

Saobraćajne nesreće u Francuskoj od 2005. do 2016. godine

Jovan Ležaja

473/2018

Matematički fakultet, Beograd

navoj96@gmail.com

Aleksandar Vračarević

434/2016

Matematički fakultet, Beograd

vracarevicaleksandar@gmail.com

September 2, 2019

1 Uvod

Ovaj rad se fokusira na analizu skupa podataka o saobraćajnim nesrećama u Francuskoj od 2005. do 2006. godine. Pozabavićemo se opisom, analizom i pretprocesiranjem datih podataka, a potom ćemo različitim algoritmima pokušati da pronadjemo pravila pridruživanja (eng. *Association rules*) koristeći se alatima koje nudi IBM SPSS Modeler.

2 Opis podataka

Podaci su preuzeti sa <https://www.kaggle.com/ahmedlhlou/accidents-in-france-from-2005-to-2016> i predstavljaju podatke o saobraćajnim nesrećama u Francuskoj prikupljene u period od 2005. do 2016. godine. Kako bismo uopšte pristupili istraživanju skrivenih pravila u okviru ovog skupa, najpre se moramo upoznati sa istim. Naime, skup se sastoji od 5 tabela u .csv formatu. U nastavku ćemo opisati attribute svake od njih.

- `characteristics.csv`
 - **Num_Acc** : identifikator nesreće - numerički
 - **jour** : dan u mesecu - numerički [1-31]
 - **mois** : mesec - numerički [1-12]
 - **an** : poslednje dve cifre godine - numerički [5-16]
 - **hrmn** : vreme u formatu (ssmm) - numerički [1-2.36k]

- **lum** : osvetljenje u trenutku nesreće brojevi [1-5] kodirani na sledeći način:
 - * 1 - dan
 - * 2 - sumrak/zora
 - * 3 - noć bez prisutnog javnog osvetljenja
 - * 4 - noć sa isključenim javnim osvetljenjem
 - * 5 - noć sa uključenim javnim osvetljenjem
- **dep** : INSEE kod odeljenja praćen nulom
- **com** : kod opštine izdat od strane INSEE
- **agg** :
 - * 1 - izvan gradske sredine
 - * 2 - unutar gradske sredine
- **int** : tip raskrsnice [1-9] kodirani na sledeći način:
 - * 1 - van raskrsnice
 - * 2 - X raskrsnica
 - * 3 - T raskrsnica
 - * 4 - Y raskrsnica
 - * 5 - raskrsnica sa više od 4 kraka
 - * 6 - kružni tok
 - * 7 - place
 - * 8 - pružni prelaz
 - * 9 - ostalo
- **atm** : atmosferski uslovi [1-9] kodirani na sledeći način:
 - * 1 - normalni
 - * 2 - slaba kiša
 - * 3 - jaka kiša
 - * 4 - sneg/grâd
 - * 5 - magla/dim
 - * 6 - jak vetar/oluja
 - * 7 - zaslepljujuće vreme
 - * 8 - oblačno
 - * 9 - ostalo
- **col** : tip sudara [1-7] kodiran na sledeći način:
 - * 1 - čeon sudar
 - * 2 - sudar otpozadi
 - * 3 - sudar sa strane
 - * 4 - lančani sudar
 - * 5 - višestruki sudari (više vozila i više sudara)
 - * 6 - drugi sudari
 - * 7 - nesreća bez sudara
- **adr** : poštanska adresa - niska (popunjava se samo za gradske sredine)

