

Analiza podataka o korisnicima telekomunikacijskih usluga

Seminarski rad u okviru kursa
Istraživanje podataka
Matematički fakultet
Univerziteta u Beogradu

Strahinja Mitrić
mi151710@alas.matf.bg.ac.rs

Predmetni profesor: Nenad Mitić Predmetni asistent: Mirjana Maljković

Sadržaj

1	Uvod	2
2	Pretprocesiranje i analiza podataka	2
2.1	Opis skupa podataka	2
2.2	Pretprocesiranje	3
3	Vizuelizacija	4
4	Klasifikacija	6
4.1	Stablo odlučivanja	7
4.2	K najbližih suseda	9
4.3	Naivni Bajesov algoritam	10
4.4	Neuronske mreže	10
5	Klasterovanje	12
5.1	Algoritam K sredina	12
5.2	Hijerarhijsko klasterovanje	14
6	Pravila pridruživanja	15

1 Uvod

Skup podataka *Telco customer churn* sadrži podatke o korisnicima telekomunikacijskih usluga, izvesne kompanije. Podaci su preuzeti sa sledeće internet adrese: <https://www.kaggle.com/blastchar/telco-customer-churn>.

Generalno, svrha ovog skupa podataka je da nam pomogne da uspemo da predvidimo koji od korisnika usluga firme ima veću verovatnoću da će raskinuti ugovor sa firmom u budućnosti, odnosno odustati od korišćenja njenih usluga. Kada bismo uspeali da to uspešno predvidimo, pokušali bismo da nuđenjem određenih beneficija tom korisniku – promenimo njegovu odluku i nastavimo saradnju sa njim. Pored toga, pokušaćemo da dodemo i do drugih saznanja koja nam mogu biti od značaja.

Za analizu korišćen je alat KNIME.

2 Pretprocesiranje i analiza podataka

2.1 Opis skupa podataka

Svaki red u tabeli sadrži podatke o jednom od korisnika. Podatke možemo svrstati u 3 grupe podataka: usluge na koje je korisnik (bio) pretplaćen, podaci o nalogu korisnika (detalji o ugovoru, dužina ugovora itd.) i demografski podaci o korisniku.

Atributi koji opisuju korisnika su sledeći:

- CustomerID - jedinstvena šifra korisnika
- gender - pol korisnika
- SeniorCitizen - da li je korisnik starija osoba
- Partner - da li ima partnera
- Dependents - da li izdržava nekoga
- tenure - koliko meseci korisnik koristi usluge kompanije
- PhoneService - da li korisnik koristi telefonske usluge kompanije
- InternetService - da li korisnik koristi internet usluge kompanije
- MultipleLines - da li korisnik koristi više telefonskih linija
- OnlineSecurity - da li korisnik koristi usluge kompanije vezane za internet sigurnost
- OnlineBackup - da li korisnik koristi Online Backup usluge kompanije
- DeviceProtection - da li korisnik koristi Device Protection usluge
- TechSupport - da li korisnik koristi usluge korisničke podrške
- StreamingTV - da li korisnik gleda TV program preko interneta ili ne
- StreamingMovies - da li korisnik gleda filmove preko interneta ili ne
- Contract - označava tip ugovora klijenta
- PaperlessBilling - da li korisnik koristi Paperless Billing usluge
- PaymentMethod - način na koji korisnik izmiruje svoje obaveze
- MonthlyCharges - iznos mesečnog računa korisnika, izraženo u USD
- TotalCharges - suma iznosa koje je korisnik do sada uplatio na račun kompanije
- Churn - označava da li je korisnik odustao od usluga kompanije ili ne


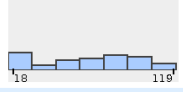
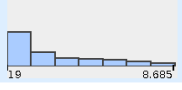
Da bismo stekli bolji utisak kako ti podaci izgledaju, na slikama 1 i 2 mogu se videti prvih 5 redova tabele. Osnovne statističke informacije o numeričkim atributima u tabeli su vidljive na slici 3.

Row ID	S custome...	S gender	I Senior...	S Partner	S Depen...	I tenure
Row0	7590-VHVEG	Female	0	Yes	No	1
Row1	5575-GNVDE	Male	0	No	No	34
Row2	3668-QPYBK	Male	0	No	No	2
Row3	7795-CFOCW	Male	0	No	No	45
Row4	9237-HQITU	Female	0	No	No	2
Row5	9305-CDSKC	Female	0	No	No	8

Slika 1: Prikaz prvih 5 redova tabele (1).

Row ID	S Phone...	S MultipleLines	S Intern...	S Online...	S Online...	S Device...	S TechS...	S Strea...	S Strea...	S Contract	S Paperl...	S PaymentMet...	D Monthl...	D TotalC...	S Cl
Row0	No	No phone service	DSL	No	Yes	No	No	No	No	Month-to-month	Yes	Electronic check	29.85	29.85	No
Row1	Yes	No	DSL	Yes	No	Yes	No	No	No	One year	No	Mailed check	56.95	1,889.5	No
Row2	Yes	No	DSL	Yes	Yes	No	No	No	No	Month-to-month	Yes	Mailed check	53.85	108.15	Yes
Row3	No	No phone service	DSL	Yes	No	Yes	Yes	No	No	One year	No	Bank transfer (...)	42.3	1,840.75	No
Row4	Yes	No	Fiber optic	No	No	No	No	No	No	Month-to-month	Yes	Electronic check	70.7	151.65	Yes
Row5	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	Month-to-month	Yes	Electronic check	99.65	820.5	Yes

Slika 2: Prikaz prvih 5 redova tabele (2).

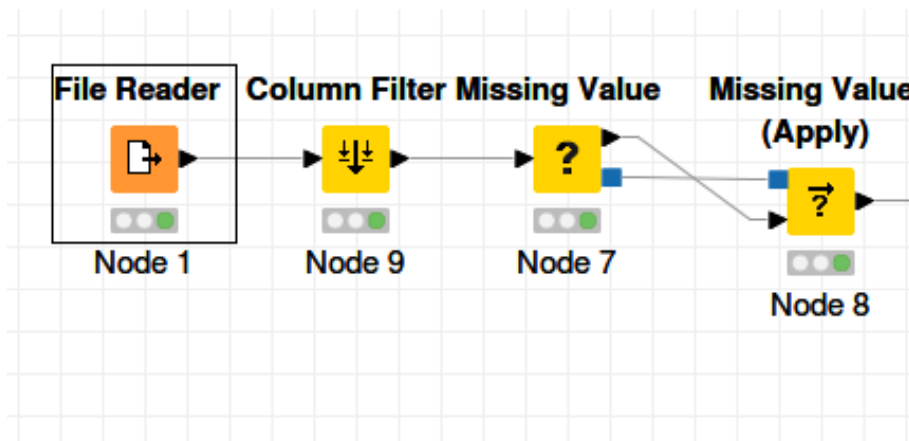
Column	Min	Mean	Median	Max	Std. Dev.	Skewness	Kurtosis	No. Missing	No. +∞	No. -∞	Histogram
SeniorCitizen	0.0	0.1624	0.0	1	0.3688	1.8311	1.3533	0	0	0	
tenure	1	32.4218	29	72	24.5459	0.2377	-1.3878	0	0	0	
MonthlyCharges	18.25	64.7982	70.35	118.75	30.086	-0.2221	-1.2562	0	0	0	
TotalCharges	18.8	2,283.3004	1,397.475	8,684.8	2,266.7714	0.9616	-0.2318	0	0	0	

Slika 3: Prikaz numeričkih veličina i njihovih statističkih parametara.

2.2 Pretprocesiranje

Jedna od osnovnih etapa kod analize podataka jeste njihovo pretprocesiranje, odnosno otklanjanje nedostataka i dovođenje podataka u adekvatan oblik za analizu.

Nakon što smo učitali podatke, primećujemo da naša tabela ima 7043 redova (nepretprocesiranih) podataka. Za analizu su nam nepotrebni customerID podaci, jer su zapisani u internom obliku. Osim toga, treba i izbaciti redove u kojima ima neostajućih vrednosti. Ovim postupkom izbacujemo 11 redova i prvu kolonu. Ovaj postupak izvodimo koristeći sledeće čvorove: File Reader, Column Filter, Missing Value i Missing Value (Apply).

