

# Analiza podataka iz skupa Pokemon

Seminarski rad u okviru kursa  
Istraživanje podataka  
Matematički fakultet

Dušan Savić, 104/2015, [mi15104@matf.bg.ac.rs](mailto:mi15104@matf.bg.ac.rs)

Avgust 2018

## Sadržaj

1 Uvod	1
2 Analiza i pretprocesiranje podataka	2
2.2 Cilj	2
2.3 Analiza podataka	2
2.3 Obrada podataka	3
2.3.1 Eliminisanje nedostajućih vrednosti	3
2.3.2 Podela podataka	3
2.4 Analiza ekstremnih vrednosti	4
3 Pravila pridruživanja	5
4 Klasterovanje	6
5 Klasifikacija	8
5.1 Prva podela po kategorijama	8
5.2 Druga podela po kategorijama	8
6 Zaključak	9

## 1 Uvod

Imaginarni svet u kome žive pokemoni nastao je 1995. godine u Japanu. Kreator ovog serijala Satoshi Tajiri predstavio je pokemone kao životinje sa super sposobnostima koje se slobodno kreću ili predstavljaju ljubimce ljudima. Pojedinaac koji poseduje pokemona smatra se pokemon trenerom. Cilj svih pokemon trenera je da u saradnji sa svojim pokemonima oforme tim od jednog čoveka i tri pokemona koji će se nadmetati u borbi sa drugim trenerima. Pokemon trener sarađuje sa pokemonima, uči o njihovom načinu borbe kao i o super sposobnostima koje taj pokemon poseduje. Nakon što se čovek i pokemon zbliže, pokemon postigne visok nivo utreniranosti ili se desi neka nepredviđena opasna situacija, pokemon može evoluirati u želji da pomogne svom vlasniku u njegovom daljem putu kad vrhu.

Podaci su preuzeti sa adrese: <https://www.kaggle.com/abcsds/pokemon>

## 2 Analiza i pretprocesiranje podataka

### 2.1 Cilj:

Želeli smo da proverimo nekoliko stvari vezanih za postojeće podatke:

- \*ukupni skor po tipu pokemona
- \*ukupni skor po generaciji kojoj pokemon pripada
- \*kako određene karakteristike pokemona utiču na njegovu jačinu
- \*karakteristike pokemona po generaciji

### 2.2 Analiza podataka

Iz skupa podataka smo izbacili pokemone koji su legendary jer se oni nikada ne javljaju kao pratioci nekog trenera pa nisu bitni za trenera koji pokušava da skupi što bolje pokemone.

Name	Ime
Type	Tip
Total	Jačina
HP	Životni poeni
Attack	Jačina fizičkog napada
Defense	Jačine odbrane od fizičkog napada
Sp.Attack	Jačina magičnog napada
Sp.Defense	Jačina odbrane od magičnog napada
Speed	Brzina
Generation	Generacija kojoj pokemon pripada