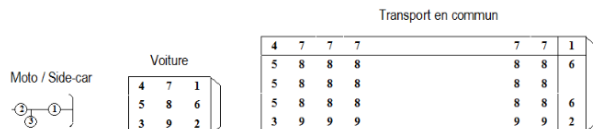
- **gps** : GPS kod - jedan karakter:
 - * M - Métropole
 - * A - Antilles (Martinique or Guadeloupe)
 - * G = Guyane
 - * R = Réunion
 - * Y = Mayotte
- **lat** : geografska širina izražena u broju stepeni
- **long** : geografska dužina izražena u broju stepeni
- **holidays.csv**
 - **ds** : datum nesreće u formatu godina-mesec-dan
 - **holiday** : naziv praznika
- **places.csv**
 - **Num_Acc** : identifikator nesreće - numerički
 - **catr** : kategorija puta [1-9] kodirani na sledeći način:
 - * 1 - autoput
 - * 2 - državni put
 - * 3 - departmentalni putevi
 - * 4 - komunalni putevi
 - * 5 - mreža puteva zabranjena za javnost
 - * 6 - javni parking
 - * 9 - ostalo
 - **voie** : broj puta - numerički
 - **V1** : numerički indeks broja puta (na primer: 2 bis, 3 ter itd.)
 - **V2** : alfanumerički indeks puta
 - **circ** : tip saobraćanja [1-4] kodiran na sledeći način:
 - * 1 - jednosmerna ulica
 - * 2 - dvosmerna ulica
 - * 3 - razdvojen kolovoz
 - * 4 -
 - **nbv** : ukupan broj traka na putu - numerički
 - **vosp** : indikator postojanja rezervisane trake [1-3], nezavisno od toga da li se nesreća dogodila u toj traci, kodiran na sledeći način:
 - * 1 - biciklistička traka
 - * 2 - parking za bicikle
 - * 3 - rezervisan kanal
 - **prof** : kategorije puta [1-4] zavisno od nagiba puta, kodirane na sledeći:
 - * 1 - “dish”
 - * 2 - nizbrdica

- * 3 - vrh brda
- * 4 - dno brda
- **pr** : PR broj kuće - numerička vrednost
- **pr1** : udaljenost od najbližeg PR broja izražena u metrima - numerička vrednost
- **plan** : izgled puta na mapi [1-4], kodirano na sledeći način:
 - * 1 - prav put
 - * 2 - zakrivljen ulevo
 - * 3 - zakrivljen udesno
 - * 4 - “S” oblika
- **lartpc** : širina ostrva na ulici, ako postoji - niska
- **larrout** : širina puta namenjena za saobraćaj - niska
- **surf** : stanje terena [1-9], kodiran na sledeći način:
 - * 1 - normalan
 - * 2 - vlažan
 - * 3 - teren
 - * 4 - potopljen
 - * 5 - sneg na terenu
 - * 6 - blatnjav
 - * 7 - poledica na terenu
 - * 8 - masan/zauljen teren
 - * 9 - ostalo
- **infra** : infrastruktura puteva [1-7], kodirana na sledeći način:
 - * 1 - podzemni tunel
 - * 2 - most/nadvožnjak
 - * 3 - uključenje
 - * 4 - pruga
 - * 5 - “carrefour arranged”
 - * 6 - pešačka zona
 - * 7 - ostalo
- **situ** : pozicija nesreće [1-5], kodirana na sledeći način:
 - * 1 - na putu
 - * 2 - u zaustavnoj traci
 - * 3 - na ivičnjaku
 - * 4 - na trotoaru
 - * 5 - na biciklističkoj stazi
- **env1** : locirano blizu škole - numerička vrednost

● **users.csv**

- **Acc_number** : identifikator nesreće - numerički
- **Num_Veh** : identifikator vozila - alfanumerički

- **place** : pozicija osobe u vozilu u vreme nesreće, kodirano u skladu sa sledećom slikom:



- **catu** : uloga osobe u saobraćaju u trenutku nesreće [1-4], kodirano na sledeći način:
 - * 1 - vozač
 - * 2 - putnik
 - * 3 - pešak
 - * 4 - pešak na rolerima ili skuteru
- **grav** : ozbiljnost povrede [1-4], kodirana na sledeći način:
 - * 1 - neozledjen
 - * 2 - ubijen
 - * 3 - hospitalizovan
 - * 4 - blaga ozleda
- **sex** : pol osobe:
 - * 1 - muško
 - * 2 - žensko
- **Year.on** : godina rođenja - numerički
- **trip** : razlog putovanja [1-9], kodiran na sledeći način:
 - * 1 - kuća-posao
 - * 2 - posao-kuća
 - * 3 - kupovina
 - * 4 - poslovni put
 - * 5 - razonoda
 - * 9 - ostalo
- **secu** : niska koja se sastoji od 2 broja. Prvi označava postojanje sigurnosne opreme [1-9], kodirano na sledeći način:
 - * 1 - pojas za vezivanje
 - * 2 - kaciga
 - * 3 - sedeljka za decu
 - * 4 - reflektujuća oprema
 - * 9 - ostalo

