Pravila pridruživanja o FIFA 18 igračima

Teodora Heldrih, 99/2015

**Uvod**

Skup o FIFA 18 igračima, preuzet sa adrese <https://www.kaggle.com/thec03u5/fifa-18-demo-player-dataset>, sastoji se od 3 tabele. Tabela Player Personal Data na prvi pogled izgleda ovako:



Slika 1: Tabela Player Personal Data

Field1 i Unnamed: 0 imaju identične vrednosti i označavaju redni broj igrača, ali pošto imamo ID igrača u tabeli, njega ćemo koristiti za identifikaciju igrača, a ova dva atributa ćemo ukloniti. Uklonićemo i atribute Photo, Flag i Club Logo, jer su to adrese slika igrača i neće nam biti korisni u daljoj analizi, kao i atribut Special, koji nema nikakvu posebnu namenu. To ćemo uraditi pomoću Filter taba iz Var čvora.

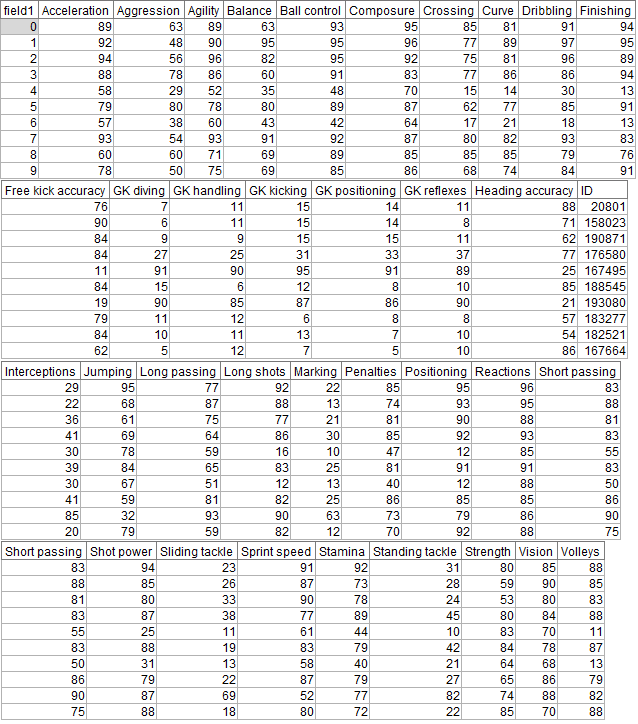
Potential je atribut koji takođe nećemo koristiti u skupu, ali on označava ocenu koju igrač može dostići ako se odgovarajuće trenira u Carrer mode.

Kako bi nam atributi Value i Wage bili korisniji u analizi, uklonićemo oznake za evro, hiljadu i milion i pretvoriti ih u neprekidne atribute i nazvaćemo ih Value in Euro i Wage in Euro. Atribut Value označava vrednost igrača, a Wage njegovu platu.

U tabeli se nalaze i atributi:

* Name, koji označava ime igrača
* Age, godine igrača
* Nationality, nacionalnost igrača
* Overall, njegovu ukupnu ocenu u FIFI
* Club, klub za koji igrač nastupa

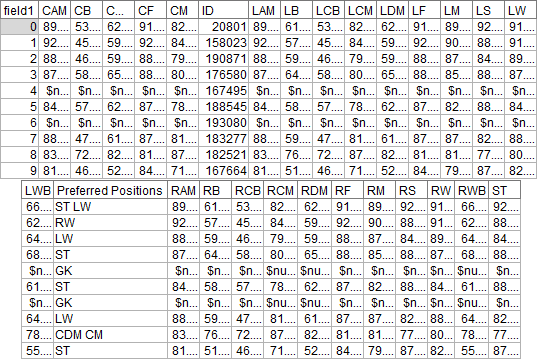
Tabela Player Attribute Data izgleda ovako:



Slika 2: Tabela Player Attribute Data

Ova tabela sadrži fizičke, tehničke, mentalne i golmanske sposobnosti igrača.

Tabela Player Playing Position Data izgleda ovako:



Ilustracija 3: Player Playing Position Data

Ova tabela sadrži ocene igrača na svim pozicijama, kao i njegove omiljene pozicije. Iz Preferred Positions ćemo izvući samo prvu i preimenovaćemo atribut u Position.