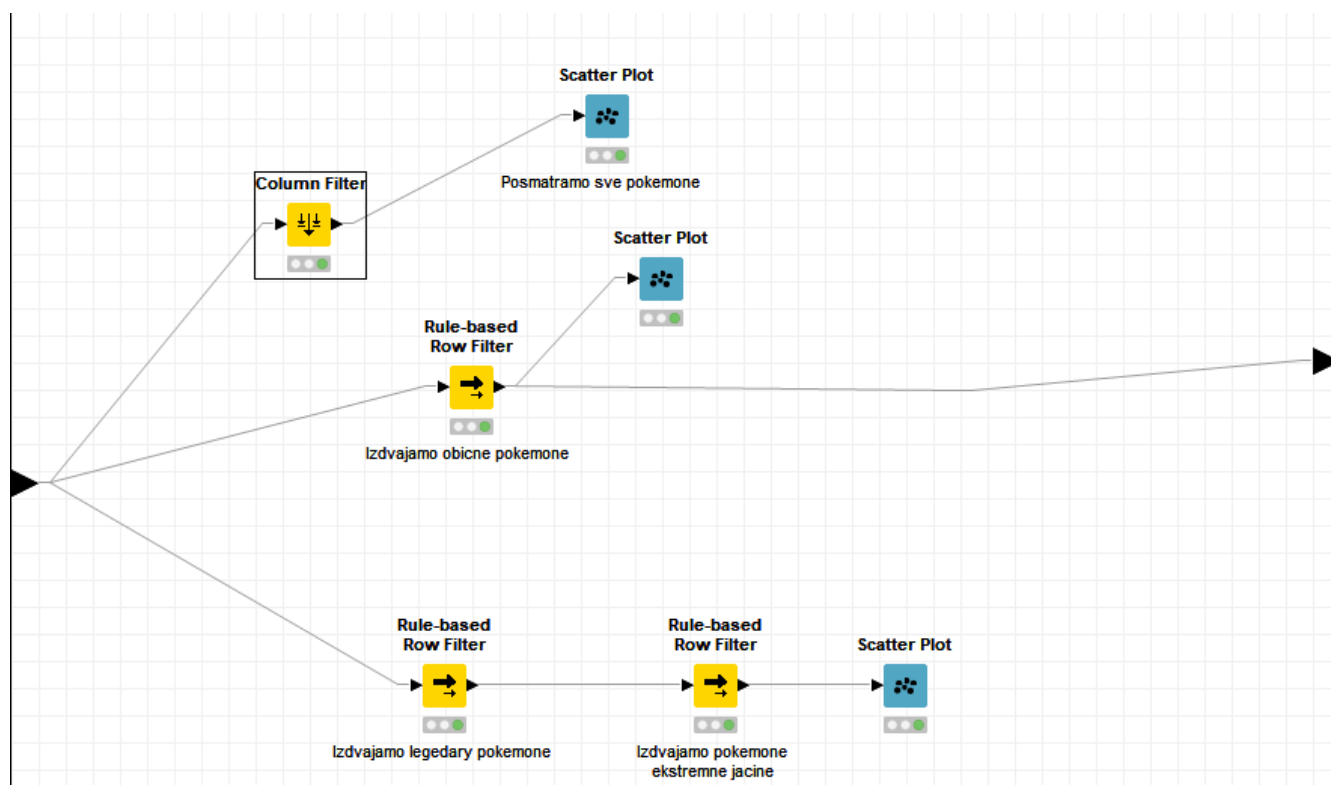
## 2.3 Obrada prodataka

### 2.3.1 Eliminisanje nedostajućih vrednosti

Kolonu type2 nismo uzimali u obzir jer više od 60% vrednosti nedostaje. Vrednosti koje nedostaju u daljem radu sa podacima smo regulisali uzoračkom sredinom.

### 2.3.2 Podela podataka

Podatke smo podelili prema tome da li je pokemon legendary ili nije, i u daljem radu koristili samo pokemone koji nisu legendary.



## 2.4 Analiza ekstremnih vrednosti

Izvukli smo skup najačih pokemona. U ovom skupu nema normalnih pokemona, svi su legendary.