Drugi označava korišćenje sigurnosne opreme [1-3], kodirano na sledeći način:

 - * 1 - oprema je korišćena
 - * 2 - oprema nije korišćena
 - * 3 - neodređeno
- **locp** : pozicija pešaka [1-8], kodirano na sledeći način:

- * 1 - više od 50 metara od pešačkog prelaza
- * 2 - manje od 50 metara od pešačkog prelaza
- * 3 - na pešačkom prelazu sa semaforom
- * 4 - na pešačkom prelazu bez semafora
- * 5 - na trotoaru
- * 6 - na ivičnjaku
- * 7 - pod zaklonom
- * 8 - u prolazu
- **actp** : akcija pešaka [0-9], kodirano na sledeći način:
 - * 0 - neodređeno
 - * 1 - kreće se u istom smeru kao i vozilo sa kojim se dogodio sudar
 - * 2 - kreće se u suprotnom smeru kao i vozilo sa kojim se dogodio sudar
 - * 3 - prelazak ulice
 - * 4 - zaklonjen
 - * 5 - u trku
 - * 6 - sa životinjom
 - * 9 - ostalo
- **etatp** : kategorička vrednost koja određuje da li je pešak bio u društvu drugih ljudi ili ne, kodirano na sledeći način:
 - * 1 - sam
 - * 2 - sa saputnikom
 - * 3 - u grupi ljudi
- **vehicles.csv**
 - **Num_Acc** : identifikator nesreće - numerički
 - **Num_veh** : identifikator vozila - alfanumerički kod
 - **GP** :
 - **CATV** : kategorija vozila [01 - 13]
 - * 01 - bicikl
 - * 02 - moped ; 50 kubika
 - * 03 - kvadricikl sa motorom
 - * 04 - suvišno od 2006. (registrovani skuter)
 - * 05 - suvišno od 2006. (motocikl)
 - * 06 - suvišno od 2006. (putnička prikolica za motocikl)
 - * 07 - VL
 - * 08 - neupotrebljena kategorija (VL i karavan)
 - * 09 - neupotrebljena kategorija (VL i prikolica)
 - * 10 - VU
 - * 11 - najviše korišćeno posle 2006. godine (VU(10) + karavan)
 - * 12 - najviše korišćeno posle 2006. godine (VU(10) + prikolica)
 - * 13 - PL samo 3.5T
 - * 14 -

3 Analiza i pretprocesiranje podataka

Prilikom učitavanja tabele *characteristics* smo uočili da je usled loše formatirane datoteke došlo do pogrešne reprezentacije podataka, što smo razrešili jednostavnom *Python* skriptom. Analizirajući tabelu *characteristics* uočili smo da atributi *gps*, *lat* i *long* imaju značajan broj nedostajućih vrednosti (preko 50%), a s obzirom da zamena nekom konkretnom vrednošću nema smisla zato što nemamo dovoljno validnih vrednosti u koloni da njihova zamena bude smisljena, odlučili smo da ih uklonimo, jer smatramo da nam nisu bitni za dalju analizu. Kada je reč o atributima *atm* i *col*, zbog izuzetno malog broja nedostajućih vrednosti (atributi su bili kompletni blizu 100%), u čvoru *Type* smo ih odbacili, jer ne gubimo ništa odbacivanjem tako malog broja podataka. U tabeli se isto tako nalaze i atributi vezani za lokaciju nesreća (ulica, opština, itd.), ali dodatnim posmatranjem smo primetili da je format zapisa tih podataka dosta nekonzistentan, tako da je njihova korisnost dovedena u pitanje, pošto bez iscrpnog analiziranja teksta ne bismo mogli da izvučemo korisne informacije, što je dovelo do odluke da preko čvora *Type* tim atributima postavimo ulogu (‘‘Role’’) na vrednost *None*.

Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Empty String	White Space	Blank Value
Num_Acc	Continuous	0	0 None	Never	Fixed		100	839985	0	0	0	0
an	Continuous	0	0 None	Never	Fixed		100	839985	0	0	0	0
mois	Continuous	0	0 None	Never	Fixed		100	839985	0	0	0	0
jour	Continuous	0	0 None	Never	Fixed		100	839985	0	0	0	0
hrmm	Continuous	0	0 None	Never	Fixed		100	839985	0	0	0	0
lum	Continuous	0	0 None	Never	Fixed		100	839985	0	0	0	0
agg	Continuous	0	0 None	Never	Fixed		100	839985	0	0	0	0
int	Continuous	20242	0 None	Never	Fixed		100	839985	0	0	0	0
atm	Continuous	41494	0 None	Never	Fixed		99.993	839930	55	0	0	0
col	Continuous	0	0 None	Never	Fixed		99.999	839974	11	0	0	0
com	Continuous	4990	0 None	Never	Fixed		100	839983	2	0	0	0
adr	Categorical	--	--	Never	Fixed		83.268	699438	0	140547	140547	0
gps	Categorical	--	--	Never	Fixed		43.599	366226	0	473759	473759	0
lat	Continuous	0	0 None	Never	Fixed		43.152	362471	477514	0	0	0
long	Continuous	0	2 None	Never	Fixed		42.77	359258	480727	0	0	0
dep	Continuous	0	0 None	Never	Fixed		100	839985	0	0	0	0

Figure 1: Sadržaj *Data Audit* čvora za tabelu *characteristics*

Analizom skupa podataka *users* uočili smo da atributi *locp*, *actp* i *etatp*, koji predstavljaju informacije vezane za pešaka, imaju značajan broj neodređenih vrednosti (preko 50%), tako da smo te kolone izbacili iz skupa podataka *users*. Kada je u pitanju atribut *secu*, čije su vrednosti predstavljene kao dva broja, uočili smo nekonzistentnost odredenih polja sa zadatim opisom reprezentacije tog atributa, tako da smo te nekonzistentne vrednosti preimenovali u *NA* (neodređenu vrednost). Za svaki atribut koji je imao 0 kao vrednost, a nije bilo definisano šta ta vrednost predstavlja, 0 je zamenjena sa *NA*. Za atribut *place* smo sve vrednosti ostavili kakve jesu, pošto je šema koja predstavlja kodiranje bila nedovoljno jasna.

Skup podataka *vehicles* smo analizirali i zaključili da sve slogove koji sadrže nedostajuće i neodređene vrednosti možemo da odbacimo. Nismo naišli ni na kakve nepravilnosti koje iziskuju detaljnije procesiranje.

Nakon što smo uvideli da kolone *v1* i *v2* skupa *places* sadrže ogroman broj nedostajućih vrednosti, odbacili smo ih. Pošto se u kolonama *pr* i *pr1* javlja preko 50% nedostajućih vrednosti, a smatramo da ne postoji smislen način da te

