Podatkovna analitika u poslovanju

Saliu, Leonora

SVEUČILIŠTE U ZAGREBU

FAKULTET ORGANIZACIJE I INFORMATIKE

V A R A Ž D I N

Leonora Saliu

PODATKOVNA ANALITIKA U POSLOVANJU

ZAVRŠNI RAD

Varaždin, 2021.

SVEUČILIŠTE U ZAGREBU

FAKULTET ORGANIZACIJE I INFORMATIKE

V A R A Ž D I N

Leonora Saliu

JMBAG: 0016130141

Studij: Informacijski sustavi

PODATKOVNA ANALITIKA U POSLOVANJU

ZAVRŠNI RAD

Mentorica:

Doc. dr. sc. Dijana Oreški

Varaždin, srpanj 2021.

Leonora Saliu

Izjava o izvornosti

Izjavljujem da je moj završni 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

Završni rad bavi se analizom podataka u poslovanju, odnosno modelom predviđanja odlaska klijenata iz telekomunikacijskih tvrtki. Na samom početku, opisana je problematika odlaženja klijenata te nekoliko istraživanja koja su se bavila sličnim problemima. Prije same izrade i analiziranja modela, opisuje se odabrani skup podataka nad kojim se provode analize te metode kojima se skup podataka tretira. Metode odabrane za analizu u završnom radu su klaster analiza, stablo odlučivanja i neuronske mreže.

Klaster analiza služi grupiranju podataka u skupine koje dijele mnoge sličnosti, ali i omogućava lakšu usporedbu skupina na način da su uočljivije razlike između klastera, a ne samo sličnosti unutar klastera. Stablo odlučivanja je prediktivna metoda koja svojom strukturom omogućava lakše predviđanje donošenja odluka kroz grananje (svako grananje – donošenje jedne odluke). Neuronske mreže su također prediktivna metoda analize podataka koje pokazuju još i međusobnu povezanost odnosno ovisnost atributa jednima o drugima te njihov utjecaj na donošenje odluka.

Ključne riječi: model; klaster; stablo odlučivanja; neuronska mreža; pouzdanost; važnost atributa; odlazak;

Sadržaj

1. Uvod

Temeljni cilj ovog rada je analiza odabranog skupa podataka i izrada modela predviđanja korištenjem istog. Izrada modela može se svrstati u praktični dio ovog rada, no prije samog kreiranja modela, potrebno je dobro objasniti problematiku, skup podataka i metode koje će se provesti nad podacima. S obzirom na to da je današnje stanje u svijetu takvo da je svatko svakome konkurencija, interes svakog poduzeća je istovremeno pridobivanje novih klijenata, ali i zadržavanje postojećih. Problem nastaje kada poduzeća ne mogu spriječiti odlazak svojih klijenata jer ne znaju iz kojih razloga oni odlaze. Tu u pomoć dolazi model predviđanja koji daje odgovore na takva pitanja.

U ovom radu nakon teorijske obrade skupa podataka i metoda analize, slijedi praktičan dio. U praktičnom dijelu prvo se klaster analizom podaci iz skupa grupiraju. Nakon toga slijede prediktivne metode stabla odlučivanja i neuronske mreže. To su metode koje pokazuju koji atributi najviše utječu na donošenje odluka kod klijenata te kako su ti atributi međusobno povezani i koje će pomoći u donošenju zaključaka vezano za problematiku odlaska klijenata u ovom slučaju iz telekomunikacijskih tvrtki.

2. Metode i tehnike rada

Za uspješnu izradu modela predviđanja odlaska klijenata korišten je online alat BigML. BigML je skalabilna platforma za strojno učenje koja olakšava rješavanje i automatizaciju zadataka klasifikacije, predviđanja, analiza brojnih klastera, otkrivanje anomalija i dr. Koriste ga brojni analitičari, programeri i znanstvenici diljem svijeta. Ovaj alat omogućuje pretvaranje podataka u djelotvorne modele koji se kasnije koriste kao udaljeni servisi ili se ugrađuju u programe za predviđanje [13].

Pomoću ovog alata, odabrani skup podataka pretvoren je u model predviđanja odlaska klijenata iz telekomunikacijskih tvrtki. Prvi korak bio je kreiranje i analiza klastera, zatim kreiranje stabla odlučivanja te pregled pouzdanosti određenog pravila odabirom grupe u stablu i za kraj izrada neuronske mreže za krajnje predviđanje i izvođenje zaključaka o razlozima odlaska.

Alat je dostupan na poveznici: https://bigml.com/.

3. Opis sličnih istraživanja

Tvrtke diljem svijeta neovisno o njihovom području rada susreću se s problemom odlaska klijenata. Stoga, provode se brojna istraživanja kako bi se ustanovili najčešći razlozi odlazaka i time u budućnosti spriječili isti. Jedan od primjera takvih istraživanja proveden je na telekomunikacijskoj tvrtki u Estoniji [2]. Temeljni cilj ovog istraživanja bio je doći do krucijalnih čimbenika koji utječu na privrženost, odnosno faktore koji govore o odanosti klijenata i mogućem povećanju razine iste. Temeljem rezultata istraživanja donesen je zaključak koji govori da tretiranje svih korisnika jednako nije najprecizniji potez, odnosno da se identične metode povećavanja odanosti klijenata ne mogu koristiti kod svakog pojedinca. Najvažniji faktor koji utječe na odanost klijenata pokazalo se da je pouzdanost samog proizvoda i pružatelja usluga iz čega se može zaključiti da porastom kvalitete i pouzdanosti proporcionalno raste i odanost klijenata.

Sljedeći primjer je istraživanje provedeno od strane JD Power koje proučava koliko klijenti zapravo vjeruju svojim primarnim bankama [3]. Istraživanje otkriva kako četiri od pet klijenata smatra da predstavnici u bankama ne posvećuju dovoljno vremena što dovodi do problema prepoznavanja potreba klijenata. Komunikacija s klijentima važan je faktor u procesu prepoznavanja potreba, ali još važniji u izgradnji povjerenja klijenata. Također, četiri od pet ispitanika tvrdi da nisu dovoljno upoznati s uslugama, odnosno da ih ne razumiju u potpunosti. Može se zaključiti da je komunikacija predstavnika u banci vrlo važna za održavanje odnosa s klijentima, odnosno održavanje poželjne razine zadovoljstva klijenata uslugom.

Posljednji primjer je istraživanje provedeno od strane TechSee čiji je temeljni cilj bio utvrditi razloge odlaska klijenata u telekomunikacijskoj industriji [4]. Ispitanici bili su Amerikanci koji su otkazali ugovor s pojedinom tvrtkom u posljednje dvije godine. Istraživanje pokazuje kako je 39% ispitanika kao primarni razlog navelo nezadovoljstvo službom za korisnike (predugo čekanje za rješavanje problema, agenti s nedovoljno znanja, klijenti moraju zvati više od jednom). Kao i u prethodnom primjeru može se reći da je komunikacija između agenta koji pruža pomoć korisnicima i korisnika vrlo važna isto kao i dostupnost, učinkovitost i brzina u rješavanju problema.

4. Opis problema i skupa podataka

Odlazak klijenata iz tvrtke predstavlja veliki problem u cijelom svijetu. Kako bi se utvrdilo koji faktori utječu na odlazak klijenata, provedeno je mnogo istraživanja temeljem kojih su sakupljeni brojni podaci. Proučavanjem skupa podataka može se doći do zaključaka i izrade modela za predviđanje odlaska klijenata iz tvrtki te mogućnosti sprječavanja istih. Sam problem odlaska i skup podataka s kojim će se raditi u ovom radu biti će opisani u sljedeća dva potpoglavlja.

4.1. Problem

U današnje vrijeme kada je svatko svakome konkurencija, vrlo je važno pokušati zadržati postojeće klijente i pronaći razloge iz kojih određeni pojedinci odlaze. Svaki razvoj tehnologija i povezivanje ima svoje prednosti i mane. Koliko god je širenje informacija putem interneta donijelo dobra marketinškom odjelu tvrtke, u isto vrijeme može biti i minus to što korisnik može dobiti raznovrsne ponude i informacije o uslugama konkurenata kroz samo nekoliko klikova. Telekomunikacijske tvrtke dobar su primjer jer za korištenje njihovih usluga, korisnika zanimaju određeni parametri koje lako može usporediti s parametrima koje nudi konkurent te se na temelju usporedbe odlučiti koje odnosno čije usluge želi koristiti.

Činjenica je da je za svaku promjenu potrebna prilagodba kako bi tvrtka mogla opstati na tržištu. Prelazak korisnika na korištenje usluga konkurenata doslovno preko noći veliki je problem s kojim se bore sve telekomunikacijske tvrtke diljem svijeta. S obzirom na to da je ulaganje u marketing za privlačenje novih klijenata puno skuplji, razvila se metoda za predviđanje odlaska klijenata kako bi se što više odlazaka spriječilo.

4.2. Skup podataka

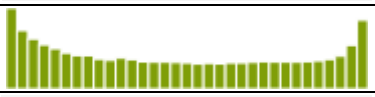
Skup podataka korišten u ovom radu na engleskom je jeziku i sastoji se od 7043 instance i 21 varijable. Prvi atribut je customerID koji govori koliko instanci korisnika postoji. Atribut gender pokazuje kojeg je korisnik spola te za razliku od customerID-a, puno je korisniji za izradu modela. Pomoću njega može se saznati koliku ulogu spol igra prilikom donošenja odluke promjene davatelja usluge. Pomoću atributa SeniorCitizen davatelji usluge mogu preciznije napraviti ponudu kada posjeduju informacije u kakvoj su dobi korisnici na određenom području njihova tržišta. Atributi Partner i Dependents pokazuju jesu li osobe koje su same i ne skrbe ni o kome sklone promjeni davatelja usluga. Atribut tenure govori koliko je dugo klijent već korisnik usluga u tvrtki. Atributi koji pokazuju koje usluge korisnik koristi:

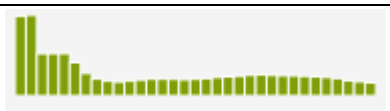
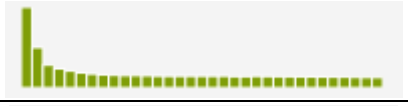
PhoneService, MultipleLines, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV i StreamingMovies. Atribut Contract govori koliko je korisnik pod ugovorom što pokazuje koliko je dugoročnu obvezu plaćanja usluge korisnik spreman potpisati.Atributi PaperlessBilling i PaymentMethod pokazuju kakvom su obliku plaćanja klijenti skloni. Atributi MonthlyCharges i TotalCharges pokazuju koliko korisnik u određenom periodu ili ukupno plaća za usluge koje koristi. Posljednji atribut u skupu podataka je za izradu modela najzanimljiviji. Naime, on pokazuje koliko je klijenata napustilo određenu tvrtku, odnosno koliko ih je i dalje vjerno prvobitnoj tvrtki. Svi atributi i neka njihova svojstva (minimalna i maksimalna vrijednost, tip podatka, prosječna vrijednost i srednja vrijednost-mod) prikazani su u tablici koja slijedi.