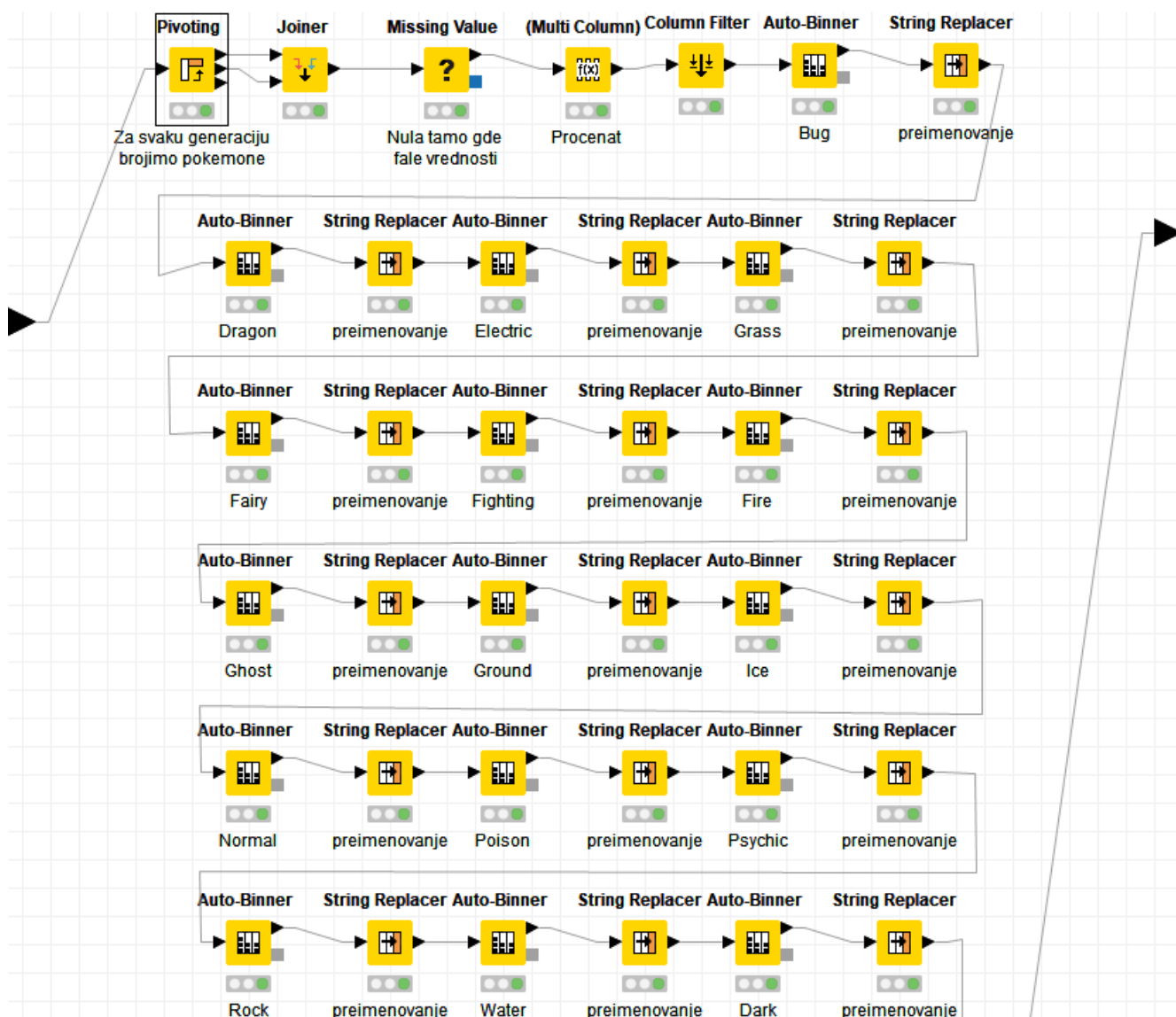
Row ID	S Name	S Type 1	I Total
493	Arceus	Normal	720
150	Mewtwo	Psychic	680
249	Lugia	Psychic	680
250	Ho-oh	Fire	680
384	Rayquaza	Dragon	680
483	Dialga	Steel	680
484	Palkia	Water	680
487	GiratinaOrig...	Ghost	680
643	Reshiram	Dragon	680
644	Zekrom	Dragon	680
716	Xerneas	Fairy	680
717	Yveltal	Dark	680
382	Kyogre	Water	670
383	Groudon	Ground	670
486	Regigigas	Normal	670
646	Kyurem	Dragon	660

*Pokemoni ekstremne jačine*

### 3 Pravila pridruživanja

Zanimalo nas je kolika je bila jačina pokemona u svakoj generaciji i kolika je bila jačina određenog tipa pokemona u svakoj generaciji.