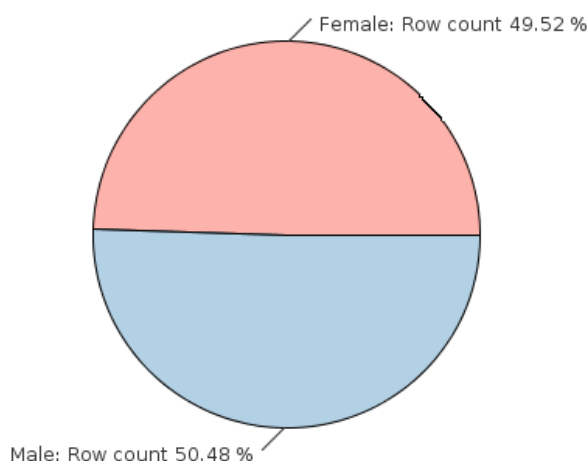


Slika 4: Obrada podataka.

U kasnijim etapama analize podataka, biće nam neophodno da podatke u tabeli prikažemo na nešto drugačiji način nego što su prikazani u originalnom obliku. U jednom slučaju, trebaće nam podaci u numeričkom obliku, a u drugom slučaju, podatke ćemo morati da prevedemo u ekvivalentan oblik originalnom, ali tako da se elementi međusobno razlikuju bez obzira kojoj koloni tabele pripadali, kako ne bi dolazilo do zabune iz koje kolone dolazi element na izlazu. Na početku poglavlja koje koriste skup podataka koji se nalazi u drugačijem obliku od početnog to će biti posebno naglašeno i biće objašnjeno na koji način su podaci bivali “prevođeni” u adekvatan oblik.

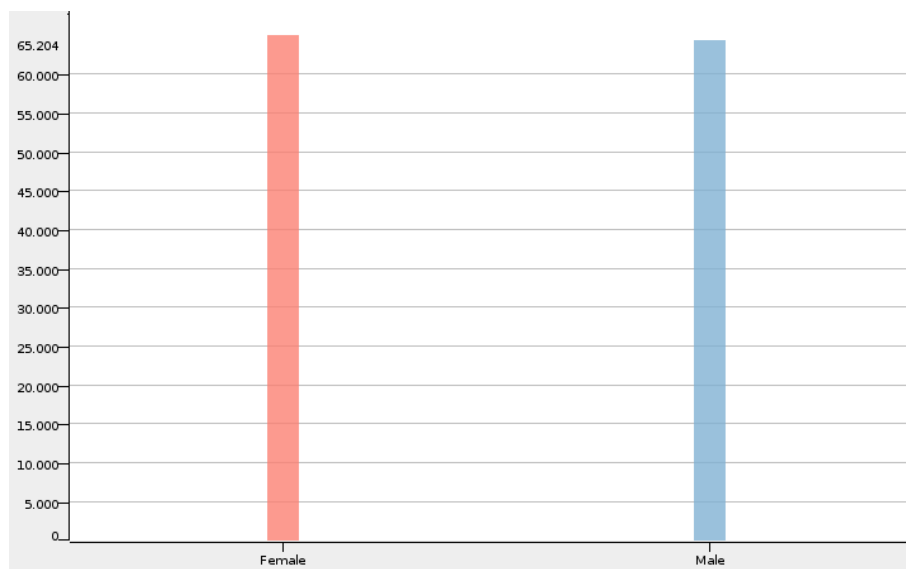
3 Vizuelizacija

Vizuelizacija predstavlja jedan od bitnih aspekata procesa istraživanja podataka. Ovaj postupak nam pomaže da bolje razumemo podatke i lakše odredimo algoritme koje ćemo primeniti na njih. U ovoj sekciji ćemo prikazati nekoliko najbitnijih karakteristika naših podataka. Za vizuelizaciju podataka koristićemo sledeće čvorove: *Scatter Plot*, *Pie chart*, *Histogram* i *Box Plot*.

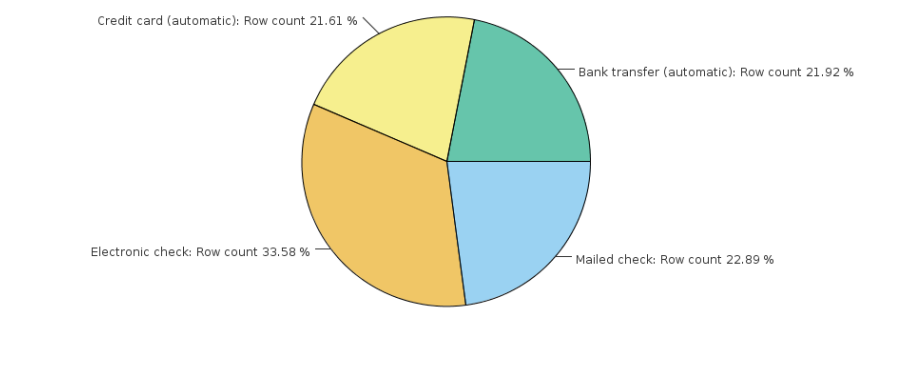


Slika 5: Odnos broja muškaraca i žena.

Kao što možemo primetiti na slici 5, odnos muškaraca i žena u podacima se razlikuje vrlo malo što nam svakako odgovara, jer će informacije koje dobijamo iz ovog skupa podataka biti merodavnije za oba pola. Na slici 6 primećujemo da su i prosečni mesečni računi muškaraca i žena prilično jednaki.



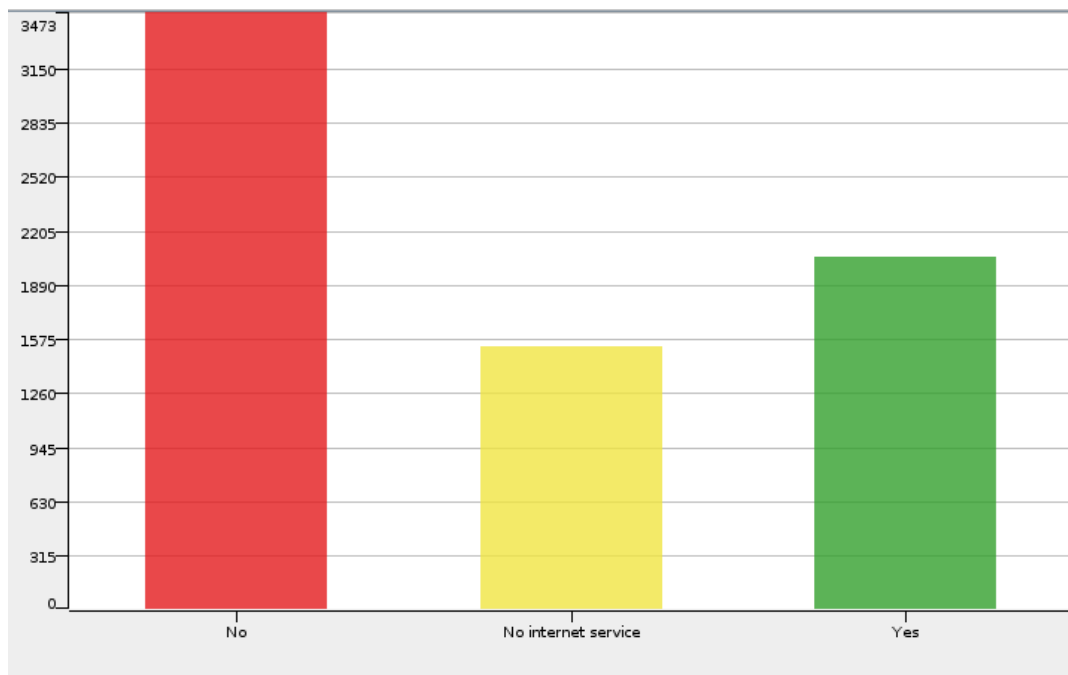
Slika 6: Prosečni iznosi zaduženja muškaraca i žena.