Tablica 1: Prikaz atributa skupa podataka

Atributi	Tip podatka	Min. vrijednost	Max. vrijednost	Prosječna vrijednost	Srednje vrijednosti (mod)
CustomerID	Tekst				
Gender	Kategorijski				Male
SeniorCitizen	Kategorijski				0
Partner	Kategorijski				No
Dependents	Kategorijski				No
Tenure	Numerički	0	72	29	
PhoneService	Kategorijski				Yes
MultipleLines	Kategorijski				No
InternetService	Kategorijski				Fiber optic
OnlineSecurity	Kategorijski				No
OnlineBackup	Kategorijski				No
DeviceProtection	Kategorijski				No
TechSupport	Kategorijski				No
StreamingTV	Kategorijski				No
StreamingMovies	Kategorijski				No
Contract	Kategorijski				Month-to-month
PaperlessBilling	Kategorijski				Yes
PaymentMethod	Kategorijski				Electronic check
MonthlyCharges	Numerički	19	11.875	1.885	
TotalCharges	Numerički	19	867.245	36.769	
Churn	Kategorijski				No

Tablica 2: Distribucija atributa s pripadajućim histogramom

Atribut	Distribucija	Histogram
CustomerID	Uniformna	
Gender	-	
SeniorCitizen	Uniformna	
Partner	-	
Dependents	Uniformna	
Tenure	Uniformna	
PhoneService	Uniformna	
MultipleLines	Uniformna	
InternetService	Uniformna	
OnlineSecurity	Uniformna	
OnlineBackup	Uniformna	
DeviceProtection	Uniformna	
TechSupport	Uniformna	
StreamingTV	Uniformna	
StreamingMovies	Uniformna	
Contract	Uniformna	
PaperlessBilling	Uniformna	
PaymentMethod	Uniformna	

MonthlyCharges	Eksponencijalno pad	
TotalCharges	Eksponencijalno pada	
Churn	Uniformna	

5. Opis metoda za rudarenje podataka

Za što precizniju izradu modela za predviđanje potrebno je upotrijebiti različite metode. Upotreba metode klaster analize omogućuje grupiranje klijenata po određenim obilježjima. Pronalazak samog razloga odlaska omogućavaju metode poput stabla odlučivanja i neuronske mreže. Detaljnom primjenom i analizom rezultata svih triju metoda, može se doći do odgovora zašto klijenti odlaze, ali i do odgovora što tvrtke mogu učiniti kako bi spriječile što više odlazaka vlastitih klijenata. Spomenute metode detaljnije će biti opisane u sljedećim potpoglavljima.

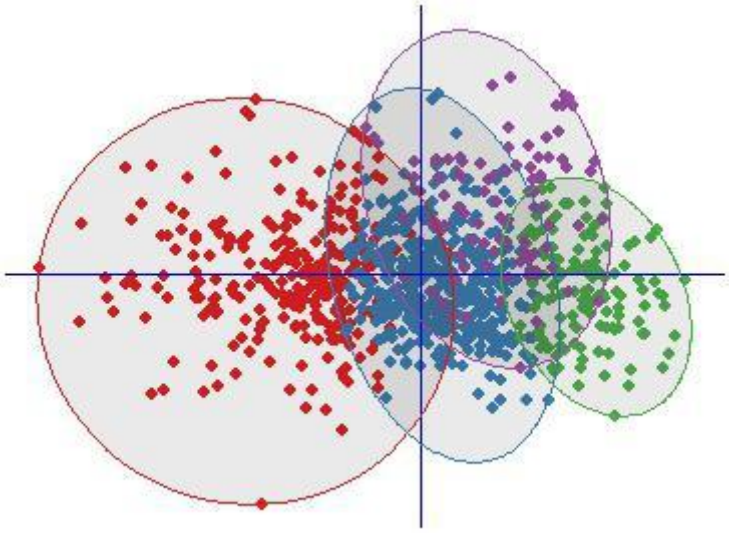
5.1. Klaster analiza

Klaster analiza ili klasterizacija je nenadgledano učenje. Ovu analizu čini skup tehnika kojima je cilj grupiranje objekata na osnovi njihovih atributa. Analiziraju se objekti, ali osobine objekata definiraju se pomoću varijabli. Za što precizniju analizu potrebno je uključiti samo one varijable koje su relevantne za istraživanje. [5] Prilikom proučavanja klastera, osim opisivanja pojedinog klastera, važno je proučiti i opisati sve ostale kako bi se izvukle sličnosti i razlike među klasterima [6, str. 199].

Prema Hair et al., 2010. (citirano u [5]) procedura klaster analize odgovara na tri važna pitanja: kako mjeriti sličnost između objekata, kako formirati klastere i kako utvrditi konačan broj klastera, a samu proceduru analize čini šest koraka (određivanje ciljeva, istraživačkog obrasca i pretpostavki, formiranje i procjena broja klastera, interpretacija klastera te procjena klaster analize i profiliranje klastera).

Prednost klaster analize je reduciranje podataka i informacija iz cijele populacije i svođenje značajki populacije na značajke reprezentativne skupine bez gubitka informacija dok nedostatak predstavlja manjak statističkog temelja jer se koriste samo jednostavni statistički postupci umjesto standardnih statističkih rezoniranja (pr. statistički značaj) [5].

Ova metoda koristi se u brojnim područjima kao što su marketing, astronomija, studije potresa i genetika.



Slika 1: Primjer prikaza klaster analize
(Izvor: https://hr.puntomarinero.com/cluster-analysis-its-method-and/)

5.2. Stablo odlučivanja

Stablo odlučivanja (eng. Decision Tree) je nadgledana tehnika učenja koja služi stvaranju modela za predviđanje vrijednosti ciljne varijable ovisno o ulaznim vrijednostima. Grafički prikazuje ishod na način da korijeni stabla predstavljaju testove i atribute, grane prikazuju rezultate ispitivanja, a listovi predstavljaju distribucije klasa [7].

Najvažniji zadatak stabla odlučivanja je prikaz svih mogućnosti i definiranje samog problema odlučivanja. Najčešće se primjenjuje prilikom donošenja odluka u pojedinim rizičnim situacijama koje su povezane s poslovnim svijetom. Sastoji se od niza odluka koje su međusobno povezane i svaka ovisi o prethodnoj [8, str.17].

Nayab i Scheid [citirano u 8, str.21] navode kako su prednosti stabla odlučivanja transparentnost, elastičnost i jednostavno korištenje, dok su nedostaci složenost, nezgrapnost, potrebno obrazovanje, troškovi te previše informacija.

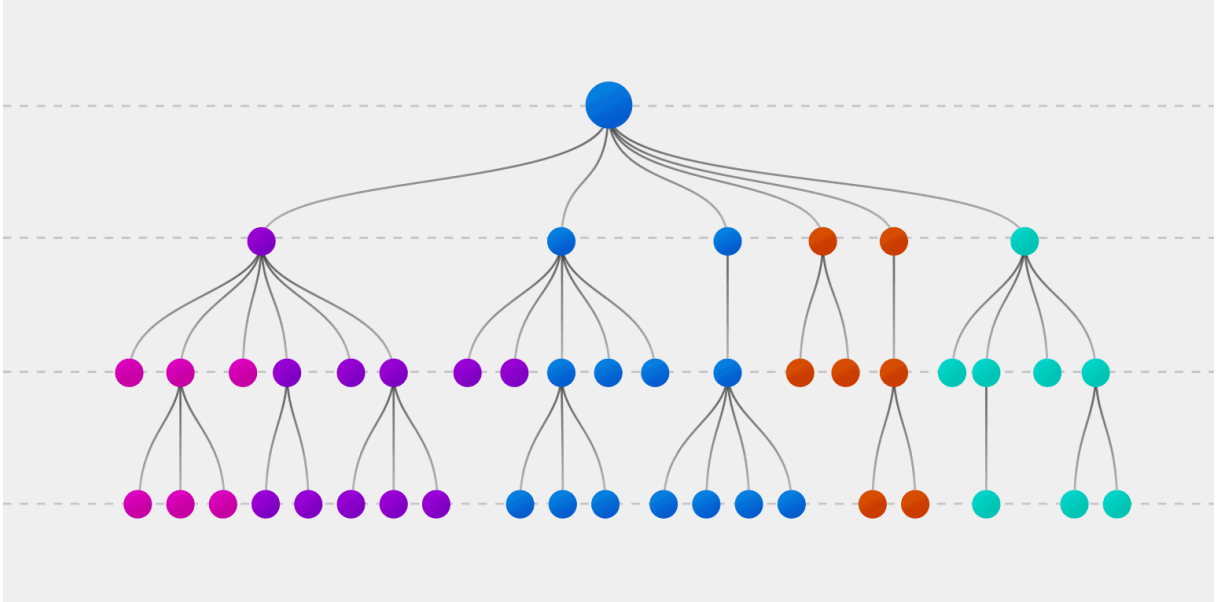
5.2.1. Prednosti

- Transparentnost olakšava samu usporedbu različitih mogućnosti s obzirom na to da prikazuje sve mogućnosti detaljno od njihovog početka do završetka.
- Koncentriranjem na same odnose između događaja stablu odlučivanja može se pripisati osobina elastičnosti koja mu omogućava da bude korišten kao dijagram utjecaja.
- Pregledan grafički prikaz omogućava jednostavnost korištenja.

5.2.2. Nedostaci

- U jednostavnim primjerima kreiranje i prikaz stabla odlučivanja je vrlo jednostavan i pregledan, no kada se radi o primjeru s velikim brojem odluka koje se u stablu manifestiraju kao veliki broj grananja, tada je kreiranje stabla odlučivanja dugotrajno i samim time prikazuje kompleksnost stabla.
- S obzirom na to da je ponekad potrebno raditi sa stablom odlučivanja koje je izuzetno kompleksno, potrebna je određena razina obrazovanja koja automatski ovu metodu čini skupom i iz tog razloga nije često primjenjivana.
- Dostupnost velike količine informacija odjednom u isto vrijeme može biti prednost, ali još veća mana. Naime, donositelj odluke mora imati pregled svih informacija kako bi na temelju njih donio najprikladniju odluku, što znači da veća količina informacija može donijeti veći broj nedoumica. Kako bi se nedoumice riješile, potrebno je proučavanje i pregled svih dostupnih informacija koje rezultiraju gubitkom vremena prilikom procesa donošenja odluka.