Prvo smo izvršili grupisanje podataka po generaciji. Koristili smo pivotiranje, kako bi smo jedinstvene vrednosti kolone type pretvorili u kolone. Za svaku generaciju smo izvršili sumiranje po kolonama, a zatim i podelu sa ukupnim brojem jačine pokemona u toj generaciji. Na taj način smo dobili normalizovane podatke po kolonama i te vrednosti predstavljaju udeo jačine pokemona u jednoj generaciji u odnosu na ukupan broj. To nam je omogućilo da napravimo kategorije u odnosu na udeo jačine pokemona.



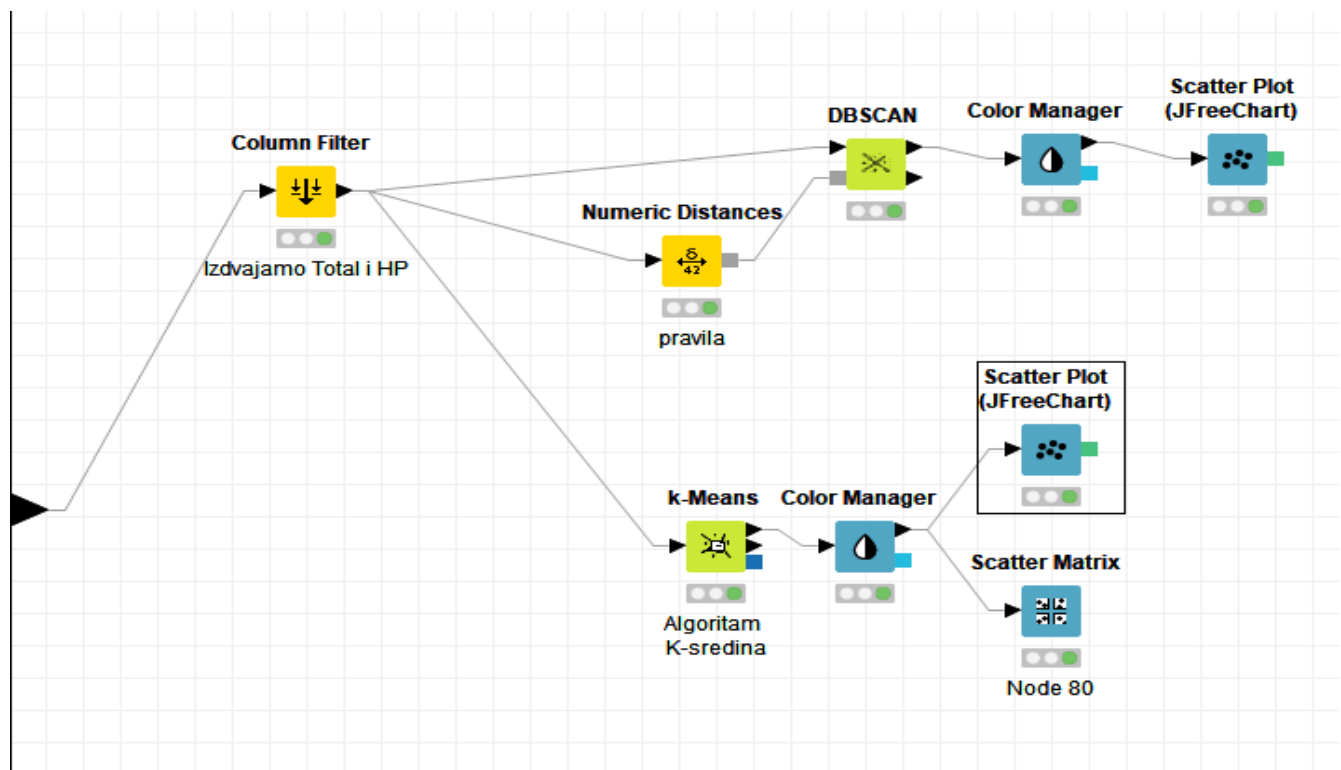
Pošto kolona generation ima 6 jedinstvenih vrednosti odlučili smo da napravimo 6 jedinstvenih kategorija za svaki od mogućih tipova.