Sada ćemo skup spojiti u jednu tabelu i izvršićemo preprocesiranje podataka.

**Preprocesiranje podataka**

**Čišćenje duplikata**

Faza čišćenja je bila zamišljena da se uradi malo kasnije u radu, ali posmatranjem skupa, primećeno je da postoje duplikati u skupu, pa je prvo urađeno čišćenje duplikata. Njih bismo svakako otkrili kasnije i da nismo slučajno primetili duplikate. To je urađeno pomoću čvora Distinct. Duplikati su eliminisani kako nam ne bi uticali na statistike, pogotovo kasnije kad radimo pravila pridruživanja.

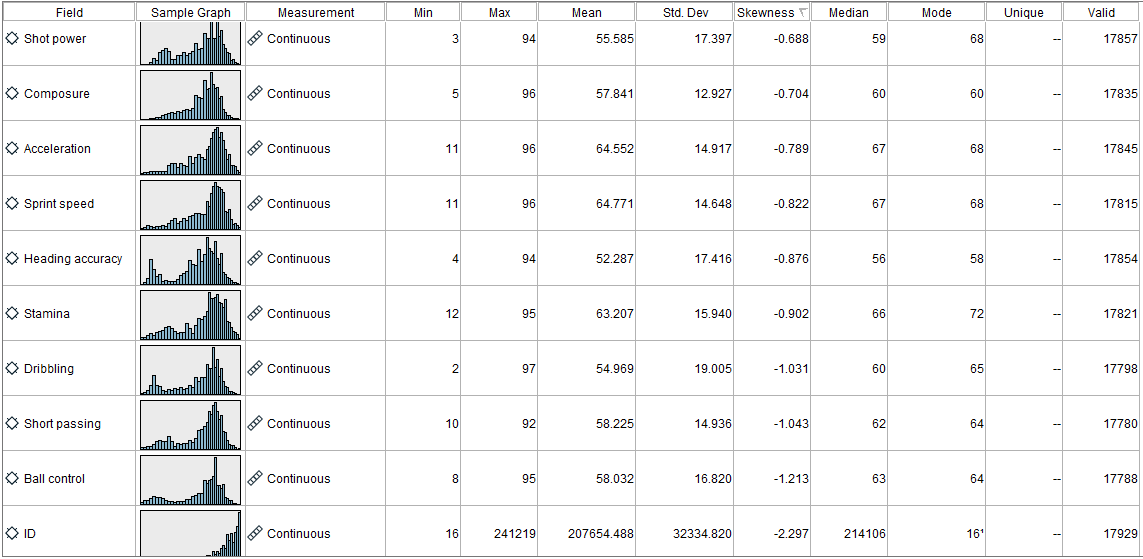


Slika 4: Duplikati

**Izdvajanje karatkeristika**

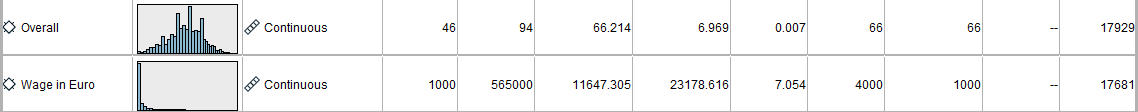
Pomoću Data Audit čvora možemo videti neke osnovne statistike.

Za neprekidne atribute imamo histograme, minimalnu, maksimalnu, srednju vrednost, standardnu devijaciju, pomerenost raspodele, medijanu i modu.