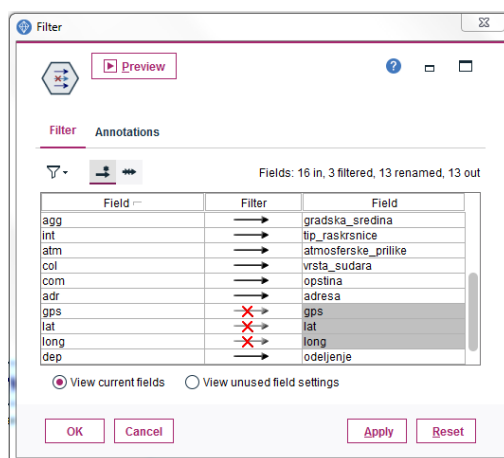


Figure 2: Sadržaj *Filter* čvora za tabelu *characteristics*

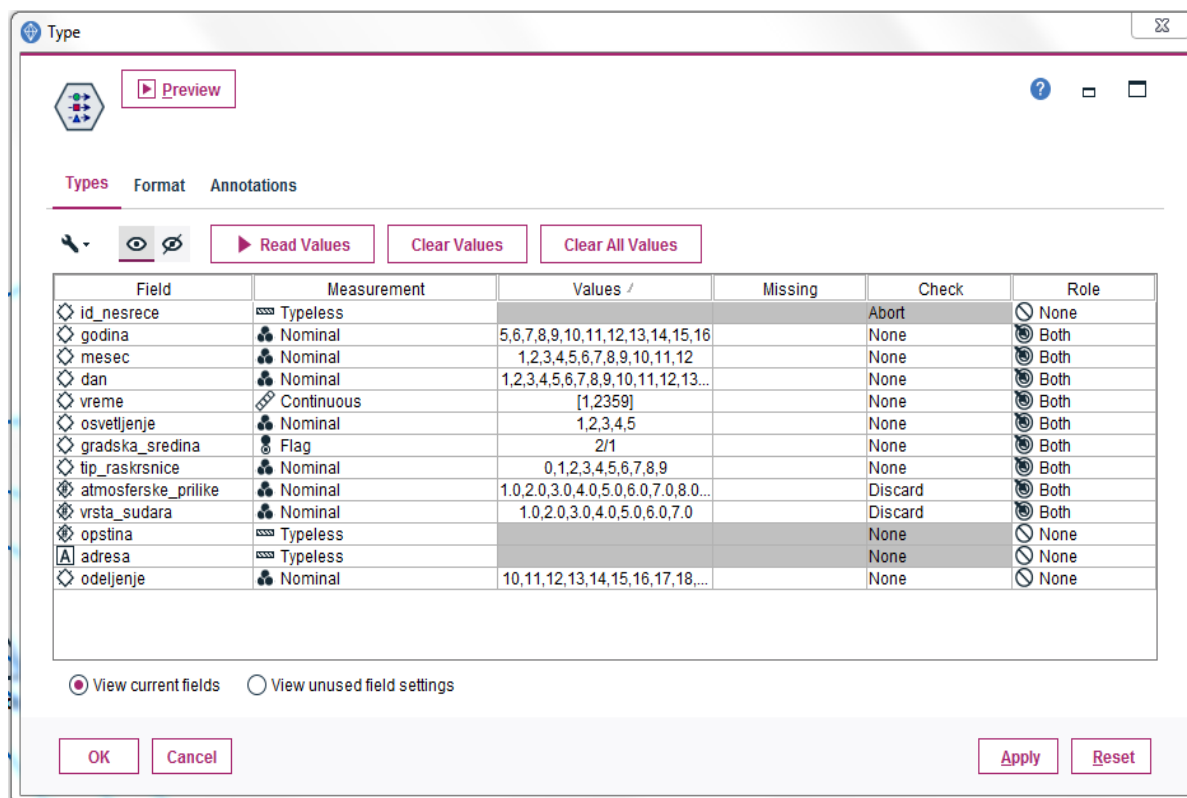


Figure 3: Sadržaj *Type* čvora za tabelu *characteristics*

vrednosti popunimo, odbacili smo i ove kolone. Kolona *env1* predstavlja predstavlja meru blizine školi, ali je zbog nejasnog kodiranja i ova kolona odbačena. Iako kolone *voie*, *vosp*, *lartpc*, *infra* i *nbv* nemaju puno nedostajućih vrednosti, gotovo svi slogovi uzimaju mali skup vrednosti za pomenute attribute pa

smo se odlučili da ni ove attribute ne koristimo u daljoj analizi. Za kolone `situ`, `prof`, `surf` i `plan` ćemo odbaciti slogove sa vrednošću nula za ove attribute. U koloni `larroust` se javljaju negativne vrednosti za širinu puta, pa ćemo i njih ukloniti.

Skup podataka `holidays` smo odlučili da ne koristimo za dalju analizu, jer ne sadrži preterano korisne informacije.

4 Pravila pridruživanja

Nakon što smo obradili skupove podataka, hteli smo da na svaki od relevantnih skupova primenimo algoritme *Apriori* i *Carma*, u nadi da ćemo uočiti neka zanimljiva pravila. Nakon primene pomenutih algoritama, cilj nam je bio da primenimo iste algoritme nad objedinjenim podacima.

4.1 Primena *Apriori* i *Carma* algoritama nad skupom characteristics

Iskoristili smo niz čvorova *Reclassify* kako bismo lakše tumačili kategoričke vrednosti. Iz tog skupa smo filtrirali one slogove čija je vrednost atributa *tip_raskrsnice* NA. Potom smo primenili *Apriori* sa podrazumevanim podešavanjima (minimalna podrška uzročnika je 10%, a minimalna pouzdanost pravila je 80%) i rezultati izvršavanja tog algoritma se mogu videti na slici. Posledice pravila se odnose na atmosferske prilike, tip raskrsnice i indikator da li se nesreća desila u gradu ili ne. Algoritam je uspeo da nadje 30 pravila, koja sve u svemu nisu zanimljiva. Naime, lift mera se kreće u opsegu od 0.998 do 1.314, što nam govori da su uzročnici u blagoj korelaciji sa posledicama. Kako bismo pripremili skup podataka za algoritam *Carma*, koristili smo čvor *SetToFlag*. *Carma* algoritam smo primenili sa istim parametrima kao i *Apriori*. Dobili smo 22 pravila, koja su gotovo identična onima koje smo dobili korišćenjem *Apriori* algoritma.