Nakon analize pravila pridruživanja dobijamo sledeću tabelu.

Row ID	D Support	D Confide...	D ▼ Lift	S Conseq...	S implies	[...] Items
rule1	0.333	1	3	Ground 5	<---	[Rock 6,Flying 1,Bug 3]
rule10	0.333	1	3	Dragon 3	<---	[Normal 6,Flying 1]
rule11	0.333	1	3	Normal 6	<---	[Flying 1,Dragon 3]
rule17	0.333	1	3	Electric 1	<---	[Fighting 2,Rock 6,Ghos...
rule18	0.333	1	3	Fighting 2	<---	[Rock 6,Ghost 3,Electric...
rule19	0.333	1	3	Ghost 3	<---	[Fighting 2,Rock 6,Elect...
rule27	0.333	1	3	Fire 1	<---	[Fairy 1,Flying 1,Dark 2...
rule36	0.333	1	3	Dragon 6	<---	[Fairy 1,Ice 6,Normal 5,...
rule38	0.333	1	3	Ice 6	<---	[Fairy 1,Dragon 6,Norm...
rule39	0.333	1	3	Normal 5	<---	[Fairy 1,Dragon 6,Ice 6...
rule43	0.333	1	3	Grass 6	<---	[Rock 3,Fairy 1,Steel 3]
rule44	0.333	1	3	Rock 3	<---	[Fairy 1,Grass 6,Steel 3]
rule0	0.333	1	2	Bug 3	<---	[Rock 6,Flying 1,Ground...
rule2	0.333	1	2	Rock 6	<---	[Flying 1,Ground 5,Bug 3]
rule4	0.333	1	2	Bug 3	<---	[Flying 1,Poison 1,Psyc...
rule7	0.333	1	2	Psychic 3	<---	[Flying 1,Poison 1,Bug 3]
rule13	0.333	0.667	2	Ice 1	<---	[Rock 6]
rule14	0.333	1	2	Rock 6	<---	[Ice 1]

Oдавde vidimo da za generaciju u kojoj je pokemon tipa Rock jak, pokemon tipa Flying slab i pokemon tipa Bug srednje jačine velika verovatnoća da je pokemon tipa Ground dosta jak.

## 4 Klasterovanje



A scatter plot showing the relationship between HP (Horizontal Axis) and Total (Vertical Axis) for 150 Pokemon. The HP axis ranges from 0 to 250, and the Total axis ranges from 200 to 650. The data points are color-coded by Pokemon type: Normal (blue), Fire (red), Water (green), Electric (yellow), Grass (cyan), Ice (purple), Fighting (brown), Poison (pink), Ground (tan), Flying (light blue), Psychic (light green), Bug (dark green), Rock (grey), Dragon (dark blue), Steel (dark grey), and Fairy (light pink). The plot shows a general positive correlation between HP and Total, with some outliers. For example, a Normal-type Pokemon (blue) has a high HP of approximately 150 and a Total of approximately 600. A Fire-type Pokemon (red) has a low HP of approximately 25 and a Total of approximately 200. A Water-type Pokemon (green) has a high HP of approximately 150 and a Total of approximately 500. A Grass-type Pokemon (cyan) has a low HP of approximately 25 and a Total of approximately 250. An Ice-type Pokemon (purple) has a low HP of approximately 25 and a Total of approximately 300. A Fighting-type Pokemon (brown) has a low HP of approximately 25 and a Total of approximately 200. A Poison-type Pokemon (pink) has a low HP of approximately 25 and a Total of approximately 200. A Ground-type Pokemon (tan) has a low HP of approximately 25 and a Total of approximately 200. A Flying-type Pokemon (light blue) has a low HP of approximately 25 and a Total of approximately 200. A Psychic-type Pokemon (light green) has a low HP of approximately 25 and a Total of approximately 200. A Bug-type Pokemon (dark green) has a low HP of approximately 25 and a Total of approximately 200. A Rock-type Pokemon (grey) has a low HP of approximately 25 and a Total of approximately 200. A Dragon-type Pokemon (dark blue) has a low HP of approximately 25 and a Total of approximately 200. A Steel-type Pokemon (dark grey) has a low HP of approximately 25 and a Total of approximately 200. A Fairy-type Pokemon (light pink) has a low HP of approximately 25 and a Total of approximately 200.