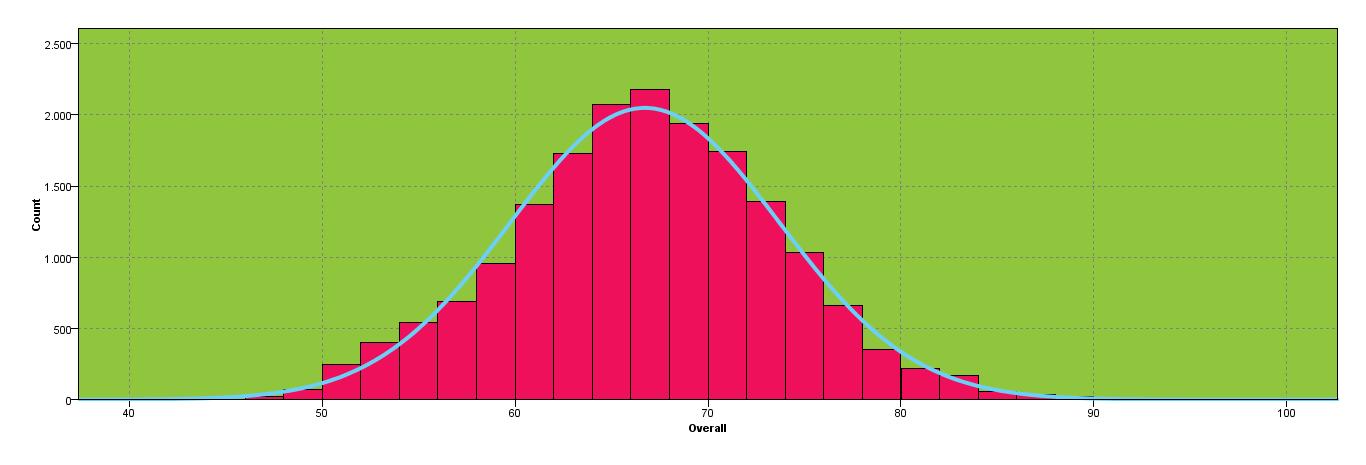
Slika 5: Data Audit Čvor

Vidimo da je najveći broj atributa pomeren ulevo, tj. levi rep raspodele je duži od desnog u odnosu na srednju vrednost.



Slika 6: najveća i najmanja pomerenost raspodele

Najmanje je pomeren rep raspodele za atribut Overall, a najviše Wage in Euro i oba su pomerena udesno.



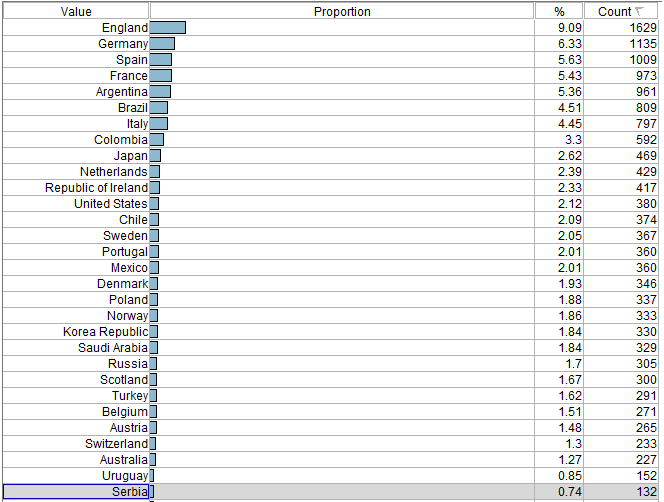
Slika 7: Histogram Overall Sa Krivom Normalne Raspodele

Ukupna ocena igrača ima neku približno normalnu raspodelu. Slično izgledaju i raspodele većine pozicija.



Slika 8: Nominalni Atribut

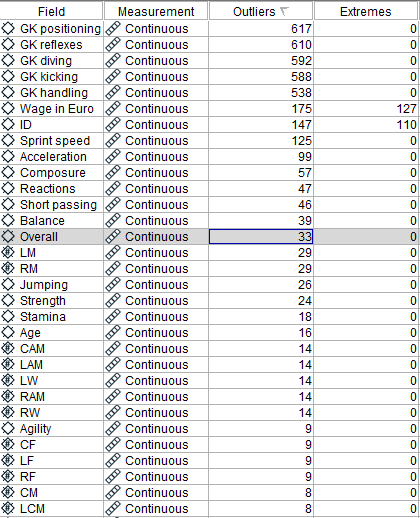
Za nominalne atribute nemamo nešto puno informacija, ali vidimo da imamo igrače iz 165 država.



Slika 9: Raspoređenost Igrača Po Nacionalnostima

Najviše igrača ima iz Engleske, Nemačke i Španije sa od po 1000 igrača, a iz Srbije ima 132 igrača u skupu.

Pomoću Quality taba možemo videti i ekstreme i autlajere.



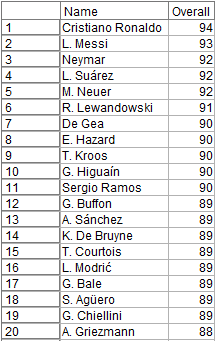
Slika 10: Autlajeri i Ekstremne Vrednosti

Ekstremne vrednosti ima samo polje plata, što je i očekivano, s obzirom na oblik raspodele. Izdvojićemo prvih nekoliko ekstrema.