Stablo odlučivanja primjenjuje se u područjima kao što su ekonomija (uvođenje smjena, prekovremenih u odnosu na potražnju), poljoprivreda (prodaja/skladištenje), građevina (kupnja novog stroja).



Slika 2: Primjer prikaza stabla odlučivanja

(Izvor: https://www.explorium.ai/blog/the-complete-guide-to-decision-trees/)

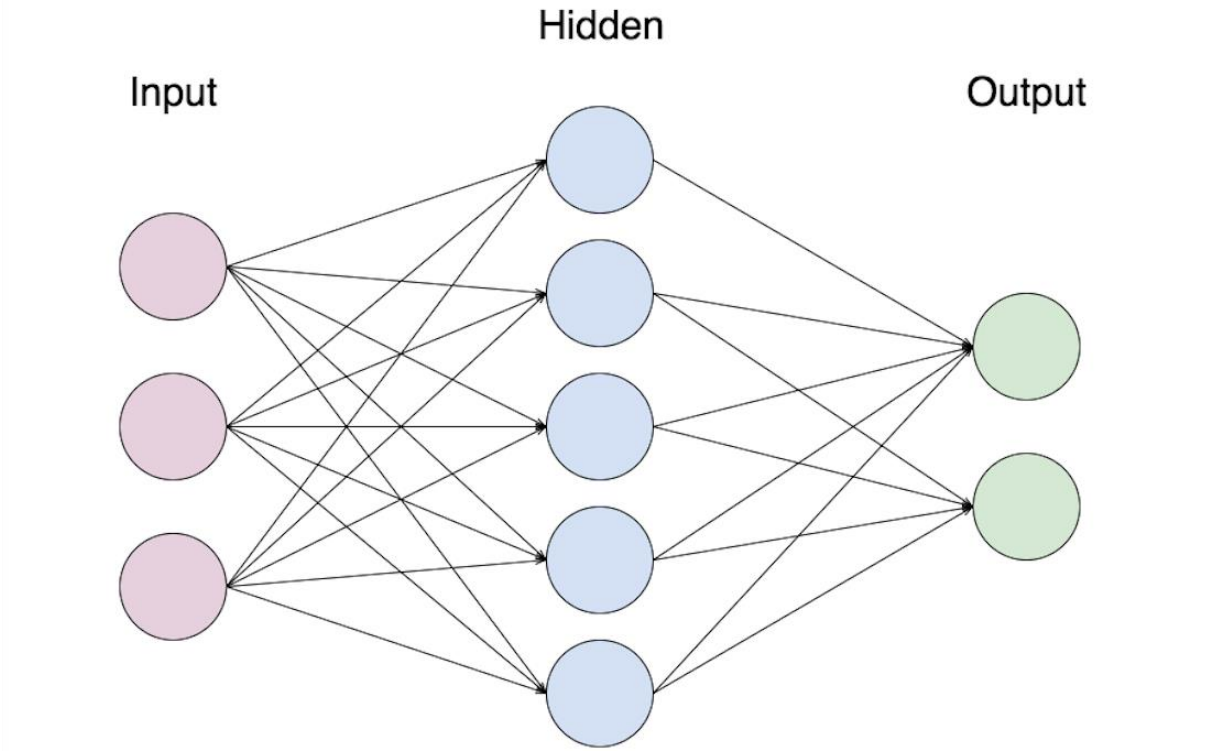
5.3. Neuronska mreža

Neuronske mreže poznate su i kao Duboko Učenje (eng. Deep Learning), ovisno o broju skrivenih slojeva [7]. Započele su 1943. godine kao linearni modeli uvođenjem modela koji je oponašao ponašanje neurona McCullocha-a i Pitts-a. Neuronske mreže su iterativni učenici za razliku od algoritama kao što je linearna regresija koji uče koeficijente tijekom samo jednog koraka obrade [6, str. 241].

Obično se umjetne neuronske mreže sastoje od hijerarhije slojeva u kojima su uzduž raspoređeni neuroni. Ulazne i izlazne slojeve čine neuroni koji su povezani s okolinom. Prilikom izrade neuronske mreže potrebno je prvo odlučiti koliko je neurona potrebno za korištenje i kako će ti neuroni biti povezani [9, str.5].

Prednosti neuronske mreže su slojevi koji omogućuju da se rezultat pojedinog sloja dodatno obrađuje te na taj način stvara kompleksni sustav.

S druge strane, kao nedostaci mogu se navesti osjetljivost na početne vrijednosti, algoritmi treniranja su dugotrajni te tako ne osiguravaju konvergenciju.



Slika 3: Primjer prikaza neuronske mreže

(Izvor: https://towardsdatascience.com/classical-neural-network-what-really-are-nodes-and-layers-ec51c6122e09)

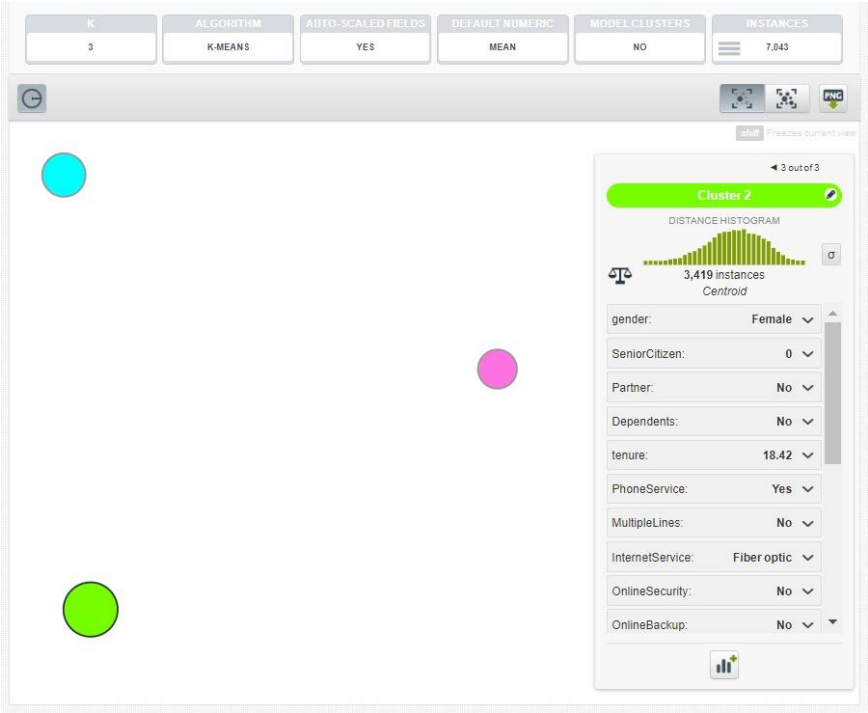
6. Modeliranje podataka

Nad odabranim skupom podataka provele su se sljedeće metode:

1. Klaster analiza
2. Stablo odlučivanja
3. Neuronska mreža

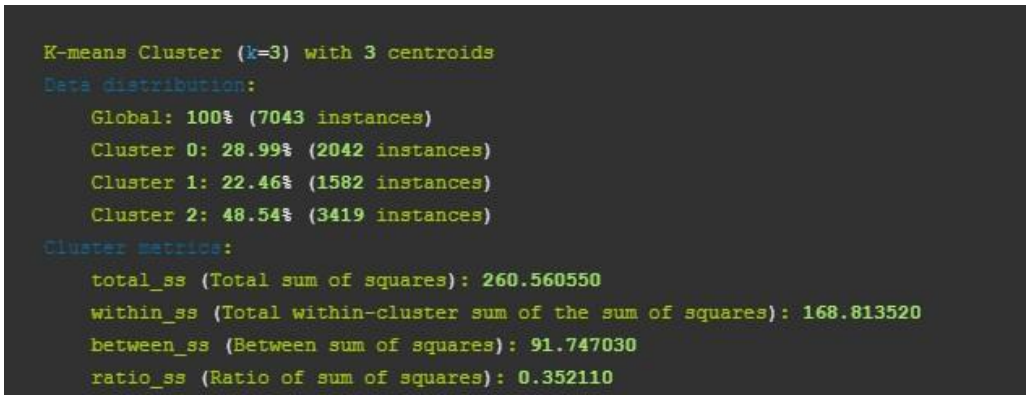
6.1. Klaster analiza

Optimalan broj klastera variranjem za odabrani skup podataka je tri. Prvi klaster koji se proučava predstavlja grupa žena koje su same, odnosno nemaju partnera. Drugi klaster predstavlja grupa muškaraca koji su sami, odnosno nemaju partnera, dok treću grupu čine muškarci i žene koji nisu sami, odnosno imaju partnera. Ovaj model može se nazvati deskriptivnim modelom jer se pomoću njega i različitih grupacija podataka opisuju skupovi podataka. Opisani podaci iz ovog modela pomažu u daljnjem predviđanju i izradi prediktivnih modela.



Slika 4: Optimalan broj klastera

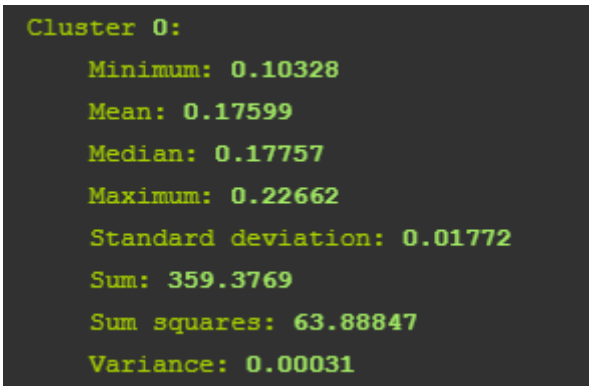
Za svaki klaster, može se prikazati izvještaj koji daje detaljniji prikaz podataka o klasteru. Osim zasebnih izvještaja o pojedinom klasteru, postoji i izvještaj koji prikazuje podatke o sva tri klastera odjednom koji je prikazan na sljedećoj slici.