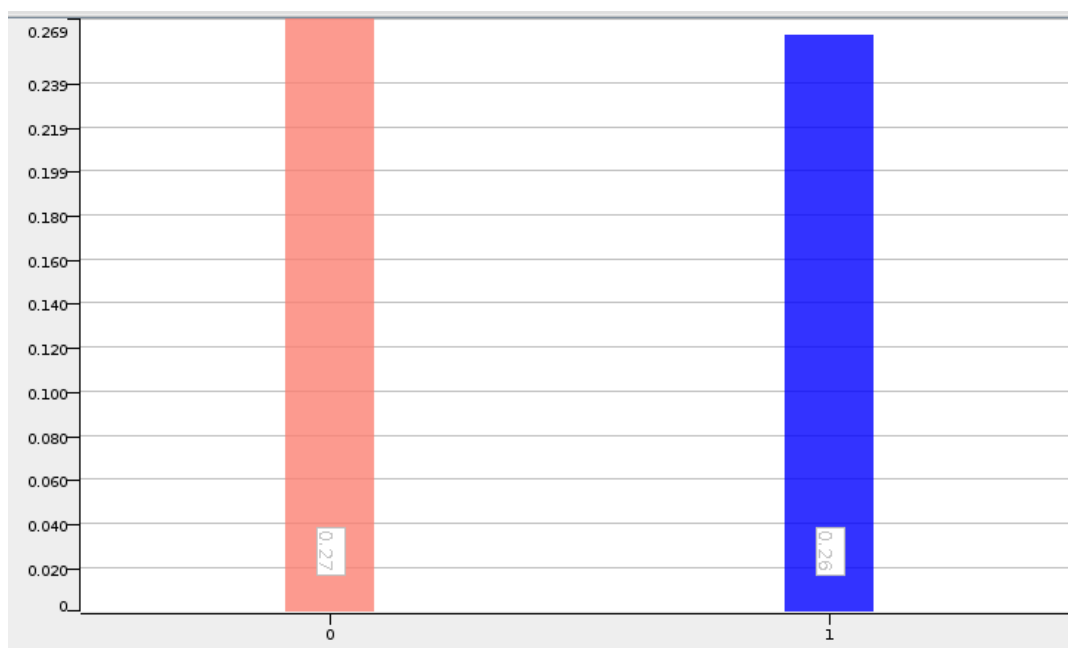
Slika 7: Udeo korisnika koji koriste svaki od načina plaćanja.

Na slici broj 7 možemo videti odnos načina plaćanja koji je zastupljen među korisnicima. Primećujemo da se korisnici u oko trećini slučajeva odlučuju za elektronskim ček, dok su preostala 3 načina plaćanja gotovo jednako zastupljena (sa blagom prednošću Mailed check-a).

Na slici 8 prikazan je broj korisnika koji su odustali od usluga kompanije, podeljenim prema tome da li su koristili usluge Korisničke Podrške za internet usluge (crveno i zeleno), ili nisu uopšte koristili internet usluge (žuto). Primećujemo da je značajno veći broj onih koji nisu zatražili pomoć Korisničke Podrške, a da su pritom odustali od usluga. Možemo pretpostaviti da postoji mogućnost da bi nuđenjem pomoći od strane Korisničke Podrške veći broj korisnika ostao sa firmom.



Slika 8: Odnos broja korisnika koji su odustali od usluga kompanije prema tome da li su koristili usluge korisničke podrške ili nisu.



Slika 9: Procenat muškaraca i žena koji odustaju od usluga kompanije.

Na slici 9 možemo videti da 27% žena i 26% muškaraca na kraju odustaje od usluga kompanije, te primećujemo da nema velike razlike u ovim podacima.

4 Klasifikacija

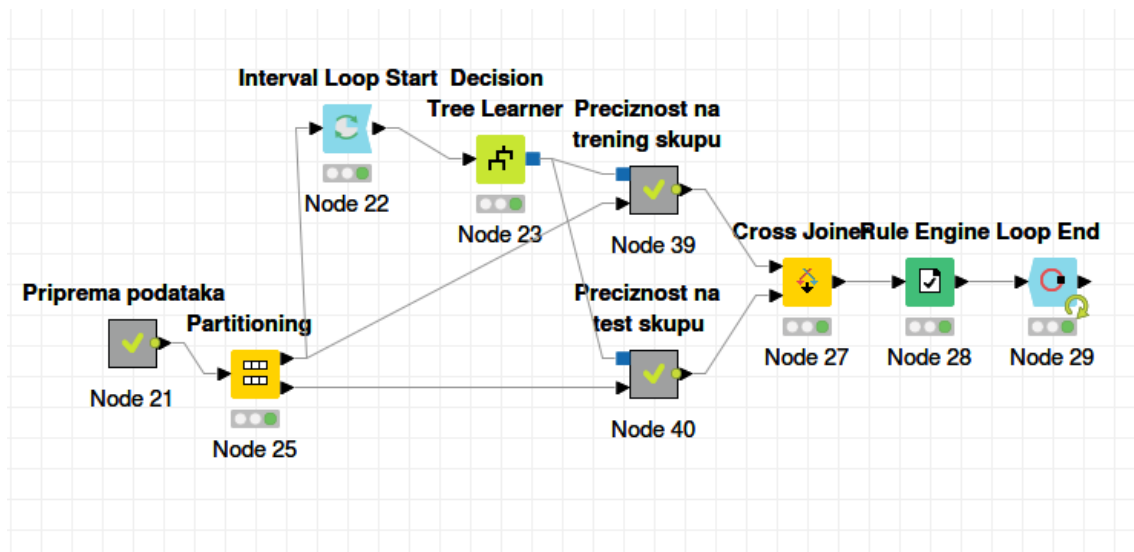
Kao što je navedeno u uvodu, u algoritmima za klasifikaciju, pogađaćemo atribut *Churn*, odnosno da li će korisnici prekinuti saradnju.

Za klasifikaciju je potrebno pripremiti podatke, glavnu obradu smo spomenuli u prethodnom odeljku i jedino što treba još uraditi jeste da se skup podeli na trening i test. Podelu na trening i

Algoritmi koje ćemo koristiti su: stablo odlučivanja, K-najbližih suseda, naivni Bajesov algoritam i neuronske mreže. Vredno napomene je i da je pokušano izvršiti klasifikaciju i SVM algoritmom, ali je zbog njegove vremenske neefikasnosti nad datim skupom podataka bilo nemoguće dobiti rezultat u razumnom vremenskom roku, te isti neće biti prikazani.

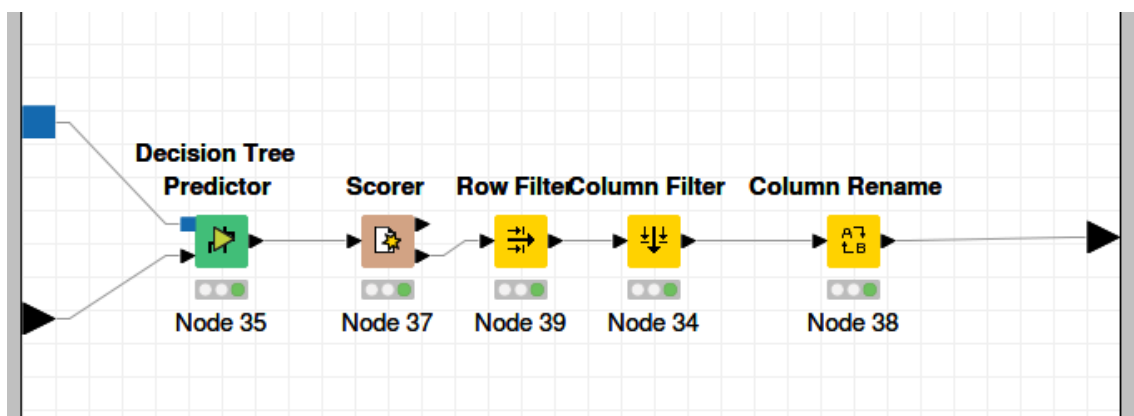
4.1 Stablo odlučivanja

Prvi algoritam za klasifikaciju je stablo odlučivanja. Kao meru kvaliteta koristimo Ginijev indeks, i podešavamo da metod odsecanja bude MDL. Za broj minimalnih elemenata u listu stabla postavljamo promenljivu kojoj menjamo vrednost u petlji i to počev od 10 do 100 sa korakom 10.