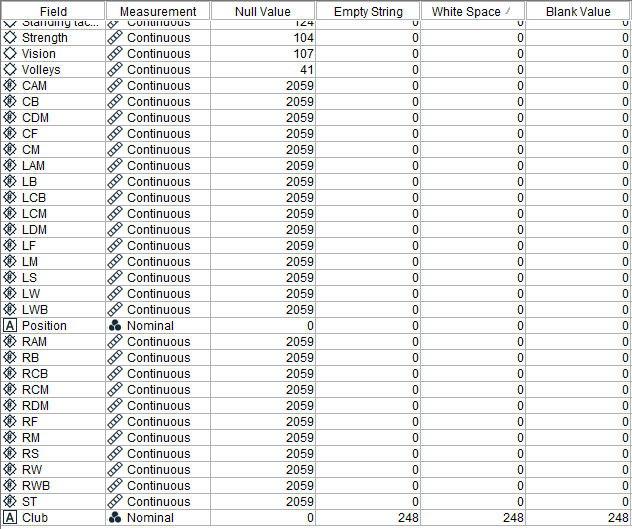
Slika 11: Ekstremi Plate

Izdvojićemo još i autlejere za Overall:



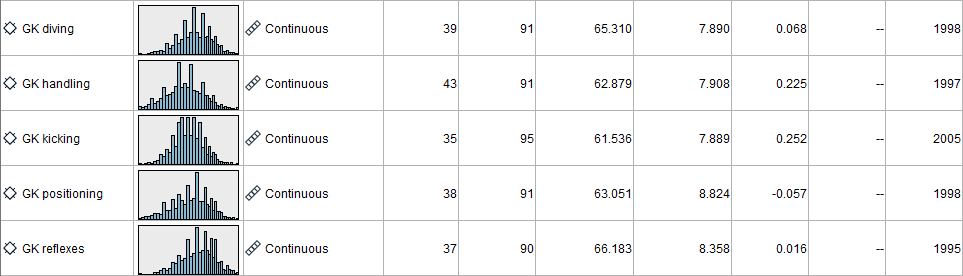
Slika 12: Autlajeri Za Ukupnu Ocenu Igrača

**Rad sa nedostajućim vrednostima**



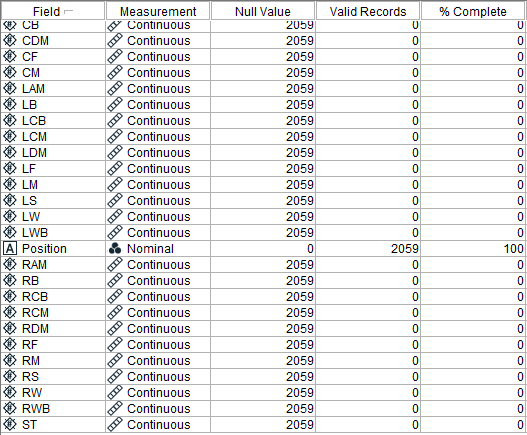
Slika 13: Nedostajuće Vrednosti i Prazni Stringovi

Isto u Quality tabu Data Audit čvora možemo videti nedostajuće vrednosti. Prvo smo tamo gde je klub nepoznat stavili Unknown. Dalje, vidimo da ocene pozicija imaju po 2059 nedostajućih vrednosti. Međutim, razgledanjem skupa smo videli da Manuel Neuer nema ocene za druge pozicije i zato ćemo prvo izvući golmane i skupa i pogledati neke njihove osobine.



Slika 14: Raspodele Golmanskih Atributa Kod Golmana

Prvo što primećujemo je da golmanske osobine sad imaju mnogo lepše i približnije normalnoj raspodele. Međutim, sada ako pogledamo nedostajuće vrednosti:



Slika 15: Nedostajuće Vrednosti Kod Golmana

Vidimo da svih onih 2059 nedostajućih vrednosti za pozicije imaju golmani! Nećemo naravno isključiti sve golmane iz skupa, samo ćemo isfiltrirati kolone sa ocenama pozicija.