Iako nema jedinstvenog slučaja klastera kod koga nema mešanja, sa slike možemo uočiti da količina životnih poena kod pokemona dosta utiče na njegovu ukupnu jačinu.

## 5 Klasifikacija

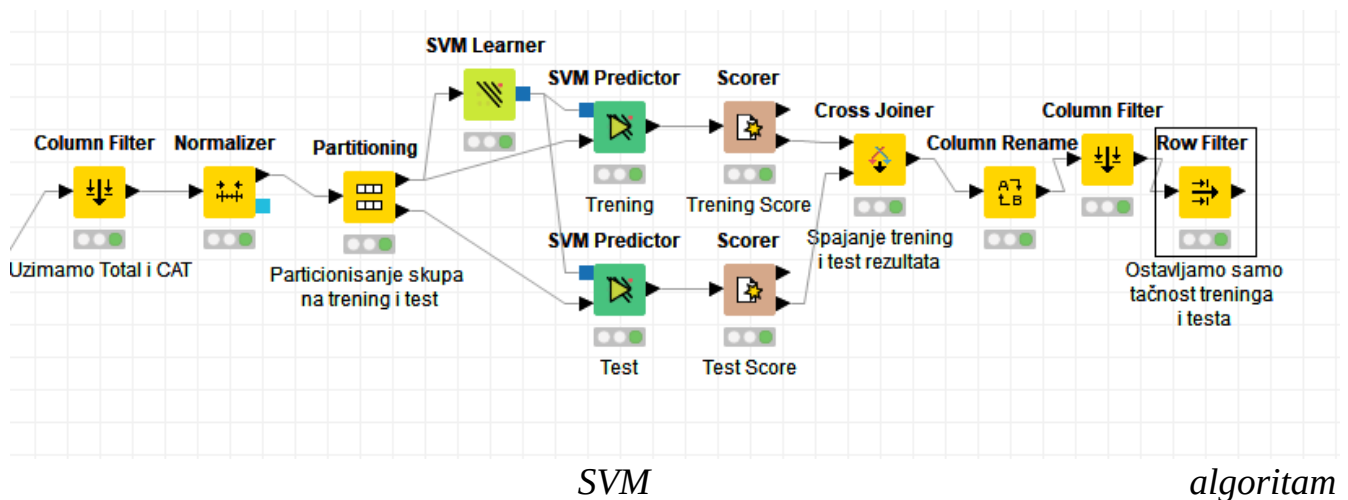
Klasifikaciju smo uradili na 2 različita pristupa. Prvo smo želeli da na osnovu ukupne jačine odredimo da li pokemon ima iznad ili ispod prosečnog broja životnih poena, a druga da na osnovu ukupne jačine pokemona odredimo koliko ja njegova jačina fizičkog napada i fizičke odbrane. U oba pokušaja smo najpre skup podataka podelili na dva dela, trening (70% podataka) i test skup (30% podataka). Ovo smo uradili koristeći čvor Partitioning.

### 5.1. Prva podela podataka po kategorijama

Prva podela podataka po kategorijama se odnosi na životne poene pokemona. Pokemoni sa brojem životnih poena iznad proseka su svrstavani u prvu a pokemoni ispod proseka u drugu grupu.

Nad ovim skupom je većina klasifikacija pokazala preciznost veću od 80%.