Slika 5: Izvještaj o klasterima

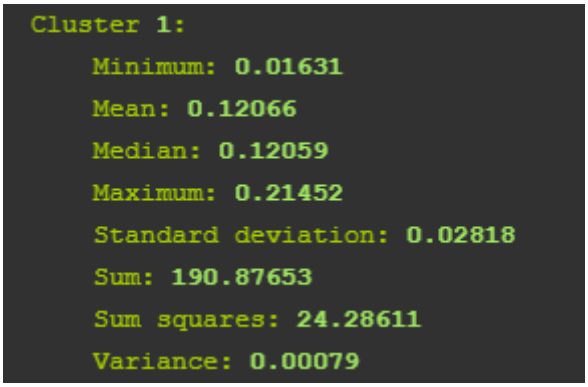
Na temelju aritmetičke sredine, instance se formiraju u klastere, što znači da svaki klaster sadrži instance koje međusobno imaju približne aritmetičke sredine. Na temelju podataka iz izvještaja vidljivo je da je najveći klaster zadnji u nizu, odnosno klaster 2. Sastoji se od 3419 instance što znači da je 3419 (čine 48.54% ukupne populacije ispitanika) ispitanika muška ili ženska osoba s partnerom. Najmanji klaster sastoji se od 1582 instance (čine 22.46% ukupne populacije ispitanika) i to je drugi klaster u nizu, odnosno klaster 1 kojeg čine muškarci bez partnera. Klaster 0 čini 28.99% ukupne populacije ispitanika i njega čine žene bez partnera. Što se tiče same kvalitete klastera, može se reći da je zadovoljavajuća s obzirom na činjenicu u prikazu klastera s optimalnim brojem klastera (tri) nema pojedinih vrijednosti koje strše. Standardna devijacija predstavlja prosječno srednje kvadratno odstupanje numeričkih vrijednosti od njihove aritmetičke sredine [10], dok je varijanca mjera disperzije veličina [11].

Detaljniji izvještaji o pojedinom klasteru i njihovim vrijednostima prikazani su na sljedećih nekoliko slika.



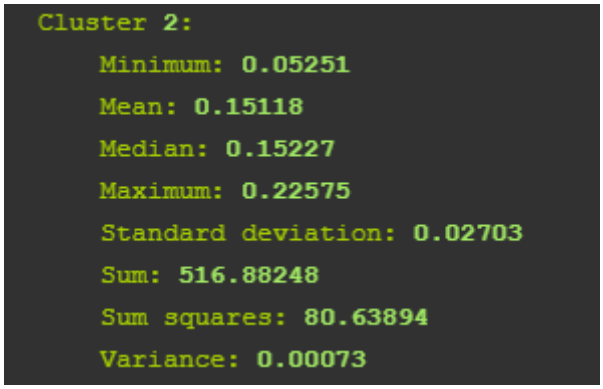
Slika 6: Izvještaj o klasteru 0

Iz izvještaja o klasteru 0 vidljivo je da je razlika između minimuma i maksimuma velika, odnosno da je maksimum i više nego dvostruko veći (Minimum: 0.10328, Maximum: 0.22662). Srednja vrijednost iznosi 0.17599, dok medijan iznosi 0.17757 iz čega se može zaključiti da razlika nije velika. Standardna devijacija iznosi 0.01772, dok varijanca iznosi 0.00031.



Slika 7: Izvještaj o klasteru 1

Iz izvještaja o klasteru 1 vidljivo je da je razlika između minimuma i maksimuma puno veća, nego što je to u slučaju kod klastera 0. Ovdje minimum iznosi 0.01631, dok maksimum iznosi 0.21452. Srednja vrijednost iznosi 0.12066, a medijan 0.12059 iz čega se može zaključiti da razlika između srednje vrijednosti i medijana nije velika s obzirom na to da se razlikuju tek u četvrtoj decimali. Standardna devijacija iznosi 0.02818, dok varijanca iznosi 0.00079.



Slika 8: Izvještaj o klasteru 2

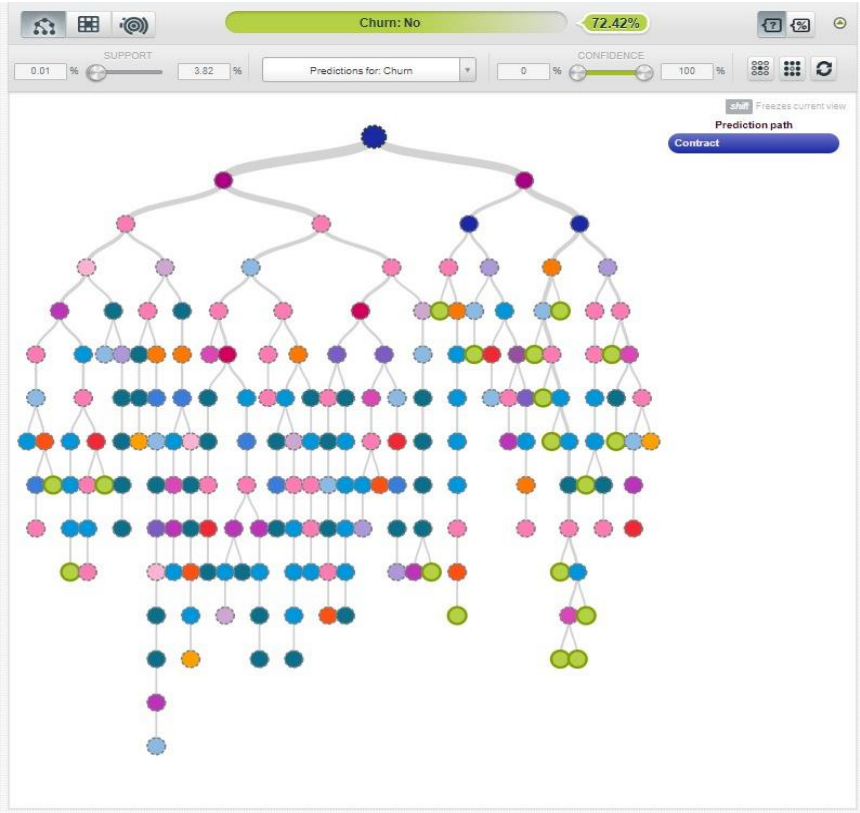
Iz izvještaja o klasteru 2 vidljivo je da je razlika između minimuma koji iznosi 0.05251 i maksimuma koji iznosi 0.22575 velika, no ne toliko koliko je u klasteru 1. Srednja vrijednost i medijan ne razlikuju se toliko, no opet dovoljno da bi razlika bila vidljiva. Srednja vrijednost

ovdje iznosi 0.15118, dok medijan iznosi 0.15227, što znači da se razlikuju već u trećoj decimali. Standardna devijacija iznosi 0.02703, dok varijanca iznosi 0.00073.

6.2. Stablo odlučivanja

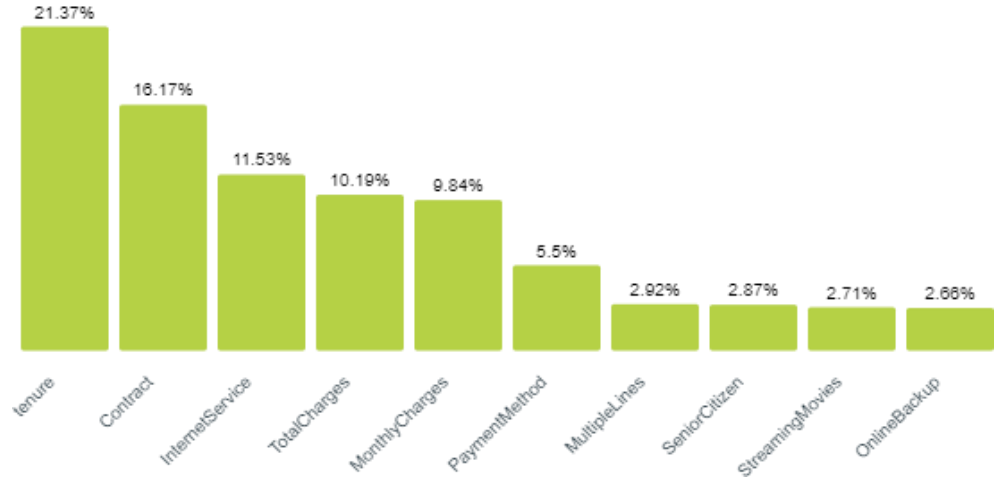
Pri izradi stabla odlučivanja, kao zavisni atribut koristi se atribut Churn koji predstavlja odlazak odnosno ostanak u telekomunikacijskoj tvrtki.

Veličina stabla odlučivanja očituje se kroz njegovo grananje. U ovom slučaju, veličina stabla jednaka je broju grananja koje iznosi 45. Pri vrhu stabla, kao početni atribut stoji atribut Contract koji daje informacije o tome koliko je korisnik pod ugovorom s telekomunikacijskom tvrtkom zatim se odvijaju različiti scenariji grananja s ostalim atributima.

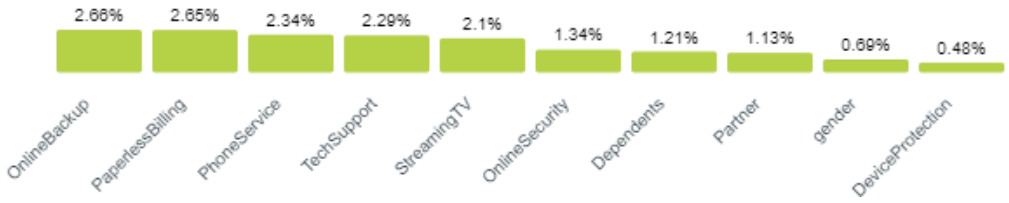


Slika 9: Stablo odlučivanja

Prilikom kreiranja stabla odlučivanja mogući je prikaz atributa po njihovoj važnosti. Konkretni postoci važnosti atributa za prethodno spomenuto stablo odlučivanja biti će prikazani u sljedeće dvije slike.