Ostale nedostajuće vrednosti ćemo zameniti sa nasumičnim brojevima iz normalne raspodele za taj atribut, za sve kolone, osim za Value i Wage in Euro. Pošto se te dve raspodele drastično razlikuju od normalne, a nisu približne ni uniformnoj raspodeli, promenili bismo oblik raspodele ako bismo nedostajuće vrednosti menjali nasumičnim vrednostima iz te dve raspodele. Takođe, ako bismo nekom konstantom menjali nedostajuće vrednosti, nagomilavale bi nam se te vrednosti u skupu. Zato ćemo, u slučaju kada u analizi koristimo platu i vrednost igrača, eliminisati one redove koji ne poseduju jedan od te dve vrednosti.

**Uzorkovanje**

Za pravila pridruživanja u kojima se koristi ocena pozicije, korišćeni su svi igrači koji nisu golmani. Tamo gde se ne koriste ocene pozicija, koriste se i svi golmani. Ako se koristi kao telo ili glava plata ili cena fudbalera, onda se koristi uzorak koji ne sadrži igrače (i golmane i ostale) koji nemaju vrednost ili Value ili Wage in Euro.

Na kraju, uzeto je pristrasno uzorkovanje, tj. urađena su pravila pridruživanja za Lige 5, 5 najbolje rangiranih liga. Izdvojeni su igrači iz klubova koji su sezoni 2017/2018 igrali u Premijer ligi, La ligi, Bundesligi, Ligi 1 i Seriji A.

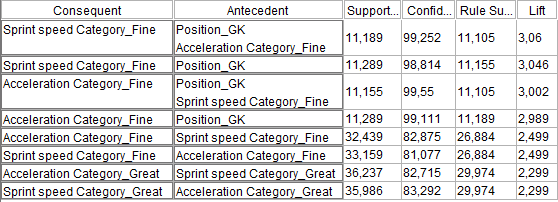
**Transformacija podataka**

Svi podaci koji su korišćeni u pravilima pridruživanja su bili u tabelarnom formatu. Izvršena je diskretizacija svih korišćenih neprekidnih podataka. Svi neprekidni podaci su podeljeni u 3 intervala, tako da svaki interval ima jednak broj elemenata. Korišćen je Binning čvor za to. Dalje su svi kategorički podaci korišćeni u pravilima pridruživanja binarizovani korišćenjem čvora SetToFlag.

**Pravila pridruživanja**

**1.**

Hoćemo da vidimo kako utiču ubrzanje (Acceleration) i brzina trčanja (Sprint speed) na poziciju koju igrač igra. Prvo je iskorišćen čvor Apriori, sa podrazumevanim vrednostima, podrška tela je 10%, a pouzdanost je 80%. Dobijeno je 8 pravila:



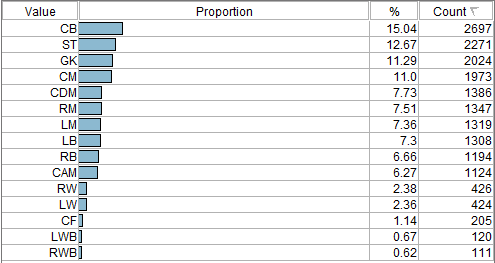
Slika 16: Podrazumevani Apriori

Najbolje pravilo je da igrač ako je golman i ima loše ubrzanje, onda on ima i lošu brzinu. U ovom skupu, 11% su golmani koji imaju loše ubrzanje, u 99% slučajeva oni imaju i lošu brzinu. Oko 11% skupa su zapravo golmani sa lošim ubrzanjem i brzinom. Lift mera od 3,06 nam govori da se 3 puta češće javlja loša brzina kod igrača koji su golmani sa lošim ubrzanjem, nego kod drugih igrača.

Takođe, oni igrači koji imaju lošu brzinu, imaju 2,5 veće šanse da imaju i loše ubrzanje. Loša brzina se javlja kod 32,5% igrača, od toga njih skoro 83% imaju i loše ubrzanje. Igrači koji imaju i lošu brzinu i ubrzanje javljaju se u oko 27% skupa.