Kao što se iz rezultata *Data Audit* čvora može primetiti, određene vrednosti nekih atributa dominiraju nad ostalim vrednostima, te stoga ne čudi što otkrivena pravila sadrže te vrednosti.

U želji da izbor pravila bude pravedniji, odlučili smo da izbalansiramo skup podataka, tako što ćemo korišćenjem čvorova *Balance* na pojedinačne kolone ublažiti efekat dominantnih vrednosti (pomenute čvorove smo generisali uz pomoć distribucija odgovarajućih kolona). Nakon toga smo redom primenjivali *Apriori* algoritam za svaku izmenjenu kolonu. Potom smo eksperimentisali sa primenom *Apriori* algoritma na ulančane *Balance* čvorove. Neki od rezultata su predstavljeni na narednim slikama.

Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
godina		Nominal	5	16	--	--	--	12	839921
mesec		Nominal	1	12	--	--	--	12	839921
dan		Nominal	1	31	--	--	--	31	839921
vreme		Continuous	1	2359	1381.087	540.758	-0.488	--	839921
osvetljenje		Nominal	--	--	--	--	--	5	839921
gradska_sredina		Flag	1	2	--	--	--	2	839921
tip_raskrsnice		Nominal	--	--	--	--	--	10	839921
atmosferske_prilike		Nominal	--	--	--	--	--	9	839921
vrsta_sudara		Nominal	--	--	--	--	--	7	839921

Figure 4: Možemo videti da se u kolonama *osvetljenje*, *tip_raskrsnice* i *atmosferske_prilike* u najvećem broju slučajeva javlja samo jedna vrednost, te ćemo te vrednosti pokušati da izbalansiramo. Još jedna kolona na koju ćemo primeniti balansiranje je *vrsta_sudara*.

tip_raskrsnice = van_raskrsnice	osvetljenje = noc_bez_osvetljenja	20.042	91.495	1.221	1.704
tip_raskrsnice = van_raskrsnice	osvetljenje = noc_bez_osvetljenja atmosferske_prilike = normalne	14.129	91.774	1.225	1.162
gradska_sredina	osvetljenje = noc_sa_osvetljenjem tip_raskrsnice = van_raskrsnice	12.364	85.211	1.402	1.829
gradska_sredina	osvetljenje = noc_sa_osvetljenjem	20.059	89.59	1.474	2.088
gradska_sredina	tip_raskrsnice = X_raskrsnica	10.605	89.611	1.475	1.102
gradska_sredina	osvetljenje = noc_sa_osvetljenjem atmosferske_prilike = normalne	15.254	89.79	1.478	1.557

Figure 5: Rezultat primene apriori algoritma na balansiranu kolonu *osvetljenje*. Izdvojena pravila jesu logična, ali nam ne otkrivaju puno interesantnih zaključaka.