Slika 10: Algoritam stablo odlučivanja.

Na slici 11 može se videti struktura metačvorova koji računaju preciznost na trening odnosno test skupu.

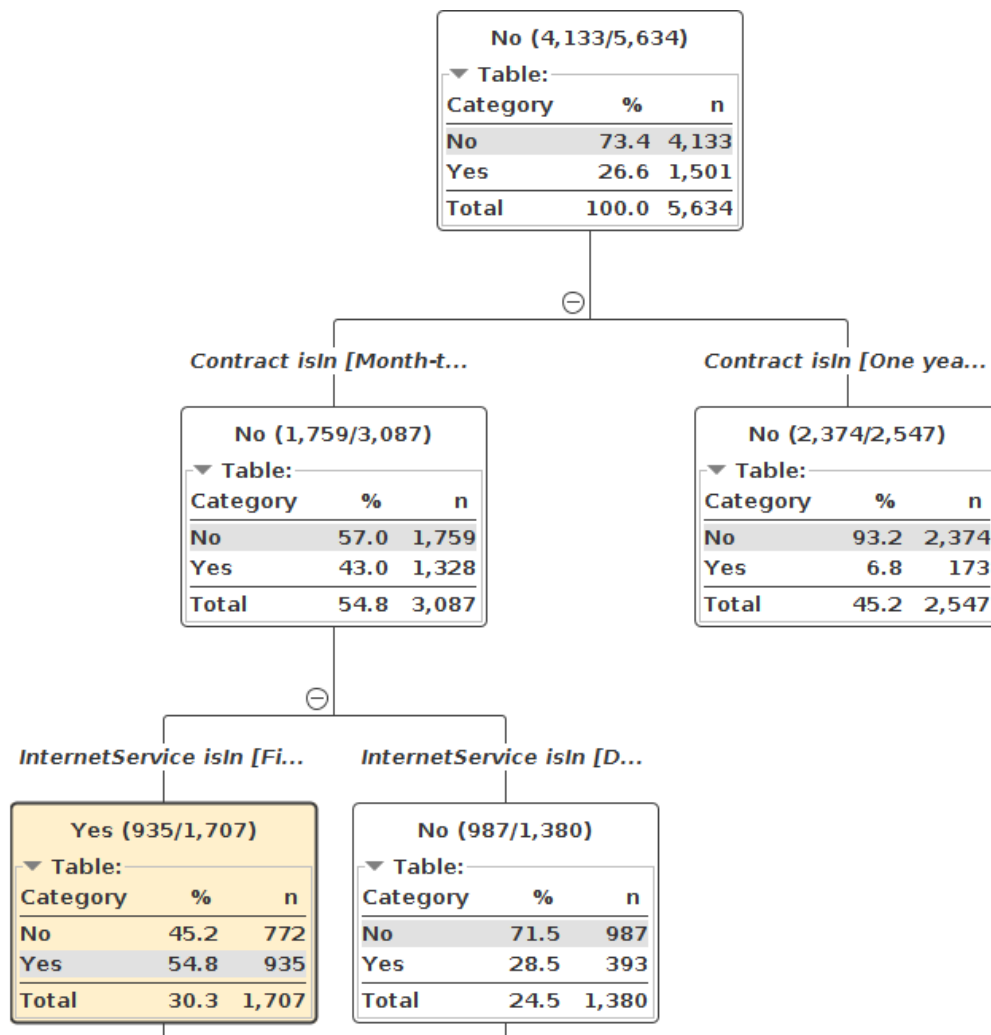


Slika 11: Metačvorovi koji računaju preciznost.

Row ID	D Training Set Accuracy	D Test Set Accuracy	I minimalanBrojCvorova	I iteration
Overall Ov...	0.828	0.798	10	0
Overall Ov...	0.819	0.803	20	1
Overall Ov...	0.813	0.803	30	2
Overall Ov...	0.81	0.806	40	3
Overall Ov...	0.807	0.803	50	4
Overall Ov...	0.805	0.806	60	5
Overall Ov...	0.804	0.8	70	6
Overall Ov...	0.803	0.803	80	7
Overall Ov...	0.802	0.8	90	8
Overall Ov...	0.802	0.794	100	9

Slika 12: Preciznosti stabla odlučivanja na trening i test skupu.

Kao što možemo videti na slici 12, preciznosti koje postizemo sa različitim minimalnim brojem instanci po cvoru su prilično slične i variraju od 79.4% do 80.6% na test skupu. Što se tiče trening skupa, očekivano je preciznost za nijansu veća, jer smo preciznost merili na istim podacima na kojim je i algoritam učio, te možemo videti preciznost od čak 82.8%, koja se dobija za 10 instanci kao minimalan broj instanci u čvoru.



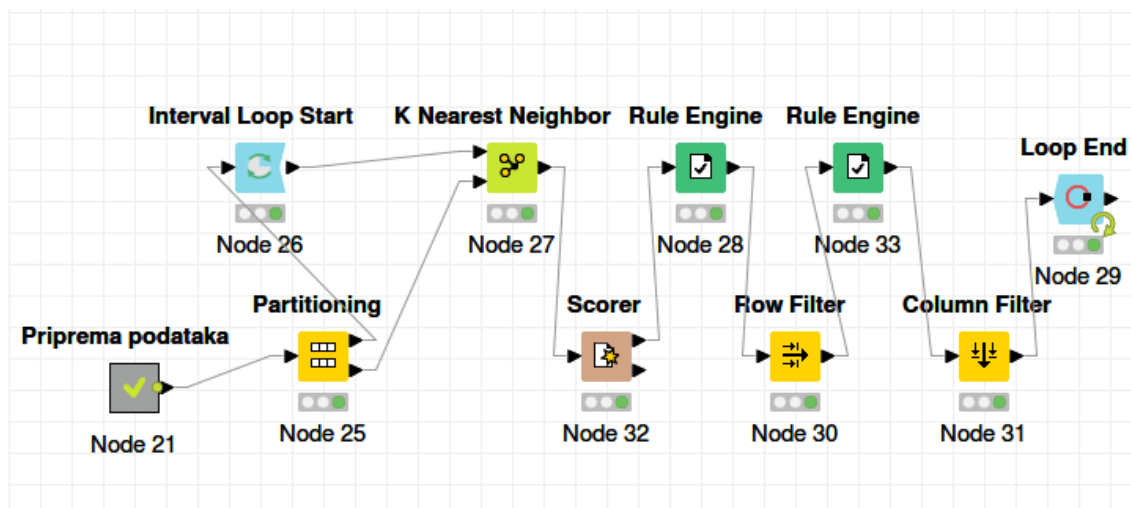
Slika 13: Prva tri nivoa stabla odlučivanja.

Na slici 13 radi preglednosti i čitljivosti, prikazana su samo prva tri nivoa stabla odlučivanja, dok je zapravo dubina stabla jednaka sedam. Primećujemo da je broj korisnika koji su odustali od

usluga firme značajno veći (7.6 puta veći) kod korisnika koji imaju tip ugovora “Month-to-month”, za razliku od korisnika koji imaju jednogodišnje ili dvogodišnje ugovore.

4.2 K najbližih suseda

Algoritam K najbližih suseda zasniva se na ideji da se isto klasifikuju vrednosti koje su bliske po nekom rastojanju. Najčešće se za to koriste Euklidsko i Menhetn rastojanje. Kako bismo iskoristili ovaj algoritam, sve atribute (za koje je to moguće) intuitivno prevodimo u numeričke, tako da nula odgovara negativnom (No), a jedinica pozitivnom (Yes) elementu. Takvim (numeričkim) atributima algoritam može da računa rastojanje. Broj komšija menjamo u petlji i to od 3 do 15.