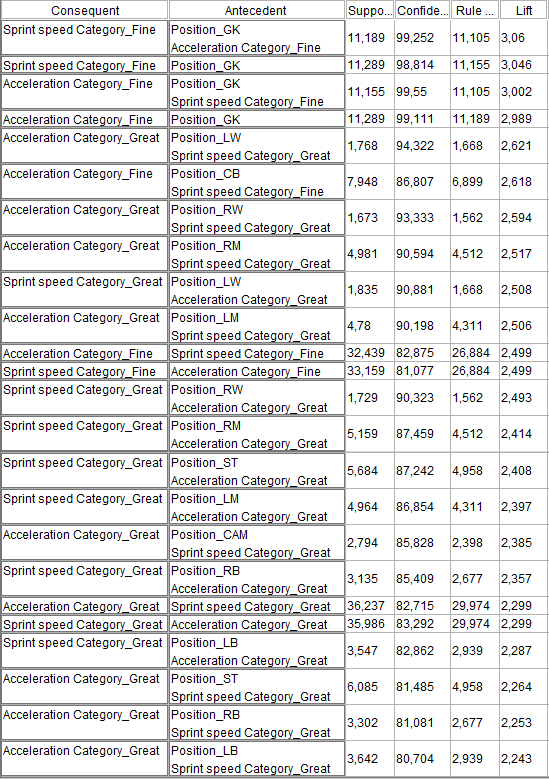
Uočimo još da ako igrač ima odlično ubrzanje, ima oko 2 puta veće šanse da ima i odličnu brzinu. Trećinu skupa čine igrači koji imaju i odličnu brzinu i odlično ubrzanje.

Zašto smo dobili ovakve rezultate? Sve pozicije i ubrzanje i brzinu smo postavili da budu i glava i telo, a u skupu nalazimo samo golmane.



Slika 17: Raspoređenost Igrača Po Pozicijama

Odgovor nalazimo kad pogledamo kako su raspoređene vrednosti pozicija. Golmana ima među najzastupljenijima u skupu, dok levi i desni krilni bek ne čine čak ni jedan posto skupa. Zato smo te pozicije i izbacili iz skupa, a podršku tela smo stavili da bude 1% da bi levo i desno krilo imali bar kakve takve šanse da se pojave u skupu. Pouzdanost smo ostavili da bude 80%. Sada smo dobili sledeće rezultate:



Slika 18: Apriori 1% Podrška Tela

Sad vidimo i neke druge pozicije u skupu. U oko 1,7% su igrači koji igraju na poziciji levog krila i imaju

odličnu brzinu. Od toga oko 93% ima i odlično ubrzanje. Levi krilni igrači sa odličnom brzinom i

ubrzanjem čine oko 1,6% skupa. Zaključujemo i da ako su igrači levo krilni i imaju odličnu brzinu, 2,5 puta je verovatnije i da imaju odlično ubrzanje. Slično je i sa igračima koji igraju desno krilo.

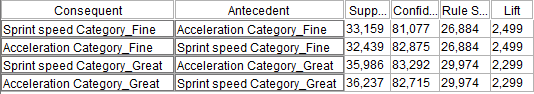
Centralni bek koji ima lošu brzinu ima 2,5 puta veću šansu da ima i loše ubrzanje nego ostali igrači.

Oko 90 posto igrača koji igraju desnog veznog i imaju odličnu brzinu imaju i odlično ubrzanje. Takvi igrači čine oko 4 posto skupa. Slično je i sa levim veznim.

Napadači sa odličnim ubrzanjem čine oko 5,5 posto skupa. Od njih, 87 posto ima i odličnu brzinu. Napadači sa obe karakteristike čine oko 5 posto skupa.

Levi bek sa odličnim ubrzanjem ima oko 2 puta veće šanse da ima i odličnu brzinu. Isto je i sa desnim bekom.

Primenili smo i Carma čvor sa podrazumevanom podrškom i pouzdanošću pravila od 20%. Sledeći rezultati su dobijeni:



Slika 19: Podrazumevana Carma

Neka pravila su, naravno ista kao što smo i dobili prvi put. Ali što smo sad dobili samo 4 pravila i ni jedno od njih ne uključuje ni jednu poziciju? Pošto je podrška pravila 20%, to bi značilo da bar na jednu četvrtinu skupa bude primenljivo pravilo. Ali ni jedna pozicija sama ne čini 20% skupa (najzastupljenija je pozicija CB sa oko 15%). Radi poređenja, stavili smo minimalnu podršku pravila za jedan manju nego što je najmanja dobijena sa Apriori algoritmom, kako bi bilo uključeno to pravilo sa najmanjom podrškom, i istu pouzdanost. Sada smo dobili 25 pravila, umesto 24. Zašto? Zato što sada imamo i pravila koja uključuju više od jedne stavke u glavi, gde kod Apriori algoritma to nismo dobili. Analiziraćemo to jedno novo pravilo:

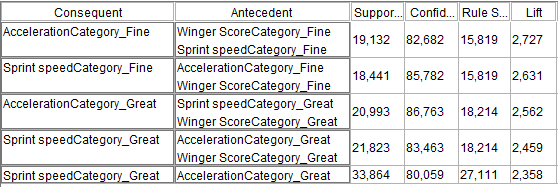


Slika 20: Carma Sa Vrednostima Dobijem Apriori Čvorom

To je zapravo i najbolje pravilo koje smo dobili do sad. Golmani čine oko 11, 3 posto skupa. Od toga, čak 98 posto golmana ima i loše ubrzanje i lošu brzinu. Čak 11% skupa čine golmani sa lošom i brzinom i ubrzanjem, dakle, ako je igrač golman, on ima 3,7 puta veću mogućnost da ima lošu brzinu i ubrzanje.

**2.**

Sada smo hteli da proverimo uticaj brzine i ubrzanja na krila. Uzeta je ocena bolje pozicije (levog ili desnog krila) za svakog igrača (isključujući golmane, jer golmani nemaju ocene za druge pozicije). Podrazumevani Apriori algoritam daje sledeće rezultate:

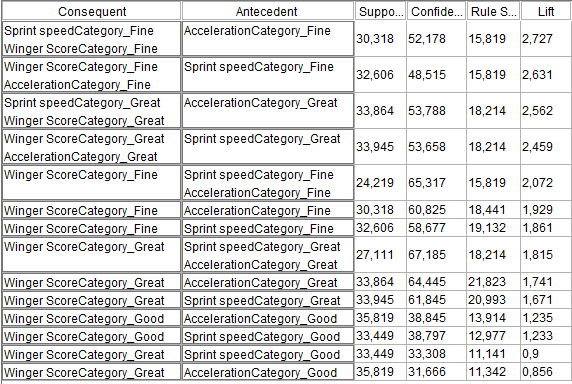


Slika 21: Podrazumevani Apriori

Dakle, ako je igrač ocenjen kao loš krilni i ima lošu brzinu, on ima 2,7 puta veću šansu da ima i loše ubrzanje. Nešto manje od 16% skupa čine igrači koji čine loša krila sa lošom brzinom i ubrzanjem.

S druge strane, ako je igrač odlično ocenjen kao krilo i ima odličnu brzinu, onda on ima 2,5 veću šansu da ima i odlično ubrzanje.

Primenom Carma algoritma sa podrazumevanim vrednostima, dobija se 27 pravila. Ali do sada nismo dobijali neka baš mnogo zanimljiva pravila, jer skoro trećinu skupa, i prvog i drugog na koji smo primenljivali pravila pridruživanja, čine zajedno odlična brzina i ubrzanje. Malo manje čine loša brzina i loše ubrzanje. Zato smo sad filtrirali model, tako da glava bude samo ocena krila. Dobili smo 14 pravila:



Slika 22: Carma, Filtrirani Podrazumevani Model

Nešto manje od 16 posto skupa čine loša krila sa lošom brzinom i ubrzanjem. Dakle, oni sa lošom brzinom i ubrzanjem su 2 puta verovanije loši krilni igrači.

Skoro 2 puta je verovatnije da su odlični krilni igrači ako imaju odličnu brzinu i ubrzanje.

Skoro 40% igrača koji imaju dobru brzinu ili ubrzanje su dobra krila.

**3.**

Tražimo odličnog zadnjeg veznog, ocene karakteristika su postavljene za telo, a za glavu je postavljena ocena koju igrač ima kao zadnji vezni (najveća od 3 pozicije koje pripadaju zadnjem veznom). Podrazumevanim Apriorijem dobijemo 1232 pravila. Filtriravši samo ona koja daju odličnog zadnjeg veznog, dobijamo 778 pravila. Zato smo ponovo pokrenuli Apriori sa podrškom tela od 20%, a pouzdanost pravila 95%. Dobili smo jedno pravilo:



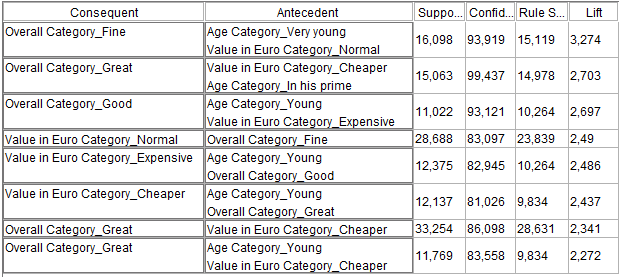
Slika 23: Apriori, Zadnji Vezni

Dakle, ako igrač odlično čita igru i presreće pasove i ima odlične duge pasove, onda je on 98% odličan zadnji vezni.

Na ovaj skup nismo primenili Carma algoritam, jer on radi na drugačiji način i ne prihvata ulogu koju treba da ima neko polje.

**4.**

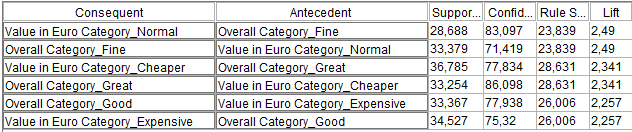
Proverili smo kako utiču godine, cena igrača i njegova ukupna ocena jedno na drugo. Koristili smo podrazumevani Apriori:



Slika 24: Podrazumevani Apriori

Vidimo da veoma mladi igrači koji imaju srednju cenu, čak 94% njih je lošije ocenjeno. Međutim, skoro svi koji su jeftini igrači, ali su malo stariji, su odlični igrači. Mladi igrači koji su najskuplji, u čak 93 posto su i dobri igrači. A oko 28% celog skupa su igrači koji su najjeftiniji, a odlični su igrači.

Sa podrazumevanim Carma algoritmom, ali sa menjanjem podrške na početku traženja čestih skupova:

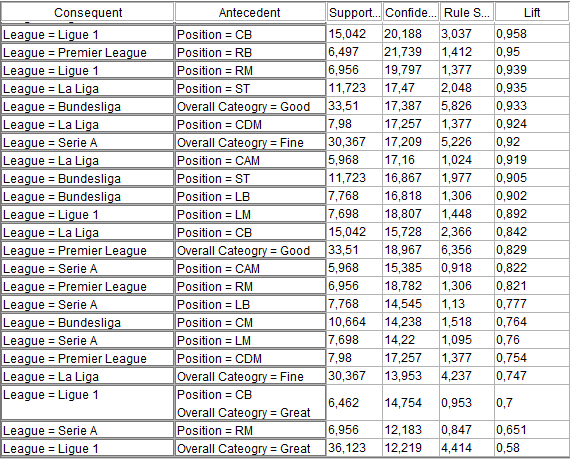


Slika 25: Carma Sa Menjanjem Podrške

Vidimo da je nešto manje od trećine skupa dobri igrači koji su skupi, a nešto manje su loši igrači koji imaju srednju cenu.

**5.**

Na kraju smo pokušali da vidimo da li ukupna ocena i pozicija utiču na Ligu. Za podršku tela je postavljeno 5%, a 10 za pouzdanost pravila. Dobijeno je 70 pravila, ali nisu toliko zanimljiva, osim ovih na dnu:



Slika 26: Apriori 5, 10

Poslednje pravilo nam kaže da ako je igrač odličan, skoro 2 puta je manje verovatno da je iz Lige 1. Slično, malo manje od toga je ako je igrač desni vezni, da igra u Seriji A.

Carma algoritmom, sa 1% podrška pravila i 10% pouzdanost pravila, dobijemo dva relativno zanimljivija pravila:



Slika 27: Carma 1, 10

Pola golmana iz Lige 1 su loši, a isto pola centralnih veznih iz Lige 1 su dobri.

**Zaključak**

Ovaj skup daje mnoge mogućnosti za igranje, ali to premašuje moje trenutne mogućnosti. Deo vremena posvećen ovom radu bio je zamenjen i igranjem FIFE, 19, doduše.

U ovom radu, malo korisnije rezultate nam je dao Carma algoritam, iako bih se ja pristrasno opredelila za Apriori, koji je nekako intuitivniji algoritam. Nekoliko urađenih pravila pridruživanja je izbačeno iz rada, pa bih Vam za kraj preporučila da ako šutirate slobodnjake, to nikako ne bude sa golmanima.