<div>  File  Generate  Preview  </div>				
<div> Model Settings Summary Annotations </div>				
<div>   Sort by: Lift    33 of 33 </div>				
Consequent	Antecedent	Support %	Confidence %	Lift
gradska_sredina	osvetljenje = noc_sa_osvetljenjem atmosferske_prilike = normalne	11.002	86.196	1.436
gradska_sredina	osvetljenje = noc_sa_osvetljenjem	14.527	85.957	1.432
tip_raskrsnice = van_raskrsnice	vrsta_sudara = lancani_sudar_3+ atmosferske_prilike = normalne	11.582	88.58	1.141
tip_raskrsnice = van_raskrsnice	vrsta_sudara = lancani_sudar_3+ osvetljenje = dan	10.8	88.383	1.139
tip_raskrsnice = van_raskrsnice	vrsta_sudara = lancani_sudar_3+	14.314	88.361	1.138
tip_raskrsnice = van_raskrsnice	vrsta_sudara = bez_sudara atmosferske_prilike = normalne	10.965	86.439	1.114
tip_raskrsnice = van_raskrsnice	vrsta_sudara = bez_sudara	14.295	85.74	1.105
atmosferske_prilike = normalne	vrsta_sudara = sudar_sa_strane osvetljenje = dan	10.591	86.741	1.079
atmosferske_prilike = normalne	gradska_sredina osvetljenje = dan tip_raskrsnice = van_raskrsnice	29.468	85.681	1.066
atmosferske_prilike = normalne	vrsta_sudara = sudar_odpozadi osvetljenje = dan	10.418	85.369	1.062
atmosferske_prilike = normalne	gradska_sredina osvetljenje = dan	42.791	85.133	1.059
atmosferske_prilike = normalne	vrsta_sudara = sudar_sa_strane gradska_sredina	11.384	85.106	1.059
tip_raskrsnice = van_raskrsnice	vrsta_sudara = ceoni_sudar	14.11	81.481	1.05
atmosferske_prilike = normalne	vrsta_sudara = sudar_sa_strane	14.455	84.248	1.048
atmosferske_prilike = normalne	vrsta_sudara = lancani_sudar_3+ osvetljenje = dan	10.8	83.85	1.044
tip_raskrsnice = van_raskrsnice	vrsta_sudara = sudar_odpozadi	11.807	80.966	1.043

Figure 6: Rezultat primene apriori algoritma na balansiranu kolonu *vrsta_sudara*. U ovom slucaju je pronadjeno 33 pravila od kojih su ona sa najvećom lift merom logična ali i dalje nam ne daju upotrebljiviji uvid u zavisnosti medju atributima.

Zaključujemo da balansiranje pojedinačnih kolona ne dovodi do željenih rezultata te smo probali sa ulančanim balansiranjem.

Consequent	Antecedent	Support %	Confidence %	Lift
gradska_sredina	osvetljenje = noc_sa_osvetljenjem	20.299	85.784	1.505

Figure 7: Rezultat primene apriori algoritma na redom izbalansirane sve kolone skupa. Izdvojeno pravilo prema lift meri jeste zanimljivo ali je opet očekivano da u gradskoj sredini postoji osvetljenje koje je uključeno.

Consequent	Antecedent	Support %	Confidence %	Lift
osvetljenje = dan	atmosferske_prilike = bljestavo_vreme	14.316	92.862	1.582
gradska_sredina	osvetljenje = noc_sa_osvetljenjem	19.314	87.437	1.319
gradska_sredina	vrsta_sudara = ostalo	17.067	82.009	1.237
gradska_sredina	atmosferske_prilike = lagana_kisa	11.662	81.935	1.236

Figure 8: Rezultat primene apriori algoritma na balansirane kolone *tip_raskrsnice*, *vrsta_sudara* i *atmosferske_prilike*.

Kako pokušaj sa balansiranjem nije prošao slavno, odlučili smo se da potpuno eliminišemo slogove koji imaju najzastupljeniju vrednost određenog atributa, i da na njega primenimo iste algoritme. Na ovaj način smo otkrili više pravila nego u prethodnim pokušajima, sa lift merama u opsegu od 0.673 do 1.421, ali sa malom pouzdanošću i podrškom. Rezultati se mogu videti na slici.