Slika 14: Algoritam K najbližih suseda.

D	Preciznost	I	brojKomsija	I	Iteration
0.735		3		0	
0.752		4		1	
0.759		5		2	
0.769		6		3	
0.759		7		4	
0.774		8		5	
0.768		9		6	
0.779		10		7	
0.776		11		8	
0.776		12		9	
0.77		13		10	
0.771		14		11	
0.771		15		12	

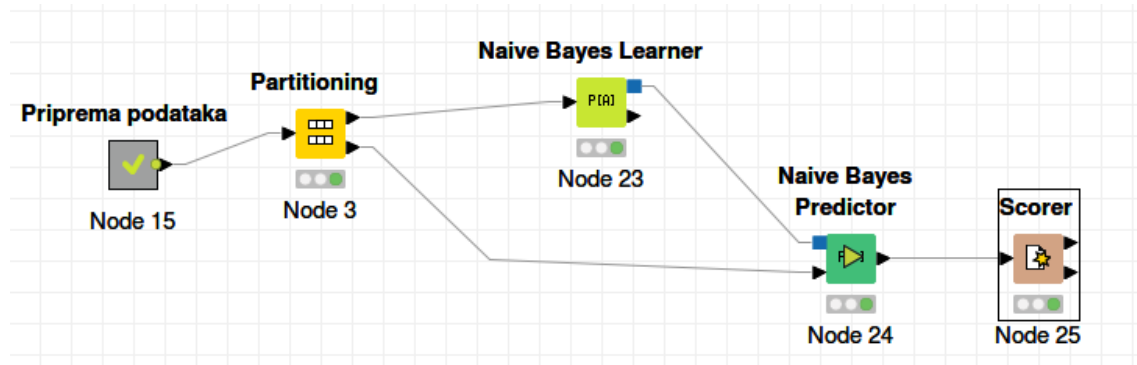
Slika 15: Preciznost algoritma K najbližih suseda.

Na slici 15 vidimo da najveću preciznost dobijamo za nešto veći broj komšija. Kada je broj komšija jednak deset, preciznost dolazi do 77.9%, a najmanja je kada je broj komšija jednak 3 – 73.5%.

4.3 Naivni Bajesov algoritam

Klasifikacija naivnim Bajesovim algoritmom pri klasifikovanju podataka služi se verovatnosnim i statističkim metodama i teoremama kao što su Bajesova i teorema o uslovnoj verovatnoći.

Što se parametara tice, u cvoru Naive Bayes Learner postavili smo kao podrazumevanu verovatnoću na 0.01.



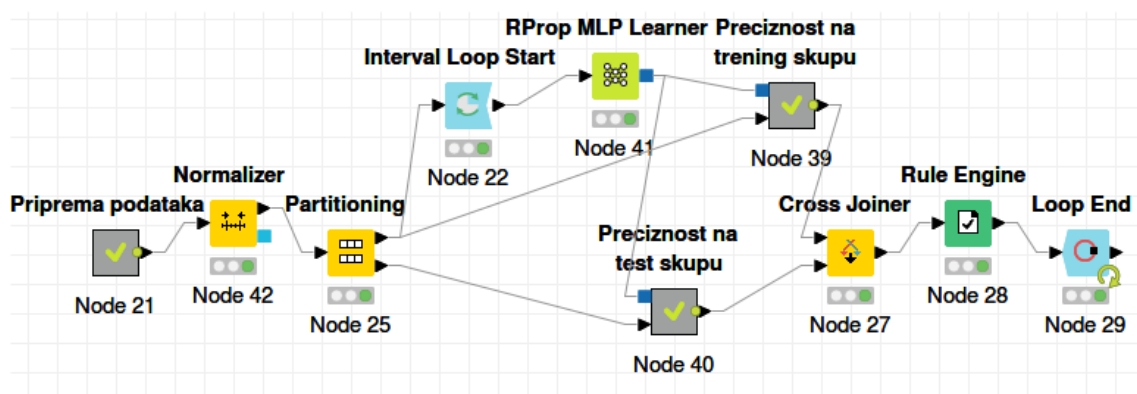
Slika 16: Naivni Bajesov algoritam.

Churn \ P...	No	Yes	
No	738	297	
Yes	80	294	
Correct classified: 1,032		Wrong classified: 377	
Accuracy: 73.243 %		Error: 26.757 %	

Slika 17: Matrica konfuzije i preciznost naivnog Bajesovog algoritma.

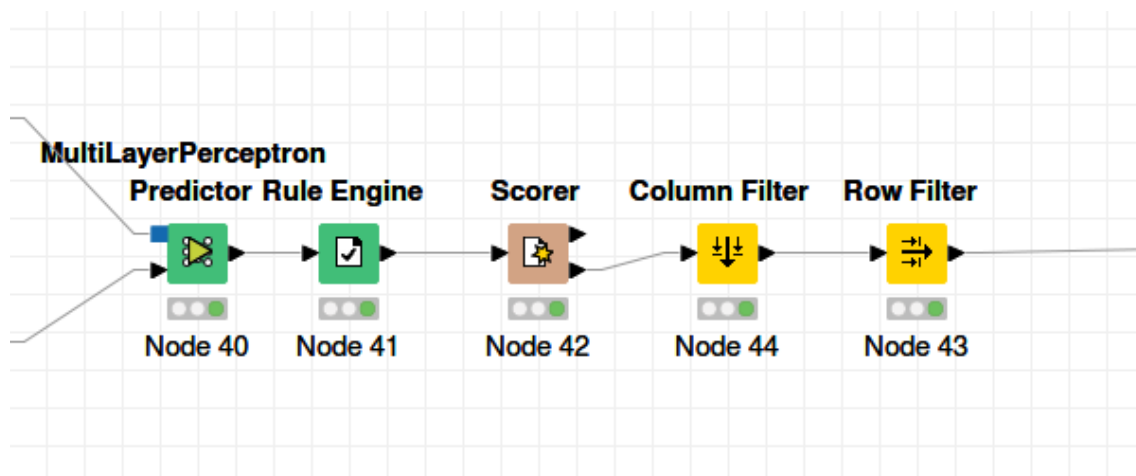
4.4 Neuronske mreže

I uz pomoć neuronskih mreža pokušavamo da klasifikujemo atribut Churn. Kao i kod algoritma “K-najbližih suseda”, sve atribute koje smo mogli da prevedemo u odgovarajuće numeričke vrednosti smo i preveli. Broj iteracija je 10.000, a broj skrivenih čvorova ide od 3 do 6 sa korakom 1. Broj čvorova u svakom unutrašnjem sloju smo postavili na 10.



Slika 18: Neuronske mreže - algoritam.

Na slici 19 može se videti struktura metačvorova korišćenih u algoritmu Neuronskih mreža.



Slika 19: Neuronske mreže - metačvorovi.

Row ID	D Accuracy	D Accuracy...	I brojUn...	I Iteration
Overall_Ov...	0.818	0.801	3	0
Overall_Ov...	0.815	0.796	4	1
Overall_Ov...	0.812	0.798	5	2
Overall_Ov...	0.817	0.803	6	3

Slika 20: Neuronske mreže - preciznost na trening i test skupovima.

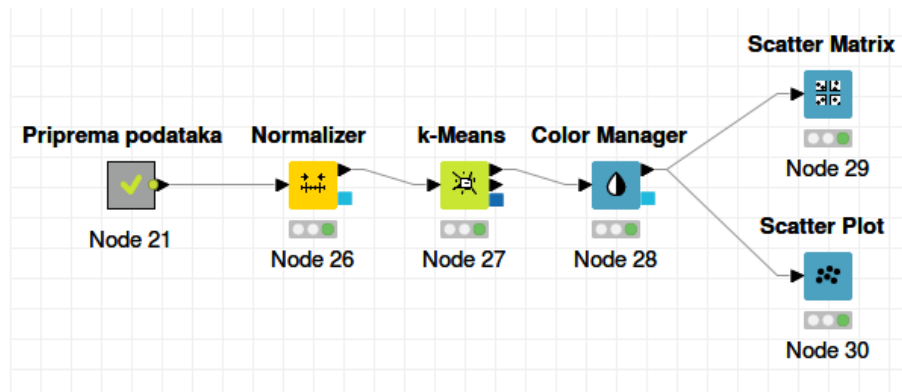
Možemo primetiti da za izabrani broj unutrašnjih čvorova, preciznost na test skupu biva skoro jednaka (između 81.2% i 81.8%). Preciznost kod test skupa takođe vrlo malo varira – od 79.6% do 80.3%.