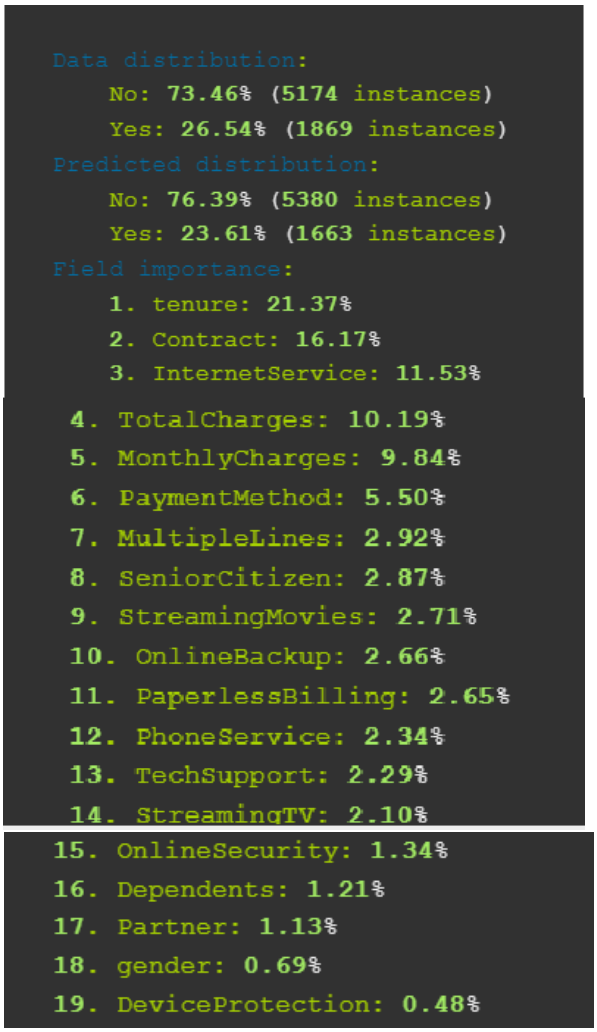
Slika 10: Važnost atributa – 1.dio



Slika 11: Važnost atributa – 2.dio

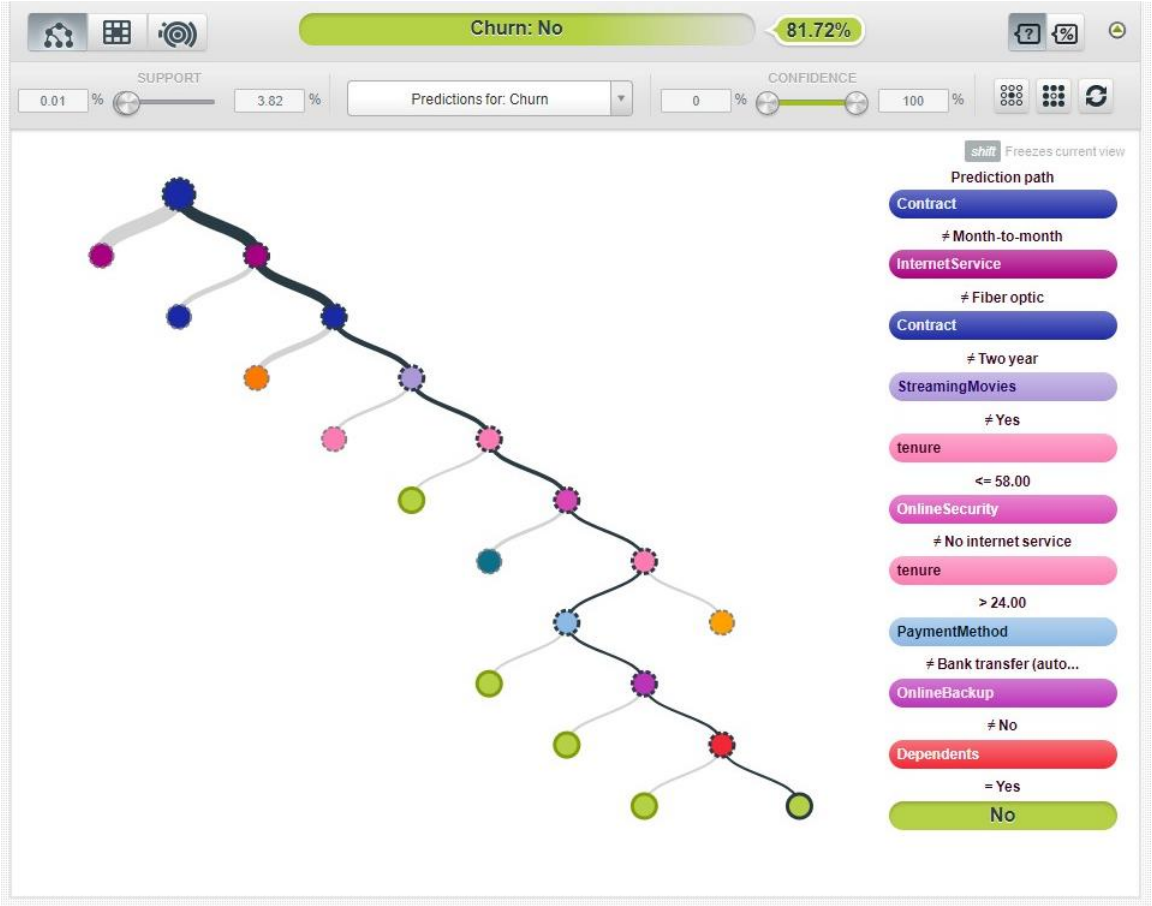
Prethodne dvije slike prikazuju važnost atributa iz skupa podataka po postocima. Jasno je vidljivo da prvih šest atributa koji su prikazani sa svojim postotkom odskaču od ostatka atributa na prvoj slici te svih ostalih atributa na drugoj slici. Bez imalo sumnje, pogledom na dijagram odmah je vidljivo da je najvažniji atribut u skupu tenure (s važnošću od 21.37%) koji govori koliko je klijent već dugo korisnik telekomunikacijskih usluga u tvrtki. Nakon tenura, sljedeći po važnosti je atribut Contract (s važnošću od 16.17%) koji govori pod kakvom je ugovornom obvezom klijent, odnosno koliko dugo je pod ugovornom obvezom. Nakon toga slijede atributi InternetService, TotalCharges i MonthlyCharges (s važnostima od 11.53%, 10.19% i 9.84%) koji govore koristi li korisnik internetsku uslugu, koliki su ukupni i mjesečni troškovi. Vidljivo je da su ta tri atributa po važnosti moglo bi se reći slični jer ne odstupaju toliko

jedan od drugog po postotku važnosti. Zadnji atribut koji odstupa od ostatka po postotku važnosti je PaymentMethod (s vrijednošću od 5.5%) koji govori o načinu na koji klijent plaća uslugu. Ostalim atributima pada postotak važnosti od 2.92% do 0.48%.



Slika 12: Detaljan izvještaj o kreiranom stablu odlučivanja

Na slici iznad prikazan je izvještaj o stablu odlučivanja koji osim postotaka važnosti atributa prikazuje i postotke o stvarnoj distribuciji podataka i postotke predviđene distribucije. S obzirom na to da je kao zavisni atribut u stablu odlučivanja uzet atribut Churn, u distribuciji podataka promatraju se postoci odgovora klijenata vezana uz taj atribut. Čak 73.46% ispitanika odgovorilo je „No“ što znači da 5174 ispitanika nije napustilo tvrtku čije su telekomunikacijske usluge koristili, dok 1869 ispitanika odnosno 26.54% jest.



Slika 13: Primjer pravila odabirom pojedine grupe

Nakon kreiranja stabla odlučivanja, postoji mogućnost pregleda određenog pravila. Pregled pravila dobije se na način da se odabere pojedina grupa, što znači da od početka stabla, na svakom grananju, odabire se jedna grana odnosno jedan put kojim se ide i tako na svakom sljedećem grananju sve dok se ne dođe do kraja stabla.

Pouzdanost pravila prikazanog na slici iznad:

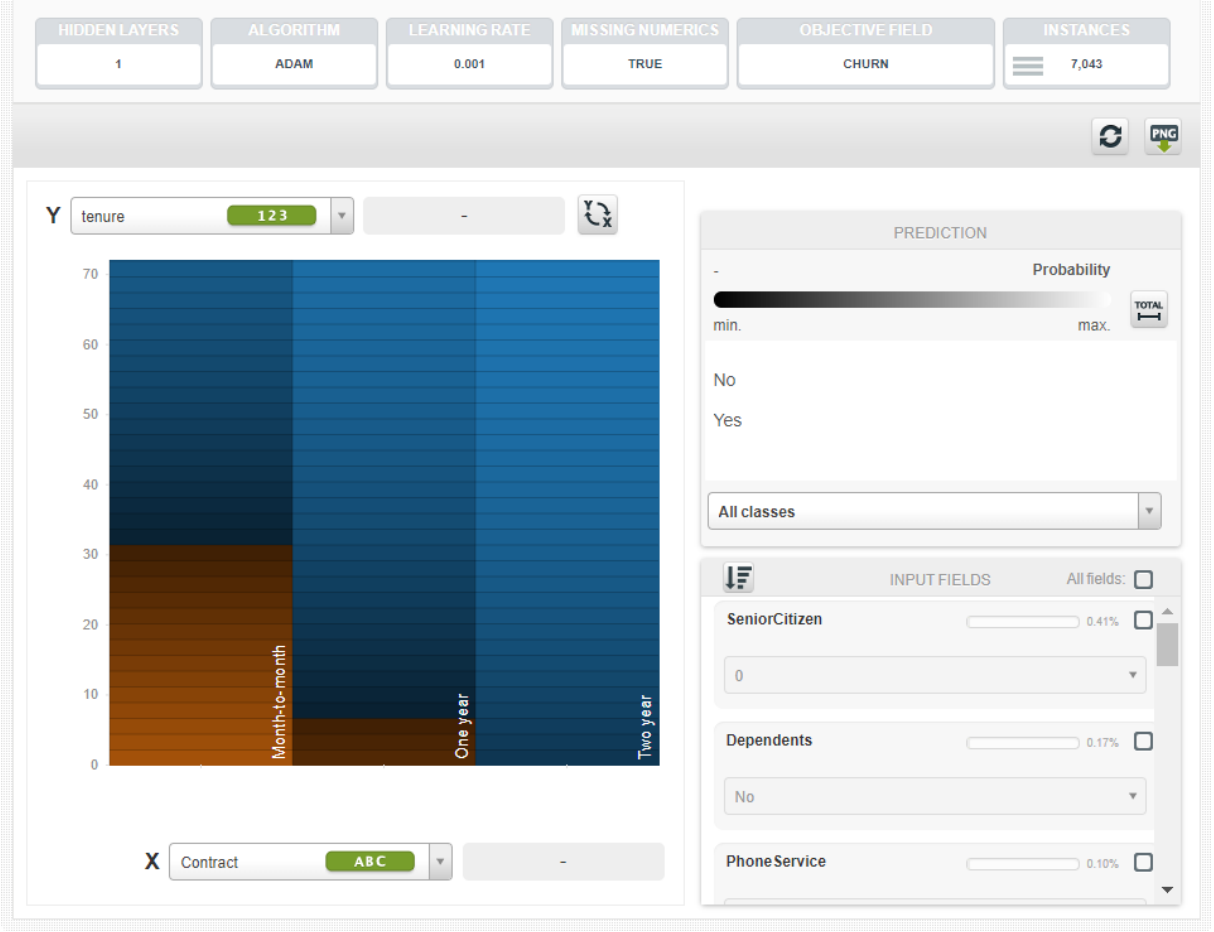
Contract != Month-to-month and InternetService !=Fiber optic and Contract != Two year and StreamingMovies != Yes and tenure <=58.00 and OnlineSecurity !=No Internet service and tenure >24 and PaymentMethod != Bank transfer and Online Backup != No and Dependents = Yes and No

