni korak u bilo kojem trenutku, jer je gotovo nemoguće iz prve dobro odabrati i pripremiti podatke koje želimo analizirati, a i time će nam se povećati znanje o samim podacima koje analiziramo.

Tema završnog rada je bila analizirati kako pojedini elementi kolegija Informatika 1 utječu na prolaz studenta na kolegiju. Korištenjem različitih metoda rudarenja podataka htjeli smo pobliže vidjeti koji element kolegija više ili manje utječe na prolaz kolegija. Metode koje smo koristili su bile klaster analiza, stablo odlučivanja, te neuronska mreža. Iz ovog projekta možemo zaključiti da je za naš slučaj najbolja metoda stablo odlučivanja koja ima pouzdanost od visokih 75%. Nakon ovog projekta smo poboljšali svoje razumijevanje koji elementi više utječu na prolaz pa sad znamo na što se treba više fokusirati, a na što manje kako bi se kolegij položio. Metoda stabla odlučivanja nam govori kako na prolaz kolegija najviše utječe „Kolokvij access“.

Iz prethodnih istraživanja možemo vidjeti kako je uvođenje sustava za e-učenje pozitivno utjecalo na prolaz studenata. Istraživanje koje je provedeno na Agronomskom fakultetu u Zagrebu nam govori kako se povećao broj studenata koji su prošli kolegij, te se smanjio broj studenata koji su imali ocjenu 2, a najčešća ocjena je bila 4 dok prije uvođenja sustava za e-učenja, najčešća ocjena je bila 3. Iz istraživanja na kojem je praćeno logiranje na sustav Moodle možemo zaključiti da su studentice aktivnije i uspješnije na samom kolegiju, te da su studenti najaktivniji u tjednima kolokvija i to najviše na dan prije kolokvija. Za istraživanje vezano uz usporedbu strojnog učenja na pristup LMS podacima koristila su se 4 pristupa. Ispostavilo se da su neuronske mreže najbolje klasificirale studente nego ostali pristupi. Zadnje istraživanje je bilo vezano za analizu ICT studenata u korištenju LMS-a te njihovog konačnog uspjeha na kolegiju. U tom istraživanju se opet dokazalo da su studentice više aktivnije i uključenije u aktivnosti LMS-a.

12. Literatura

[1] L. Jones, K. (n.d.) The Advantages of eLearning. Dostupno 3.1.2020. na https://ezinearticles.com/?The-Advantages-of-eLearning&id=603386

[2] K.Klasnić, J.Lasić-Lazić, S. Seljan; Mjerenje kvalitete integriranog sustava za e-učenje na Filozofskom fakultetu u Zagrebu iz perspektive studenata. Dostupno 3.1.2020. na http://www.academia.edu/download/42838271/Mjerenje_kvalitete_integriranog_sustava_20160219-11848-1105i4l.pdf

[3] L. Blašković, D. Mandušić; Utjecaj korištenja sustava za e-učenje na uspješnost učenja studenata agronomskog fakulteta. Dostupno 4.1.2020. na https://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=314059

[4] D. Oreški; Klaster analiza, Materijal preuzet iz kolegija Otkrivanje znanja u podacima akademske godine 2019./20., preuzeto s Moodle sustava FOI-a 8.1.2020.

[5] Dr.sc.M. Zekić-Sušac, A. Frajman- Jakšić univ. Spec. Asistent, N. Drvenkar univ. Spec. Asistent; Neuronske mreže i stabla odlučivanja za predviđanje uspješnosti studiranja. Dostupno 8.1.2020. na https://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=73924

[6] MMSCODE (2019.); „Data mining“ ili metoda „vrijedna zlata“. Dostupno 8.1.2020. na http://www.mmscode.com/2019/06/21/data-mining-ili-metoda-vrijedna-zlata/

[7] Rudarenje podataka kao metoda upravljanja znanjem. Dostupno 3.8.2020. na: https://www.bib.irb.hr/959793/download/959793.5b97edeb62c7d_Rudarenje_podataka_kao_metoda_upravljanja_znanjem.pdf

[8] Garača Ž, Jadrić M (2011). „Rudarenje podataka: različiti aspekti informacijskog društva“. Split: Ekonomski fakultet u Splitu