5 Klasterovanje

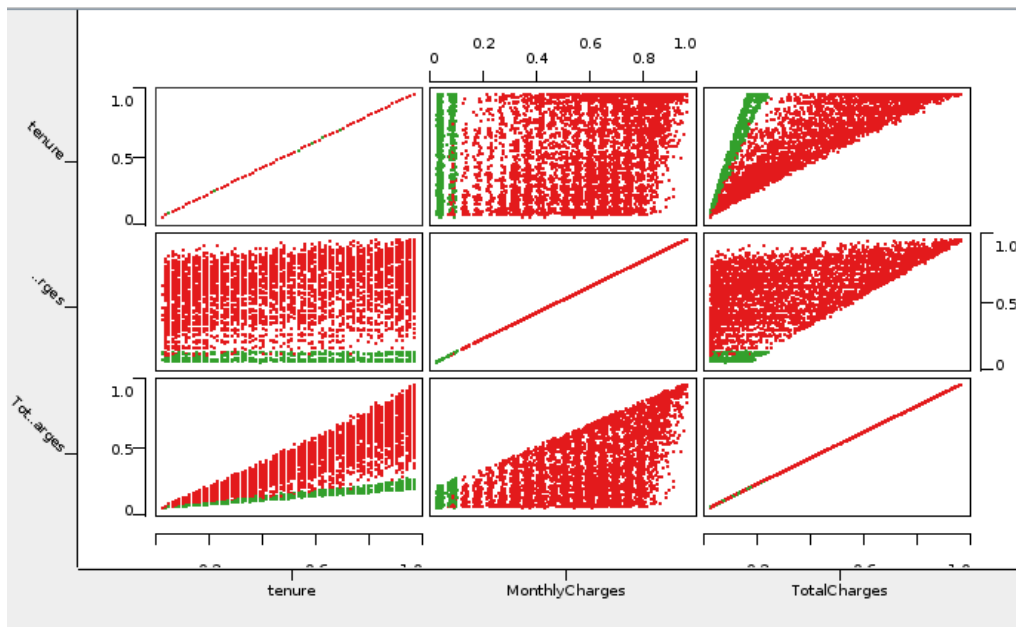
Klasterovanje predstavlja grupisanje objekata, na način da objekti iz iste grupe budu međusobno slični, odnosno da su objekti koji su u različitim grupama – međusobno različiti. Klasterovanje se vrši bez poznavanja izlaza. Ovde koristimo sledeće algoritme: algoritam K-sredina, kao i Hijerarhijsko klasterovanje. Podrazumeva se da koristimo podatke koji su prevedeni u numeričke vrednosti, a tokom postupka ćemo ih i normalizovati *min-max(0, 1)* algoritmom.

5.1 Algoritam K sredina

Kod algoritma K sredina, prvo postavljamo parametar $k=2$, kao i 99 za broj mogućih ponavljanja algoritma. Pokazaće se da jedino za $k=2$ algoritam grupiše podatke na zadovoljavajući način, tako da mi možemo primetiti neke pravilnosti. Kod svakog većeg parametra k , algoritam ne uspeva da pravilno grupiše podatke.

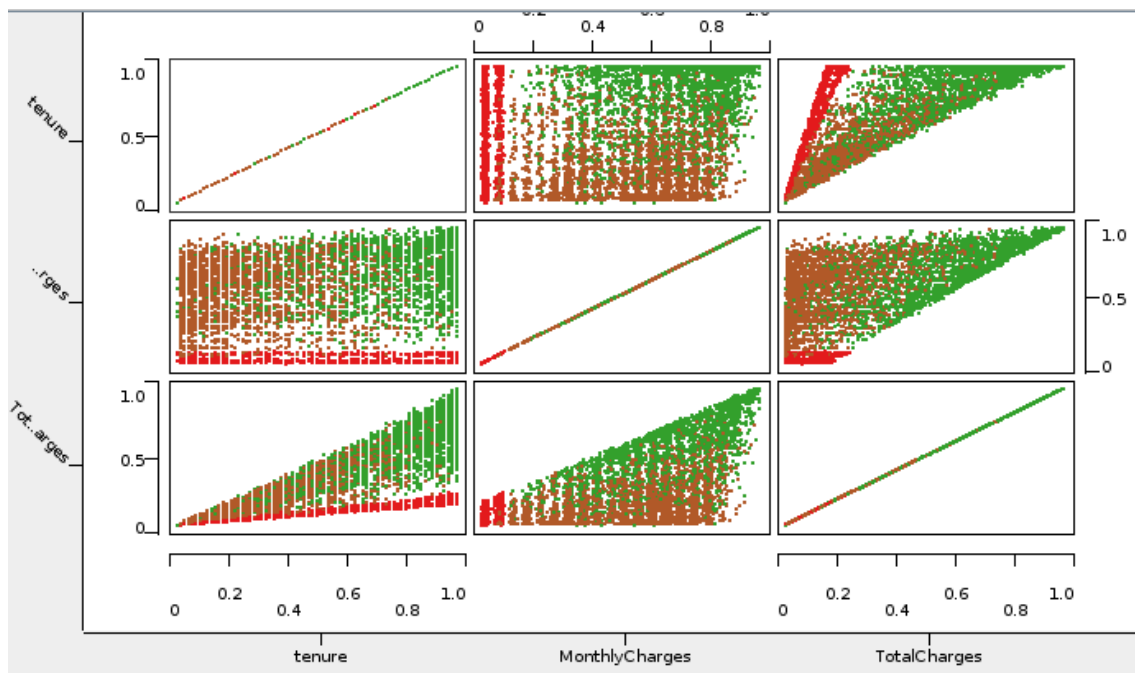


Slika 21: Klasterovanje - K sredina.



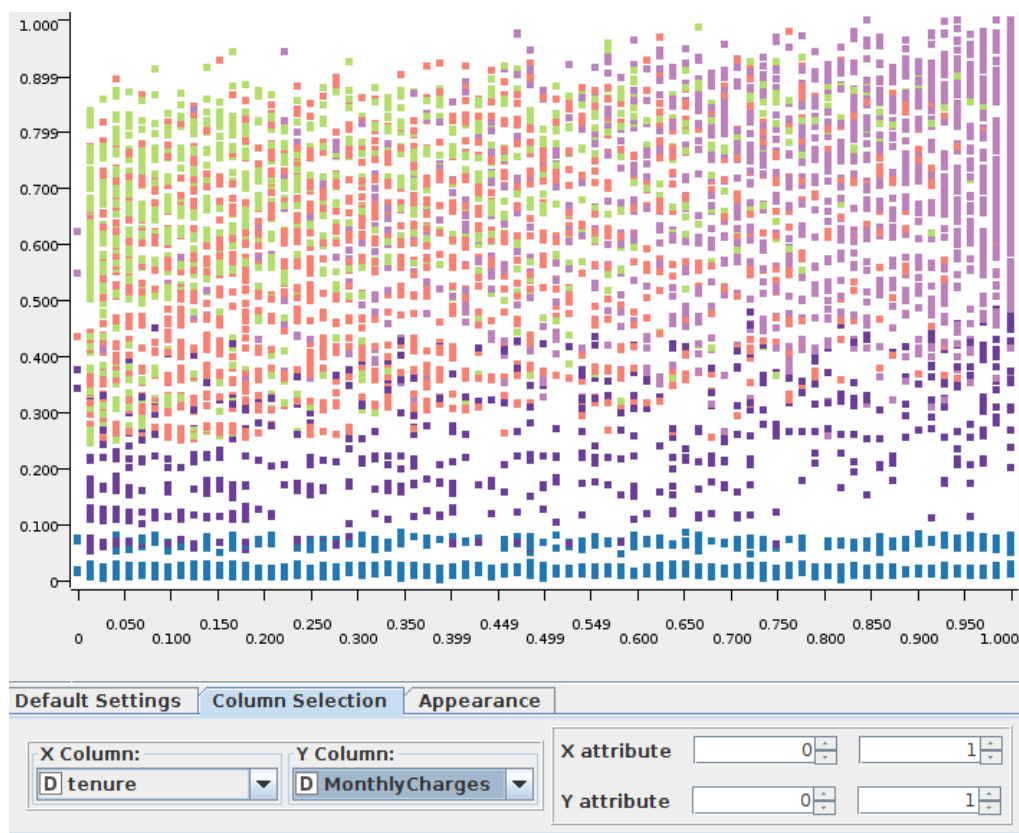
Slika 22: Prikaz klastera za $k=2$.