Odnosno ako ugovorna obveza nije od-mjeseca-do-mjeseca (što znači da klijent potpisuje ugovor na duže vremensko razdoblje) i ako internetska usluga nije putem optičkih vlakana i ako ugovorna obveza nije na dvije godine i ako klijent ne koristi uslugu streaming filmova i ako klijent koristi usluge tvrtke manje od 58 mjeseci i koristi internetsku zaštitu i ako je klijent korisnik usluga tvrtke više od 24 mjeseca i ako kao oblik plaćanja ne koristi bankovne

transakcije i ako koristi sigurnosnu kopiju na mreži i ako ima članove obitelji o kojima brine tada je odgovor za prebacivanje na korištenje usluga druge tvrtke „No“.

6.3. Neuronska mreža

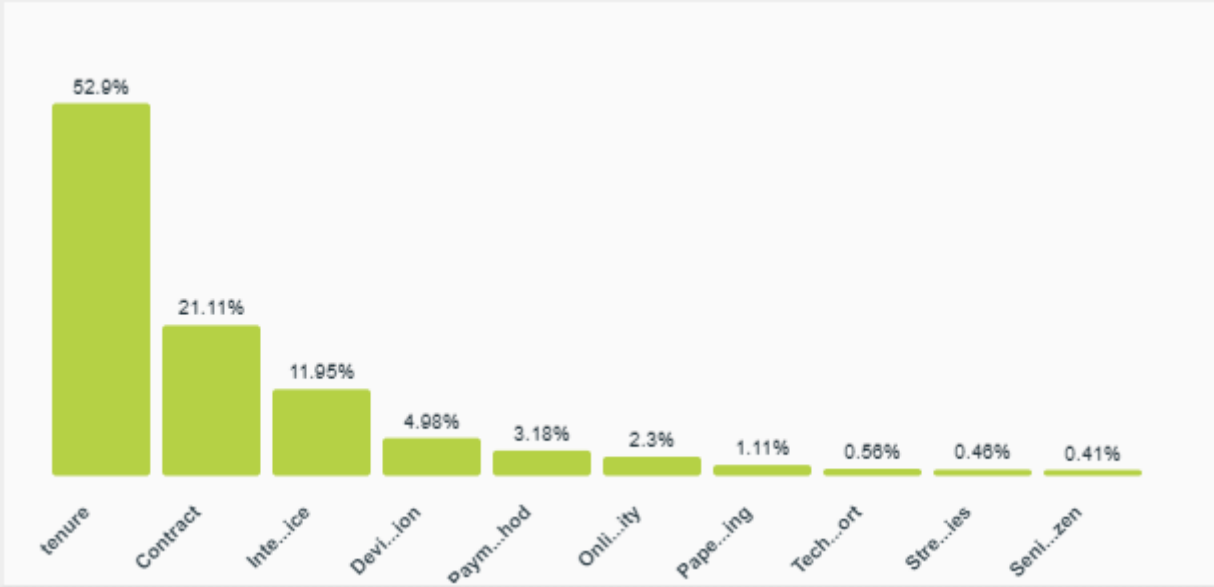
Za izradu modela neuronske mreže korišten je samo jedan skriveni sloj. Korištenje jednog skrivenog sloja karakteristika je ADAM (eng. Adaptive Moment Estimation) algoritma. ADAM algoritam skalira faktor učenja na način da se za svaki parametar računaju eksponencijalni prosjeci za diferencijale i kvadrate diferencijala [12, str. 34].



Slika 14: Postavke neuronske mreže

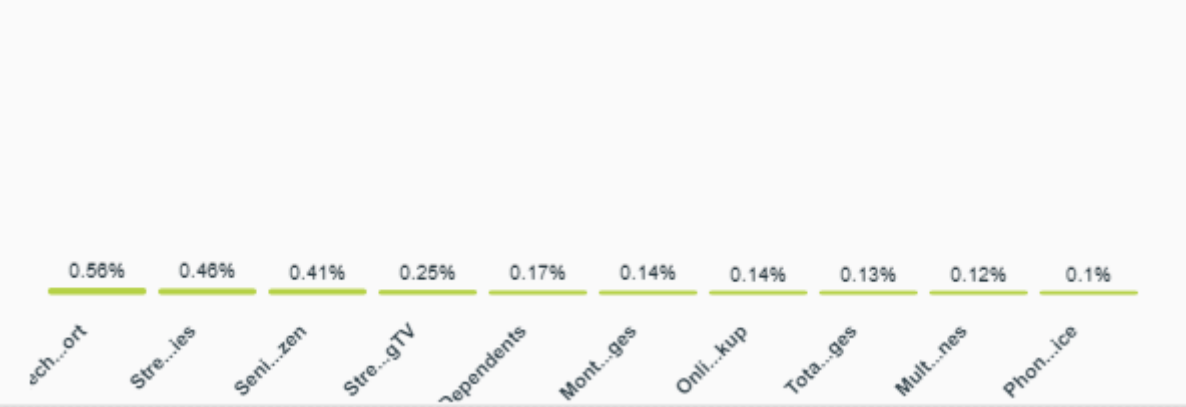
Detaljnijim proučavanjem neuronske mreže s prethodne slike, vidljive su mnoge informacije. Na X os stavljen je atribut Contract koji pokazuje vrstu ugovorne obveze, a na Y os stavljen je atribut tenure koji govori koliko mjeseci je klijent korisnik usluga tvrtke. Smeđim

nijansama označena su područja vrijednosti atributa za koje se predviđa da će više od 50% klijenata prijeći u drugu tvrtku. Dakle, predviđa se da će klijenti koji potpisuju ugovor od-mjeseca-do-mjeseca i koji su korisnici usluge do 30 mjeseci te klijenti koji potpisuju ugovor na godinu dana i korisnici su 10 mjeseci, prijeći u drugu tvrtku odnosno prestati koristiti usluge prvobitne tvrtke. Može se zaključiti kako su ugovorne obveze i dugoročnost korištenja usluga međusobno povezane na način da klijenti koji potpisuju ugovorne obveze na duži period, skloni su nastavku korištenja usluga u istoj tvrtki, umjesto prelaska u novu. Odnosno, klijenti koji potpisuju ugovore na kraće vremenske periode i kraće su korisnici usluge u tvrtki, skloni su odluci prelaska na korištenje usluga neke druge tvrtke.



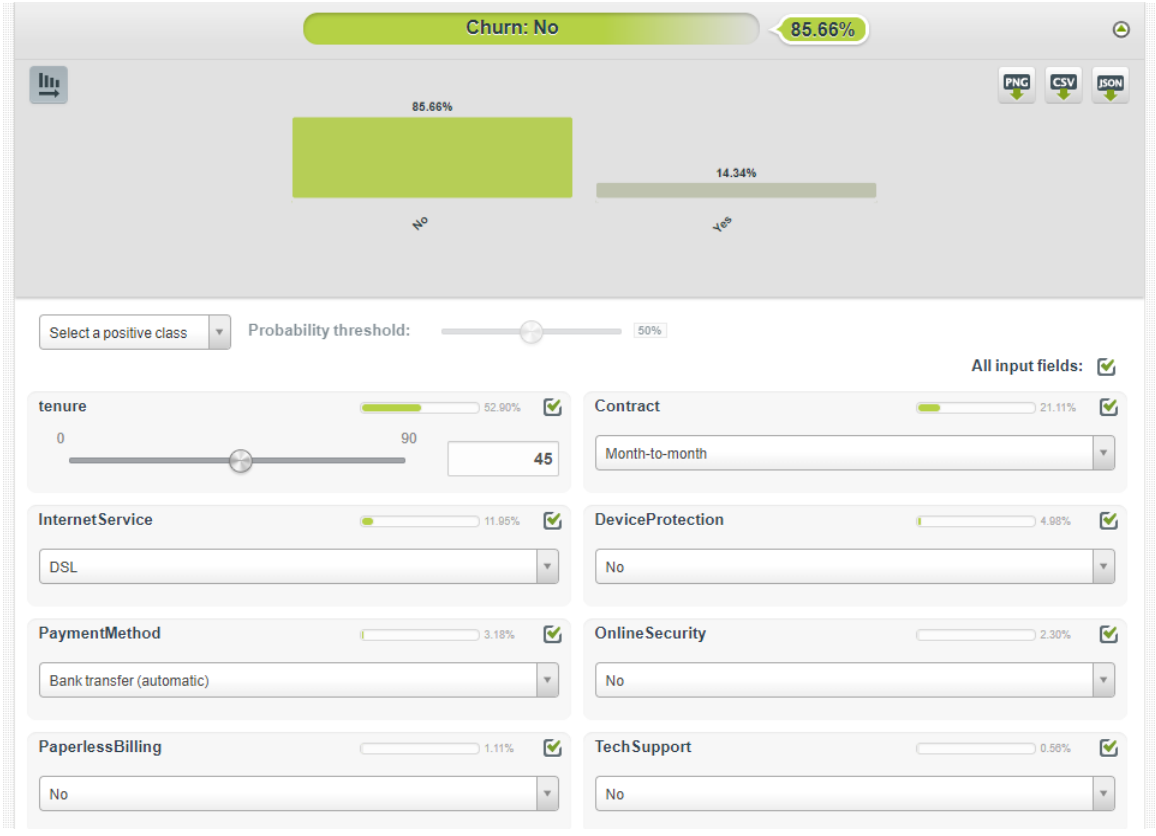
Slika 15: Važnost atributa – 1.dio

Na slici iznad prikazana je važnost atributa s njihovim postocima važnosti u obliku dijagrama. Kao tri atributa od najveće važnosti mogu se izdvojiti atributi tenure, Contract i Internet Service odnosno atributi koji pokazuju podatke o tome koliko dugo je klijent korisnik usluga tvrtke, kakav ugovor potpisuje i koristi li internetsku uslugu. Sva tri atributa uvelike odskaču od ostatka sa svojih 52.9%, 21.11% i 11.95%. Sljedeća četiri atributa po važnosti su DeviceProtection, PaymentMethod, OnlineSecurity i PaperlessBilling čiji postoci važnosti se protežu od 4.98% i padaju do 1.11%. Za ostale atribute može se reći da u ovom modelu nisu od velike važnosti obzirom na to da njihovi postoci važnosti padaju ispod 1%.