Klasifikaciju smo izvršili algoritmom SVM, naivnim Bajersovim algoritmom, algoritmom K najbližih suseda kao i pomoću stabla.



Row ID	D	Accura...	D	Accura...
Overall_Overall		0.9		0.83

Klasifikacija algoritmom SVM

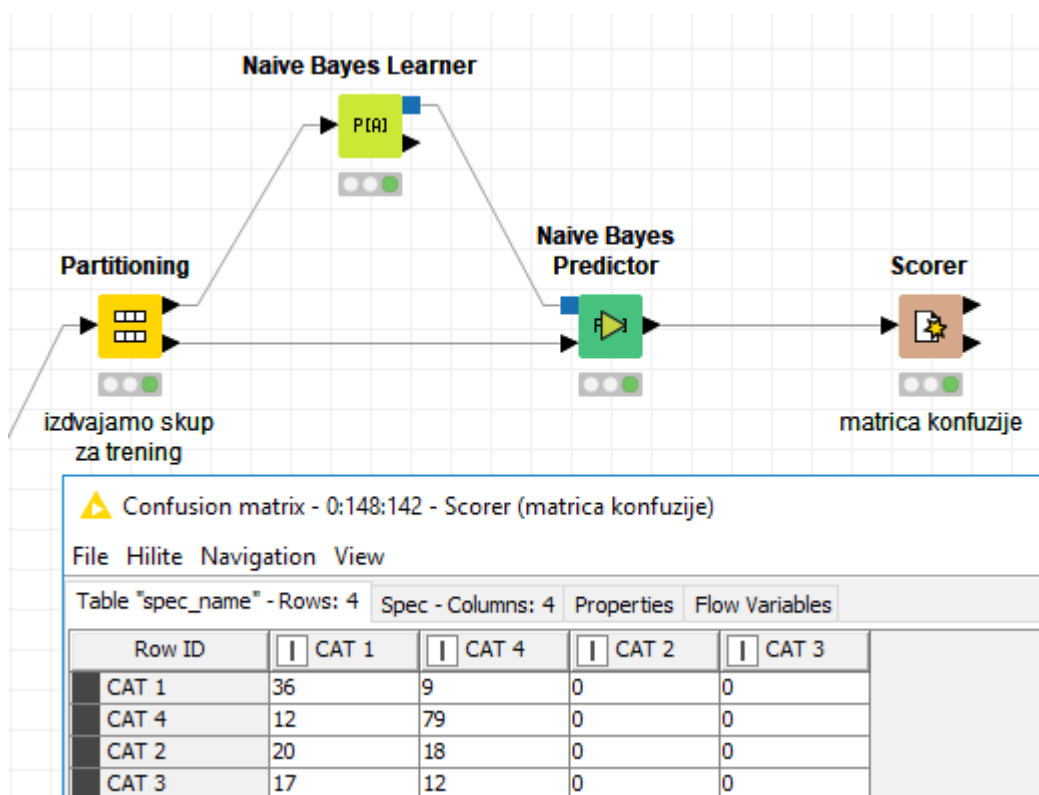
### 5.2 Druga podela po kategorijama

Prvo smo podelili pokemone u dve kategorije: iznad prosečne jačine fizičkog napada i ispod prosečne jačine fizičkog napada, a zatim te dve kategorije podelili na jos dve: iznad prosečne jačine fizičke odbrane i ispod prosečne jačine fizičke odbrane.

Klasifikaciju smo izvršili korišćenjem Stabla i naivnog Bajersovog algoritma.