Na slici 22 algoritam grupiše korisnike u dve grupe, uglavnom prema tome koliki su im mesečni računi, tako da su korisnici koji plaćaju manje mesečne racune od neke granice u jednoj grupi, a drugi korisnici u drugoj grupi. Primećujemo da je algoritam uspeo da grupiše podatke na koliko-toliko različite grupe, što neće biti slučaj sa bilo kojim $k>2$.



Slika 23: Prikaz klastera za $k=3$.

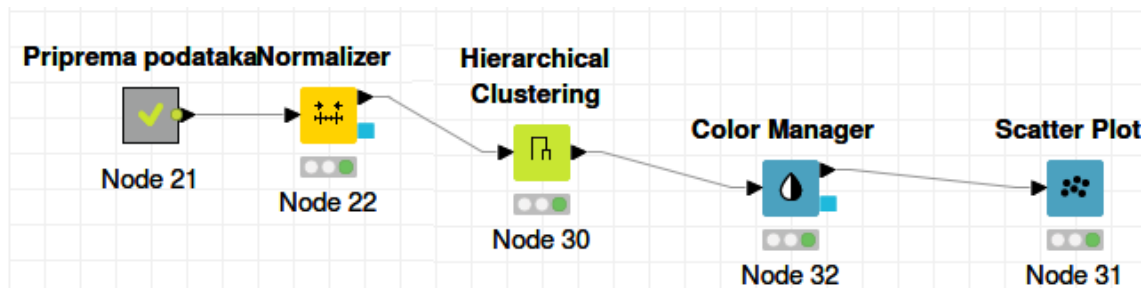
Kao što je navedeno, već kod $k=3$, na slici 23 možemo videti da dolazi do lošijeg rezultata. Za 5 klastera, očekivano, dobijamo još lošiji rezultat.



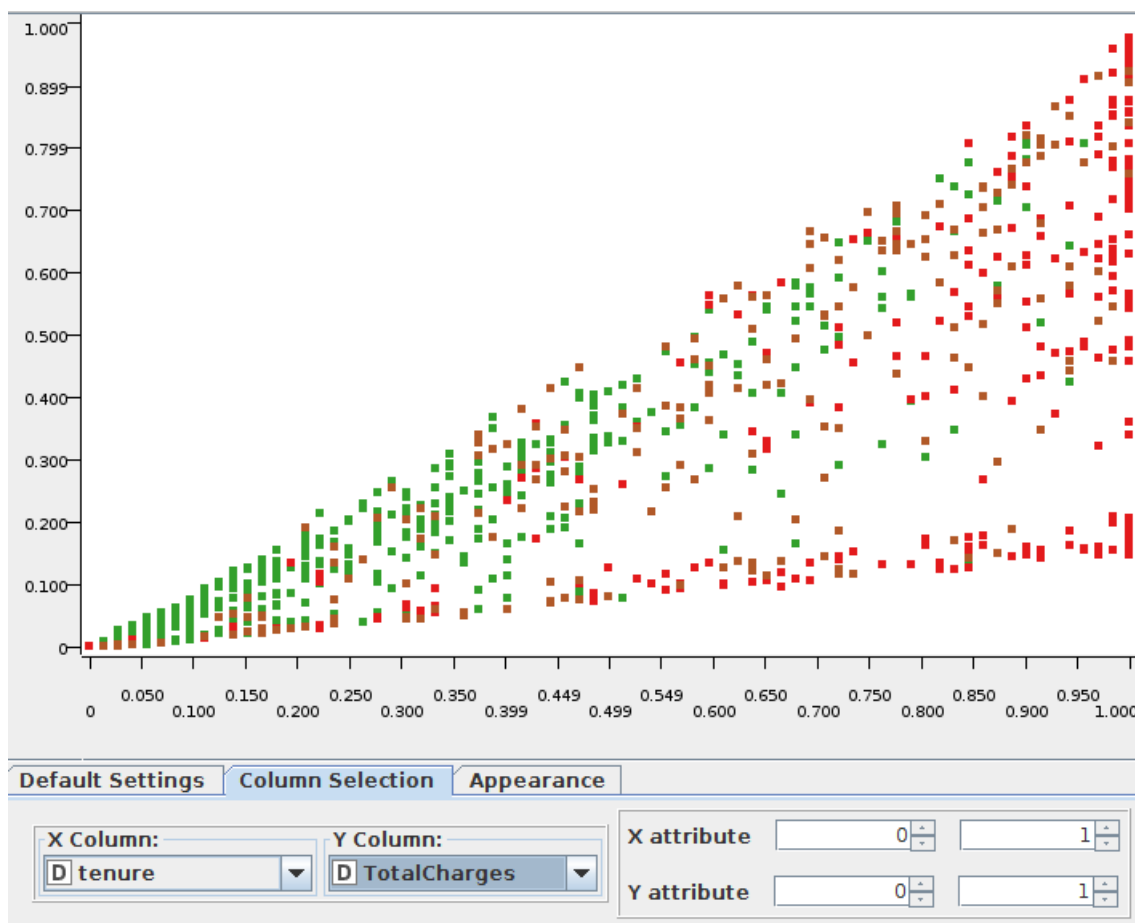
Slika 24: Prikaz klastera za $k=5$.

5.2 Hijerarhijsko klasterovanje

Za razliku od algoritma K-sredina, kod hijerarhijskog klasterovanja, ne dolazimo ni do kakvih zaključaka jer algoritam ne uspeva da klasteruje podatke. Na slici je prikazan loš pokušaj klasterovanja podataka na 3 klastera.



Slika 25: Hijerarhijsko klasterovanje.



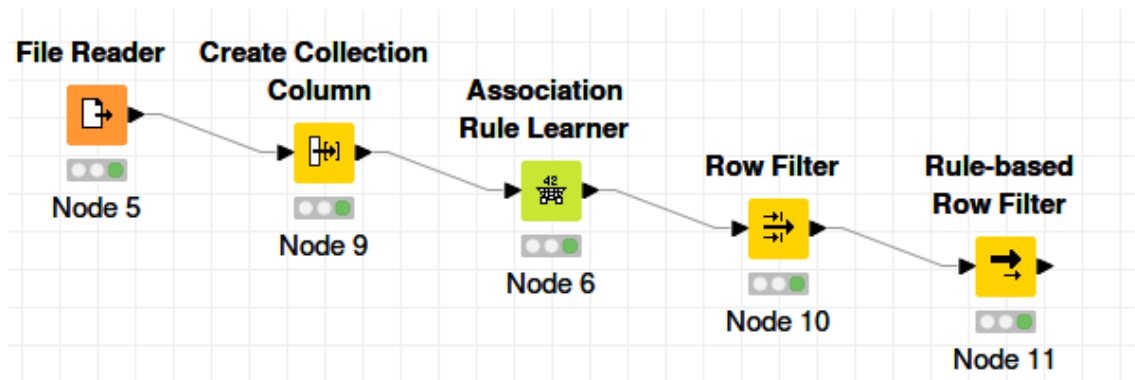
Slika 26: Rezultat hijerarhijskog klasterovanja.