Slika 16: Važnost atributa – 2.dio

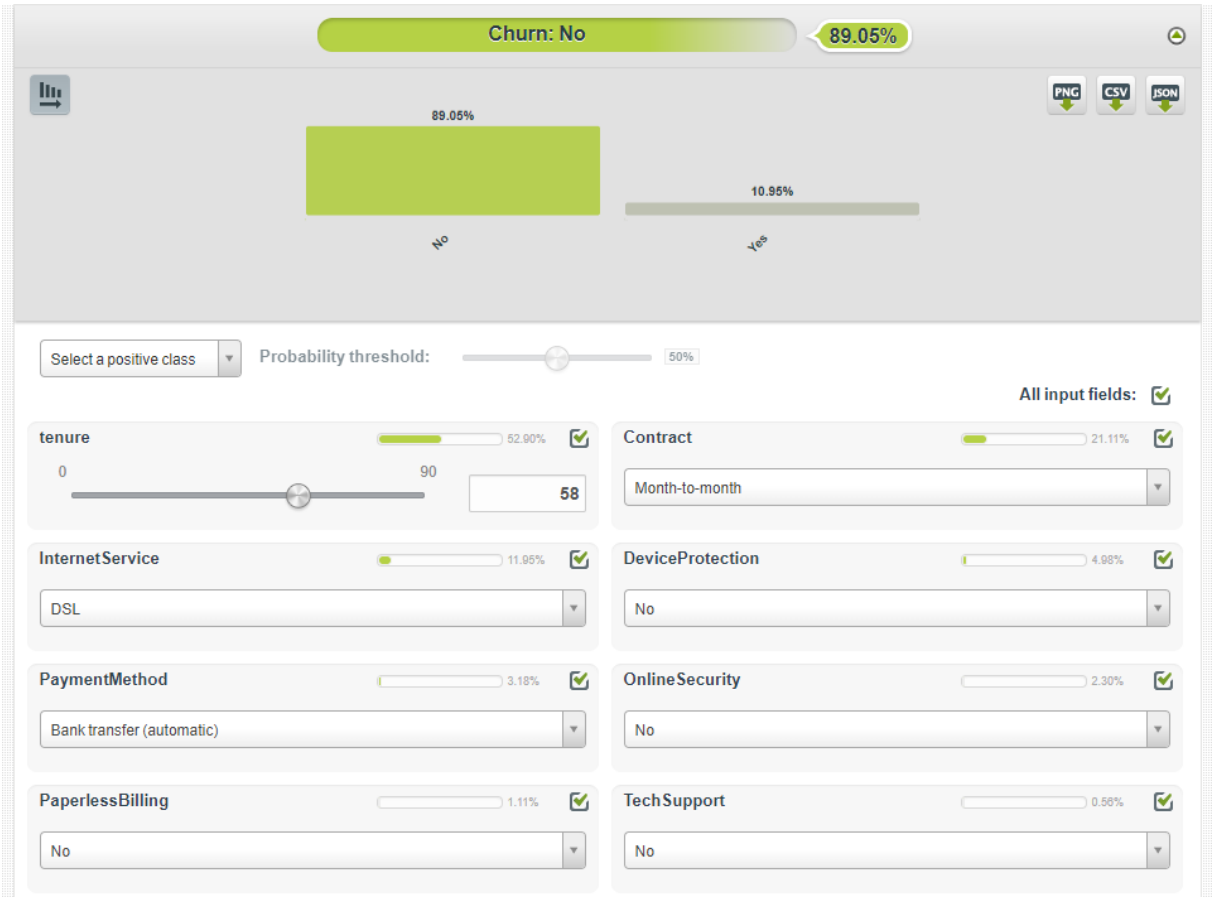
Kao i zadnjih nekoliko atributa na prvoj slici, svi atributi na slici 16 nisu od velike važnosti jer je postotak svakog pojedinog atributa ispod 1%.



Slika 17: Predviđanje odlaska klijenta - 1

Uzme li se za atribut tenure vrijednost 45 (klijent koristi usluge tvrtke 45 mjeseci), internetska usluga DSL, način plaćanja automatske bankovne transakcije, fakturiranje nije bez papira, ugovor od-mjeseca-do-mjeseca i da se pritom ne koriste usluge poput tehničke

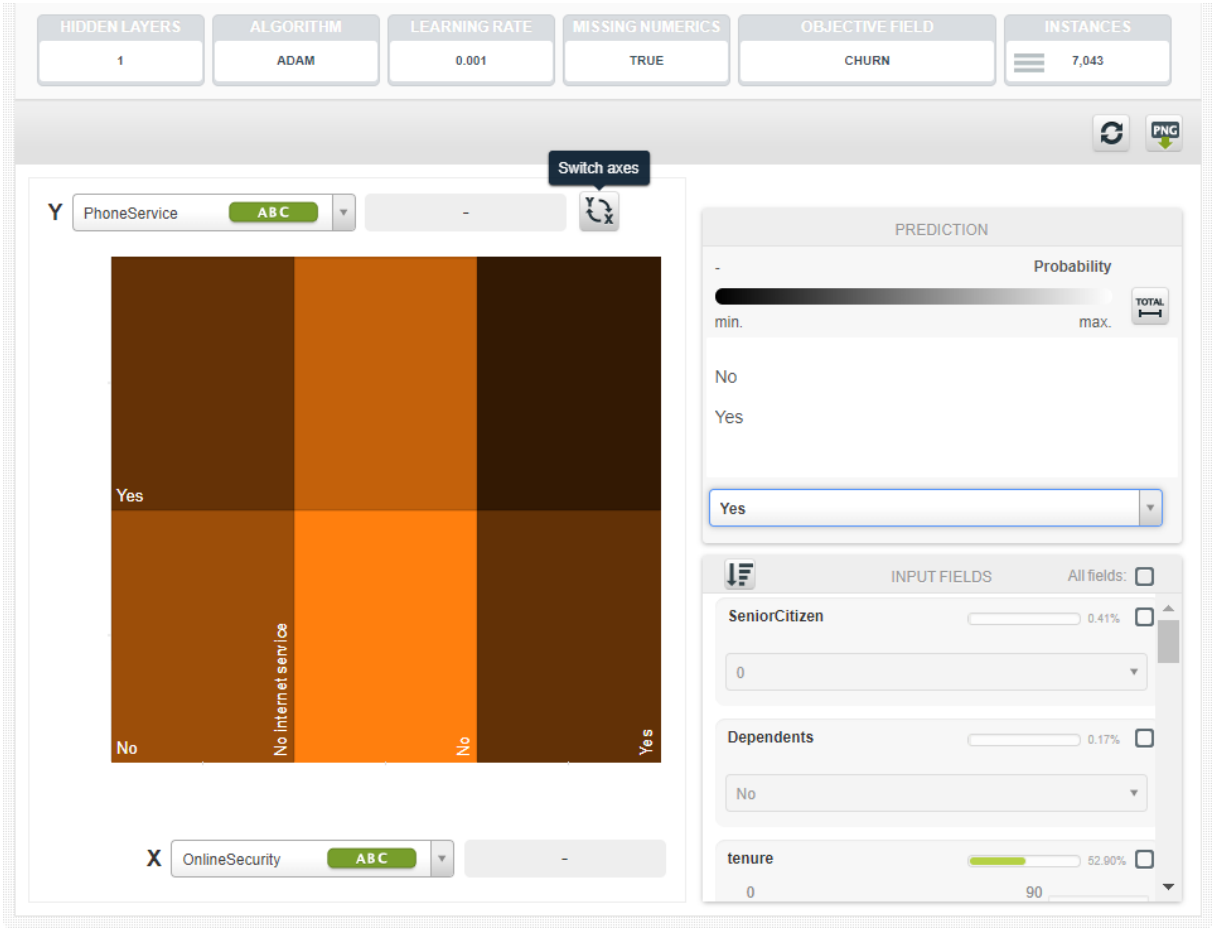
podrške, internetske sigurnosti, zaštite uređaja, streaming filmova, više telefonskih linija i ostale. Pritom da je klijent ženska osoba koja nema partnera, nije starija osoba i nema članove za koje, tada bi 85.66% klijenata trebalo reći da se neće prebaciti na korištenje usluga druge tvrtke.



Slika 18: Predviđanje odlska klijenta - 2

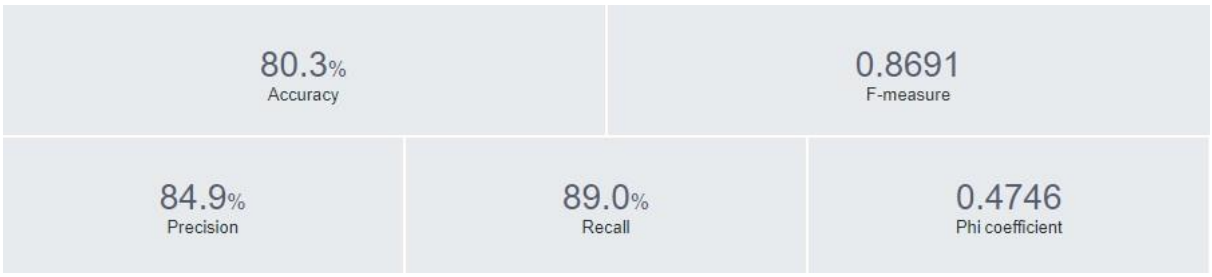
Ostave li se sve vrijednosti atributa jednake kao u prošlom primjeru (slika 17), a promijeni li se samo vrijednost atributa tenure koja govori koliko mjeseci je klijent korisnik usluga tvrtke tako da se postavi na 58 mjeseci, postotak odgovora se mijenja. U tom slučaju 89.05% klijenata trebalo bi reći da se neće prebaciti na korištenje usluga neke druge tvrtke.