Consequent	Antecedent	Support %	Confidence %	Lift
vrsta_sudara = bez_sudara	osvetljenje = noc_bez_osvetljenja atmosferske_prilike = normalne tip_raskrsnice = van_raskrsnice	6.993	38.917	1.421
vrsta_sudara = bez_sudara	tip_raskrsnice = kruzni_tok	2.202	38.626	1.41
vrsta_sudara = bez_sudara	osvetljenje = noc_bez_osvetljenja atmosferske_prilike = normalne	7.424	38.573	1.408
vrsta_sudara = bez_sudara	osvetljenje = noc_bez_osvetljenja tip_raskrsnice = van_raskrsnice	9.814	38.531	1.407
vrsta_sudara = bez_sudara	osvetljenje = noc_bez_osvetljenja	10.426	38.177	1.394
vrsta_sudara = sudar_odpozadi	tip_raskrsnice = T_raskrsnica osvetljenje = dan atmosferske_prilike = normalne	3.121	37.868	1.252
vrsta_sudara = sudar_odpozadi	tip_raskrsnice = T_raskrsnica osvetljenje = dan	3.745	37.572	1.243
vrsta_sudara = sudar_odpozadi	tip_raskrsnice = kruzni_tok	2.202	36.952	1.222
vrsta_sudara = sudar_odpozadi	atmosferske_prilike = normalne	4.209	36.385	1.203
vrsta_sudara = sudar_odpozadi	tip_raskrsnice = T_raskrsnica	5.236	35.776	1.183
vrsta_sudara = bez_sudara	atmosferske_prilike = lagana_kisa osvetljenje = noc_sa_osvetljenjem	2.162	33.791	1.233
vrsta_sudara = ceoni_sudar	tip_raskrsnice = X_raskrsnica atmosferske_prilike = normalne	4.698	33.639	1.29
vrsta_sudara = bez_sudara	osvetljenje = noc_sa_osvetljenjem tip_raskrsnice = van_raskrsnice	10.234	33.585	1.226

Figure 9: Nakon izbacivanja slogova koji bi 'prigušili' ostatak skupa, dobijeni su ovakvi rezultati. I dalje smatramo da ne postoje izuzetno zanimljiva pravila.

Model Settings Summary Annotations				
<div> Sort by: Lift of 171 </div>				
Consequent	Antecedent	Support %	Confidenc...	Lift
zastita = 11	manevar = 1 kategorija_vozila = 7 vrsta_raskrsnice = 1 kategorija_ucesnika = 1 pokretna_prepreka = 2	10.287	90.531	1.564
zastita = 11	manevar = 1 kategorija_vozila = 7 vrsta_raskrsnice = 1 osvetljenje = 1 kategorija_ucesnika = 1	10.007	89.416	1.545
zastita = 11	manevar = 1 kategorija_vozila = 7 vrsta_raskrsnice = 1 kategorija_ucesnika = 1	14.147	89.119	1.539
zastita = 11	manevar = 1 kategorija_vozila = 7 osvetljenje = 1 kategorija_ucesnika = 1 pokretna_prepreka = 2	11.629	88.83	1.534
zastita = 11	manevar = 1 kategorija_vozila = 7 vrsta_raskrsnice = 1 kategorija_ucesnika = 1 atmosferske_prilike = 1.0	11.147	88.799	1.534
zastita = 11	pol kategorija_vozila = 7 vrsta_raskrsnice = 1 kategorija_ucesnika = 1	11.426	88.695	1.532

Figure 11: Rezultati apriori algoritma sa podrazumevanim parametrima.

Model Settings Summary Annotations				
<div> Sort by: Lift 2290 of 2290 </div>				
Consequent	Antecedent	Support %	Confidence %	Lift
vrsta_raskrsnice_1 kategorija_ucesnika_1 kategorija_vozila_7	zastita_11 osvetljenje_1	41.425	49.011	1.518
ozbilnost_1 kategorija_vozila_7 pokretna_prepreka_2	atmosferske_prilike_1.0 zastita_11	46.355	43.773	1.515
atmosferske_prilike_1.0 zastita_11	ozbilnost_1 kategorija_vozila_7 pokretna_prepreka_2	28.901	70.207	1.515
kategorija_vozila_7 pokretna_prepreka_2 osvetljenje_1	atmosferske_prilike_1.0 kategorija_ucesnika_1 zastita_11	38.314	54.83	1.513
atmosferske_prilike_1.0 kategorija_ucesnika_1 zastita_11	kategorija_vozila_7 pokretna_prepreka_2 osvetljenje_1	36.238	57.972	1.513
zastita_11	vrsta_raskrsnice_1 kategorija_ucesnika_1 kategorija_vozila_7 osvetljenje_1	23.189	87.554	1.512
vrsta_raskrsnice_1 kategorija_ucesnika_1 kategorija_vozila_7 osvetljenje_1	zastita_11	57.893	35.07	1.512
atmosferske_prilike_1.0 kategorija_ucesnika_1 kategorija_vozila_7 osvetljenje_1	ozbilnost_1	45.988	43.699	1.508
ozbilnost_1	atmosferske_prilike_1.0			

Figure 12: Rezultati carma algoritma sa podrazumevanim parametrima.