Zaključujemo da su podaci u velikoj meri nepogodni za klasterovanje i da algoritmi ne mogu da pronađu zakonitosti po kojima bi ih uspešno grupisali, te su zato klasteri bili loši, a podaci loše raspoređeni.

6 Pravila pridruživanja

Da bismo koristili pravila pridruživanja i da bismo mogli da razlikujemo elemente iz rezultujuće tabele, moramo napraviti određene modifikacije početnog skupa podataka. “Yes” i “No” elemente koji se javljaju u većini kolona zamenili smo ekvivalentnim elementima, koji označavaju istu stvar, ali se i međusobno razlikuju. Na primer, u prvobitnoj tabeli, “Yes” se moglo naći i u koloni *Phone Service* i u koloni *Online security* i kao takve ih ne bismo mogli razlikovati u rezultujućoj tabeli koju dobijamo ovde. Umesto toga, “Yes” iz kolone “Phone Service” smo preimenovali u *Provided Phone Service*, a “Yes” iz *Online security* kolone u *Has Online Security*. Slične izmene smo uveli u svim kolonama u kojim je bilo neophodno to uraditi kako ne bi dolazilo do sličnih zabuna, a novi elementi koji menjaju stare elemente su intuitivno krajnje jasni i čitalac bez dileme može znati koji element je ranije stajao na tom mestu.

Što se tiče same implementacije pravila pridruživanja, nakon uspešnog učitavanja podataka, pravimo skup elemenata pomoću čvora *Create Collection Column* i to iz kolona *gender*, *PhoneService*, *InternetService*, *TechSupport*, *Contract* i *Churn*, jer su odnosi elemenata ovih kolona najzanimljiviji i najbitniji za prikaz. Prilikom konfiguracije navedenog čvora, biramo opciju da nam tip kolekcije bude skup, odnosno da se ne čuvaju duplikati elemenata.



Slika 27: Pravila pridruživanja.

Prilikom konfiguracije čvora *Association Rule Learner*, postavljamo parametre minimalne podrške na 0.2, a minimalne pouzdanosti na 0.8. Dobijamo tabelu sa slike 28.

Row ID	[D] Support	[D] Confid...	[D] ▼ Lift	[S] Consequent	[S] implies	[...] Items
rule7	0.255	0.844	1.712	No Tech Support	<---	[Month-to-month,Provided Phone Serv...
rule21	0.235	0.886	1.609	Month-to-month	<---	[Churned]
rule17	0.213	0.884	1.607	Month-to-month	<---	[Churned,Provided Phone Service]
rule8	0.255	0.805	1.464	Month-to-month	<---	[Provided Phone Service,No Tech Sup...
rule34	0.234	0.972	1.323	Not Churned	<---	[Two year]
rule30	0.211	0.971	1.321	Not Churned	<---	[Provided Phone Service,Two year]
rule26	0.201	0.926	1.26	Not Churned	<---	[No Internet Service,No Tech Support ...
rule33	0.246	0.848	1.155	Not Churned	<---	[Provided Tech Support]
rule23	0.209	0.839	1.143	Not Churned	<---	[Provided Tech Support,Provided Pho...
rule5	0.206	0.834	1.135	Not Churned	<---	[Provided Phone Service,DSL]
rule3	0.221	1	1.107	Provided Phone Service	<---	[Female,Fiber optic]
rule9	0.255	1	1.107	Provided Phone Service	<---	[Month-to-month,No Tech Support,Fib...
rule13	0.317	1	1.107	Provided Phone Service	<---	[No Tech Support,Fiber optic]
rule19	0.302	1	1.107	Provided Phone Service	<---	[Month-to-month,Fiber optic]
rule25	0.255	1	1.107	Provided Phone Service	<---	[Not Churned,Fiber optic]
rule27	0.201	1	1.107	Provided Phone Service	<---	[Not Churned,No Internet Service,No ...
rule35	0.219	1	1.107	Provided Phone Service	<---	[Male,Fiber optic]
rule39	0.44	1	1.107	Provided Phone Service	<---	[Fiber optic]
rule40	0.217	1	1.107	Provided Phone Service	<---	[No Internet Service,No Tech Support ...
rule6	0.279	0.81	1.103	Not Churned	<---	[DSL]
rule37	0.241	0.909	1.007	Provided Phone Service	<---	[Churned]
rule18	0.213	0.908	1.005	Provided Phone Service	<---	[Churned,Month-to-month]
rule4	0.448	0.905	1.002	Provided Phone Service	<---	[Female]
rule43	0.218	0.905	1.002	Provided Phone Service	<---	[Two year]
rule31	0.211	0.904	1.001	Provided Phone Service	<---	[Not Churned,Two year]
rule16	0.25	0.904	1	Provided Phone Service	<---	[Month-to-month,Male]
rule20	0.497	0.903	1	Provided Phone Service	<---	[Month-to-month]
rule1	0.247	0.902	0.999	Provided Phone Service	<---	[Month-to-month,Female]
rule2	0.326	0.902	0.998	Provided Phone Service	<---	[Not Churned,Female]
rule36	0.455	0.901	0.998	Provided Phone Service	<---	[Male]
rule32	0.662	0.901	0.998	Provided Phone Service	<---	[Not Churned]
rule22	0.336	0.901	0.997	Provided Phone Service	<---	[Not Churned,Male]
rule15	0.284	0.9	0.996	Provided Phone Service	<---	[Not Churned,Month-to-month]
rule10	0.339	0.891	0.986	Provided Phone Service	<---	[Month-to-month,No Tech Support]
rule0	0.216	0.887	0.983	Provided Phone Service	<---	[Female,No Tech Support]
rule14	0.437	0.887	0.982	Provided Phone Service	<---	[No Tech Support]
rule12	0.221	0.886	0.981	Provided Phone Service	<---	[Male,No Tech Support]

Slika 28: Pravila pridruživanja - tabela.

Primećujemo da se element *Provided Phone Service* javlja veoma često u repu pravila, što i ne treba da čudi, jer možemo pretpostaviti da je telefonska usluga osnovna usluga kompanije koja se bavi telekomunikacijama, te se daleko najviše puta javlja u skupu podataka i u kombinaciji sa većinom drugih elemenata. Ono što je nama zanimljivo je skup podataka koji dobijamo kada isključimo sve redove ove tabele u kojima se nalazi *Provided Phone Service* i među ostatkom izvučemo one koji imaju najveću podršku i pouzdanost, pomoću čvora *Rule-based row filter*, kao na slici 29.

Row ID	[D] Support	[D] Confid...	[D] ▼ Lift	[S] Consequent	[S] implies	[...] Items
rule28	0.201	1	4.615	No Internet Service	<---	[Not Churned,Provided Phone Service,No internet ...
rule29	0.201	1	4.615	No internet service	<---	[Not Churned,No Internet Service,Provided Phone ...
rule41	0.217	1	4.615	No Internet Service	<---	[Provided Phone Service,No internet service]
rule42	0.217	1	4.615	No internet service	<---	[No Internet Service,Provided Phone Service]
rule7	0.255	0.844	1.712	No Tech Support	<---	[Month-to-month,Provided Phone Service,Fiber op...
rule8	0.255	0.805	1.464	Month-to-month	<---	[Provided Phone Service,No Tech Support,Fiber o...
rule34	0.234	0.972	1.323	Not Churned	<---	[Two year]
rule30	0.211	0.971	1.321	Not Churned	<---	[Provided Phone Service,Two year]
rule26	0.201	0.926	1.26	Not Churned	<---	[No Internet Service,Provided Phone Service,No in...
rule6	0.279	0.81	1.103	Not Churned	<---	[DSL]

Slika 29: Pravila pridruživanja - tabela bez elementa *Phone Service*.

Pouzdanost merimo koliko se često se određeni element javlja u transakcijama koje sadrže drugi, a podrška meri koliko se puta ti elementi zajedno pojavljuju u odnosu na ceo skup transakcija. Jedna od stvari koja prvo zapada za oko je veoma visoka vrednost parametra *lift* u prve 4 kolone i ti podaci su svakako prvi po prioritetu za razmatranje i najzanimljiviji.