U ovom modelu moguće je vidjeti kako vrijednosti ulaza i izlaza ovise jedne o drugima, na slici koja slijedi prikazan je primjer grafičke ovisnosti ulaza i predviđene vrijednosti izlaza. Na Y os stavljen je atribut telefonske usluge, dok je na X os stavljen atribut internetske sigurnosti.



Slika 19: Predviđanje vrijednosti izlaza

Obzirom na to da su ulaz i izlaz povezani, povećanjem telefonske usluge povećava se i internetska sigurnost što je vidljivo na grafičkom prikazu, također vrijedi obrat odnosno povećanjem internetske sigurnosti slijedi povećanje telefonske usluge. Krene li se promatrati kada su klijenti skloniji odlasku iz telekomunikacijske tvrtke uspoređujući vrijednosti atributa telefonske usluge i internetske sigurnosti, može se zaključiti da je klijentima lakši odlazak pada kada ne koriste niti telefonske usluge niti internetsku sigurnost (čak 37.98%). Najmanje šanse za odlazak klijenata iz tvrtke su kada je klijent korisnik telefonskih usluga kao i internetske sigurnosti (19.83%). Veće su šanse odlaska klijenta kada je korisnik telefonske usluge, a ne internetske sigurnosti (29.54%), nego kada je korisnik internetske sigurnosti, a ne telefonske usluge (26.21%).



Slika 20: Točnost modela

S obzirom na činjenice da točnost modela iznosi 80.3% i preciznost 84.9% može se reći da je ovaj model poprilično pouzdan i točan.

7. Diskusija rezultata

Prva metoda koja je primijenjena je klaster analiza. Pomoću klaster analize instance odabranog skupa podataka podijeljene su u tri klastera. Svaki klaster predstavlja određenu grupaciju klijenata koji imaju iste atribute. Jednu grupu čine žene koje nemaju partnera. Drugu grupu čine muškarci koji nemaju partnera, a treću grupu čine klijenti koji su u paru, odnosno klijenti koji imaju partnera.

Što se tiče samih klastera, nakon grupacije, po veličinama nisu jednaki. Klaster s najvećim brojem instanci čini klaster s grupacijom klijenata koji imaju partnera (3419 instanci). Sljedeći klaster po veličini je klaster s grupacijom klijenata koje čine žene bez partnera (2042 instance). Najmanji klaster po veličini čini grupacija klijenata koju čine muškarci bez partnera (1582 instance). Uspoređuju li se vrijednosti atributa standardne devijacije, srednje vrijednosti, medijana i varijance, zanimljiv je poredak po veličini za svaku vrijednost. Usporede li se vrijednosti srednje vrijednosti i medijana, najveće vrijednosti ima grupacija žena bez partnera, zatim slijede klijenti koji imaju partnera pa zatim muškarci bez partnera. Dok s druge strane, usporede li se vrijednosti standardne devijacije i varijance, najveću vrijednost ima grupacija muškaraca bez partnera, zatim klijenti s partnerom pa tek onda žene bez partnera. Iz prethodne rečenice može se zaključiti da prema vrijednostima varijance i standardne devijacije, najviše odstupanja i disperzije ima u grupaciji muškaraca bez partnera.

Nakon primjene klaster analize primijenjena je metoda stabla odlučivanja. Kao zavisni atribut u stablu odlučivanja stavljen je atribut koji govori hoće li klijent otići iz tvrtke ili će ostati. Stablo se sastoji od 45 grananja što znači da je njegova veličina jednaka 45, a zavisnost stabla je 72.42%. Za svaki atribut u skupu podataka može se vidjeti kolika je njegova važnost u donošenju odluke u postocima. U ovom slučaju odluka za koju se gleda važnost određenog atributa je hoće li klijent otići iz tvrtke. Kao najvažniji atributi prilikom donošenja odluke mogu se uzeti atributi s važnošću iznad 10%, a to su atributi koji govore koliko dugo je klijent korisnik usluga u tvrtki (16.17%), kakvu ugovornu obvezu potpisuje (11.53%) i korištenje određenih internetskih usluga (11.53%). S obzirom na važnost atributa koji su imali ulogu u donošenju odluke odlaska, 73.46% ispitanika nije napustilo tvrtku čije usluge koriste (5174 ispitanika), dok ostatak (26.54%, 1869 ispitanika) jest.

Zadnja metoda koja je provedena je neuronska mreža. Kreirana je po ADAM algoritmu odnosno korišten je jedan skriveni sloj. Što se tiče važnosti atributa u ovoj metodi, najvažniji atributi po postocima su također atributi koji govore koliko dugo je klijent korisnik usluga tvrtke, kakvu ugovornu obvezu potpisuje te kakvu internetsku uslugu koristi. No, u ovom slučaju

atribut tenure ima važnost veću od 50% (čak 52.9%) što je kada usporedimo s postotkom važnosti atributa tenure u stablu odlučivanja više nego dvostruko. Ugovorna obveza i internetske usluge ovdje imaju važnost od 21.11% i 11.95%.

Primjenom ove metode predviđa se da će klijenti skloni odlasku biti oni klijenti koji potpisuju ugovor od-mjeseca-do-mjeseca i koji su korisnici usluga tvrtke do 30 mjeseci te oni klijenti koji imaju potpisanu ugovornu obvezu na jednu godinu, a sada su već korisnici usluga tvrtke 10 mjeseci. Kao odaniji klijenti predviđa se da su žene bez partnera. Naime, žene koje su već klijenti iste tvrtke 45 mjeseci daju postotak ostanka od 85.66%, a povećanjem mjeseci korištenja usluga tvrtke na 58 postotak se samo povećava na 89.05%, dok je kod slobodnih muškaraca manji postotak. Također, prelasku na korištenje usluga neke druge tvrtke više su skloni klijenti koji koriste samo telefonske usluge (bez internetskih usluga), dok najlakše odlaze klijenti koji nisu korisnici ni telefonskih niti internetskih usluga.

Detaljnim proučavanjem rezultata dobivenih primjenom svih triju metoda mogu se izvući zaključci (koje podupiru grafički prikazi) da se s vremenom koje klijent provede kao korisnik usluga neke određene tvrtke, razvije određena doza odanosti. Važno je napomenuti da klijenti koji su duže vrijeme klijenti u istoj tvrtki nisu voljni lako prijeći u drugu tvrtku i riskirati gubitak dosadašnjih pogodnosti zbog nečega nepoznatog što ih čeka u novoj tvrtki. S druge strane, klijenti koji nisu dugogodišnji klijenti nemaju problem s prelaskom jer nemaju razvijenu istu dozu povjerenja prema trenutnoj tvrtki čije usluge koriste kao što to imaju dugogodišnji klijenti.

8. Zaključak

Svako poduzeće u svijetu, kako veliko tako i malo, bori se s odlaskom svojih klijenata. Kako bi spriječili što više odlazaka važno je prvo saznati koji su razlozi iz kojih klijenti odlaze, odnosno čemu pridaju najviše važnosti prilikom donošenja odluke oko odlaska.

Iako bi možda prvobitan zaključak bio da je najvažnija stavka prilikom odluke odlaska ili ostanka, kvaliteta usluge, detaljnom analizom pokazalo se da najveću ulogu ipak igra vremenski period u kojem je klijent već dugo korisnik usluga neke tvrtke. Klijenti potpisivanjem ugovorne obveze na neko duže razdoblje razvijaju određenu razinu povjerenja prema svom operateru, a iz rezultata analize vidljivo je da su dugoročni korisnici manje skloni odlascima. Jedno od mogućih rješenja problema odlaska klijenata ili bar smanjenje broja odlazaka mogli bi biti ugovori na duži vremenski period s određenim pogodnostima koji privlače klijente te paket u kojem se povoljno nudi i telefonska i internetska usluga (rezultati analiza pokazuju da su klijenti koji koriste obje usluge manje skloni odlasku).

Popis literature

[1] „Telco Customer Churn“ (bez datuma) [na internetu] dostupno:

https://www.kaggle.com/blastchar/telco-customer-churn [pristupano 21.06.2021.]

[2] Andres Kuusik, Urmas Varblane, „How to avoid customers leaving: the case of the Estonian telecommunication industry“ (09.01.2009.) [na internetu] dostupno:

https://www.emerald.com/insight/content/doi/10.1108/17465260910930458/full/html [pristupano 21.06.2021.]

[3] Fintech, „4 Out of 5 of Customers Don't Think Banks Know What They Need“ (04.03.2019.) [na internetu] dostupno:

https://fintechnews.hk/8667/various/trust-bank-challenger-bank-customer-service/

[pristupano 21.06.2021.]

[4] TechSee, „Reasons for customer churn in the telecom industry: 2019 survey results“

(bez datuma) [na internetu] dostupno:

https://techsee.me/resources/surveys/2019-telecom-churn-survey/

[pristupano 23.06.2021.]

[5] Kristina Devčić, Ivana Tonković Pražić, Željko Župan „Klaster analiza: primjena u marketinškim istraživanjima“ (08.05.2012.) [na internetu] dostupno:

https://hrcak.srce.hr/file/124179 [pristupano 24.06.2021.]

[6] Dean Abbott, Applied Predictive Analytics Principles and Techniques for the Professional Data Analyst, USA:Wiley. 2014

[7] Uduak Idio Akpan, Andrew Starkey „Review of classification algorithms with changing inter-class distances“ (15.03.2021.)

[8] Mateo Kiđmet, „Primjena metode stablo odlučivanja“ [na internetu] dostupno: https://urn.nsk.hr/urn:nbn:hr:211:444358 [pristupano: 30.06.2021.]

[9] Stella Paris, „Umjetne neuronske mreže“ [na internetu] dostupno:
view (uniri.hr) [pristupano 30.06.2021.]

[10] Hrvatska enciklopedija „Standardna devijacija“ [na internetu] dostupno:

https://www.enciklopedija.hr/natuknica.aspx?ID=57758 [pristupano 30.06.2021.]

[11] Hrvatska enciklopedija „Varijanca“ [na internetu] dostupno:

https://www.enciklopedija.hr/natuknica.aspx?ID=63913 [pristupano 30.06.2021.]

[12] Jelić Nikola „Duboke Neuronske mreže“ [na internetu] dostupno:

https://repozitorij.pmf.unizg.hr/islandora/object/pmf:9024/datastream/PDF/download [pristupano 30.06.2021.]

[13] BigML „About BigML“ [na internetu] dostupno:

https://bigml.com/about/ [pristupano 01.07.2021.]

Popis slika

Popis tablica