[9] UNIZD (2020): Metode znanstvenih istraživanja. Dostupno 6.8.2020. na: http://www.unizd.hr/portals/4/nastavni_mat/1_godina/metodologija/metode_znanstvenih_istrazivanja.pdf

[10] Rudarenje podataka u bankarstvu. Dostupno 8.8.2020. na: https://hrcak.srce.hr/26220

[11] Proces dana mininga nad podacima o prodaji tekstila. Dostupno 8.8.2020. na: https://repozitorij.foi.unizg.hr/islandora/object/foi:2024

[12] Baragoin, C., Andersen, C.M., Bayerl, S., Bent, G., Lee, J., Schommer, C. (2001). Mining Your Own Business in Banking Using DB2 Intelligent Miner for Data.

[13] Klepac G. (2006). „Što je to data mining ?“. Dostupno 17.8.2020 na: http://www.goranklepac.com/index.asp?j=HR&iz=1&sa=1&vi=1&hi=1

[14] Oreški D., Kadoić N., Analysis of student behavior and succes based on log in Moodle. Dostupno 27.8.2020 na: https://ieeexplore.ieee.org/abstract/document/8400123

[15] Oreški D. Hadjin G., A Comparative Study of Machine Learning Approaches on Learning Managment System Data. Dostupno 28.8.2020 na: https://ieeexplore.ieee.org/abstract/document/9057187

[16] Oreški D., Kadoić N., Analysis of ICT student's LMS engagement and success. Dostupno 29.8.2020 na: https://bib.irb.hr/datoteka/968614.eds_rad_proec.pdf

13. Popis tablica i slika

13.1 Popis slika

Slika 1: Faze rudarenja podataka [7] 6
Slika 2: Frekvencija broja logiranja prema ocjenama [14] 12
Slika 3: Analiza po spolu [14].....14
Slika 4: Omjer logiranja po tjednima [14] 14
Slika 5: Omjer broja logiranja prema danima prije kolokvija [14] 15
Slika 6: Omjer logiranja prema satima u danu [14].....15
Slika 7: Omjer broja logiranja prema danu u tjednu sa postignutom ocjenom [14].....16
Slika 8: CRISP DM metodologija [15].....19
Slika 9: Analiza osjetljivosti [15] 22
Slika 10: Opis podataka [16].....24
Slika 11: Prosječan broj logiranja po spolu [16] 25
Slika 12: Prosječan broj logiranja prema ocjeni [16] 25
Slika 13: Omjer logiranja prema tjednima [16].....26
Slika 14: Omjer logiranja prema satima u danu [16].....26
Slika 15: Usporedba broja klastera [izrada autora].....34
Slika 16: Raspodjela instanci [izrada autora] 34
Slika 17: Sredine klastera [izrada autora].....35
Slika 18: 3D prikaz klastera [izrada autora].....36
Slika 19: Stablo odlučivanja [izrada autora] 37
Slika 20: Prikaz pouzdanosti [izrada autora].....38
Slika 21: Graf pouzdanosti [izrada autora] 38
Slika 22: Neuronska mreža s 7 neurona [izrada autora] 39
Slika 23: Rezultati neuronske mreže [izrada autora].....40

Slika 24: Prediction profiler [izrada autora].....40
Slika 25: Prediction profiler [izrada autora].....40
Slika 26: Prediction profiler [izrada autora].....40

13.2 Popis tablica

Tablica 1: Ocjene studenata prije uvođenja i primjene sustava za e-učenje [2].....11
Tablica 2: Ocjene studenata nakon uvođenja i primjene sustava za e-učenje [2].....11
Tablica 3: Korelacija između varijabli (*p<0.05) [14].....13
Tablica 4: Opis varijabli [15].....18
Tablica 5: Usporedba preciznosti [15] 20
Tablica 6: Testiranja statističke značajnosti razlika između točnosti[15] 20
Tablica 7: Testiranja statističke značajnosti razlika između vremena izvršavanja [15].....21
Tablica 8: Analiza korelacije [16].....28
Tablica 9: Osnovni podaci o varijablama [izrada autora].....30