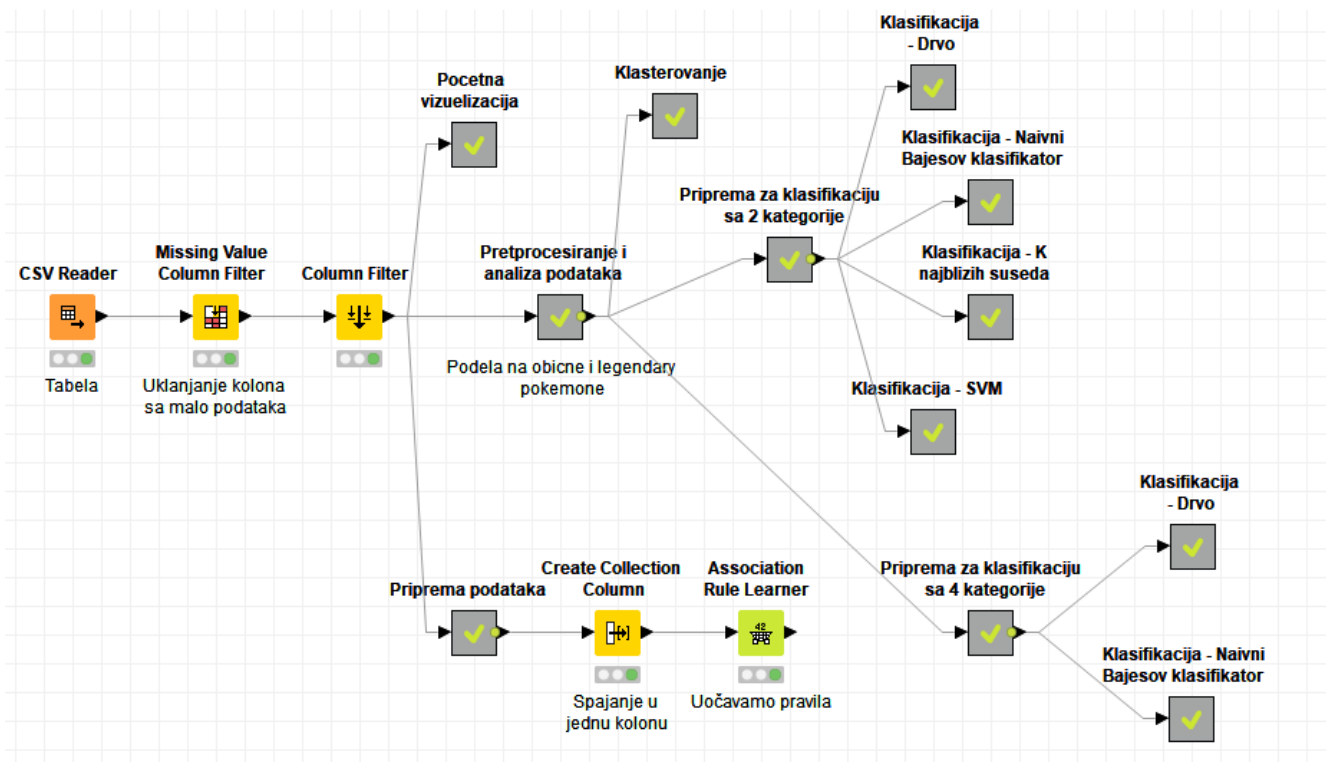
Rezultati klasifikacije sa ovakvim kategorijama su jako neprecizni.



Loša klasifikacija je nastala zbog toga što pokemon koji ima slab fizički napada i slabu fizičku odbranu i obrnuto, pa su kategorije 2 i 3 ostale prazne. Bilo kakva klasifikacija koja ne uključuje životne poene pokemona bi dala loše rezultate jer pokemon koji ima visoke fizičke atribute ima slabe magične attribute i obrnuto.

## 6 Zaključak

U ovom projektu koristili smo mnogobrojne metode za klasifikovanje i dve za klasterovanje, kao i različite načine pristupe podacima. Sa sigurnošću možemo da utvrdimo da su pokemoni koji su legendary ekstremno jaki i da na osnovu ukupne jačine nekog pokemona možmo sa tačnošću od najmanje 80% da utvrdimo da li je broj njegovih životnih poena iznad ili ispod proseka.



*Slika celog projekta*