KLASIFIKACIJA INCIDENATA SA SUDJELOVANJEM BESPILOTNIH LETJELICA ZASNOVANA NA STROJNOM UČENJU

Cestar, Toni

Master's thesis / Diplomski rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka, Faculty of Engineering / Sveučilište u Rijeci, Tehnički fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:190:186797

Rights / Prava: Attribution 4.0 International/Imenovanje 4.0 međunarodna

Download date / Datum preuzimanja: **2024-07-27**

Repository / Repozitorij:

Repository of the University of Rijeka, Faculty of Engineering

SVEUČILIŠTE U RIJECI

TEHNIČKI FAKULTET

Diplomski sveučilišni studij računarstva

Diplomski Rad

KLASIFIKACIJA INCIDENATA SA SUDJELOVANJEM BESPILOTNIH LETJELICA ZASNOVANA NA STROJNOM UČENJU

Rijeka, rujan 2023.

Toni Cestar

0165053880

SVEUČILIŠTE U RIJECI

TEHNIČKI FAKULTET

Diplomski sveučilišni studij računarstva

Diplomski Rad

KLASIFIKACIJA INCIDENATA SA SUDJELOVANJEM BESPILOTNIH LETJELICA ZASNOVANA NA STROJNOM UČENJU

Mentor: Prof. dr. sc. Renato Filjar

Rijeka, rujan 2023

Toni Cestar

0165053880

Zadatak rada

Rastuća primjena bespilotnih letjelica za vojne, komercijalne i privatne potrebe, postavlja njihovo operativno korištenje u potencijalne sporove s postojećim procesima korištenja zračnog prostora. Razumijevanje prirode i uzroka incidenata sa sudjelovanjem bespilotnih letjelica otvara mogućnosti za optimalnu regulaciju zračnog prostora, koja će smanjiti rizik od incidenata te povećati djelotvornost korištenja zračnog prostora, sigurnost svih sudionika, energetsku učinkovitost i održivost takve vrste prometa. U ovom radu potrebno je analizirati zadane baze podataka o incidentima u koje su bile uključene bespilotne letjelice na području Ujedinjenog Kraljevstva u razdoblju od 2000. godine do kraja 2016. godine i tijekom 2017. godine. Temeljem opisne statističke analize podataka, potrebno je izabrati odgovarajuće metode razvoja i razviti klasifikacijski model incidenata sa sudjelovanjem bespilotni letjelica, zasnovan na statističkom učenju. Incidente treba klasificirati sukladno mjestu, vremenu i okolnostima incidenata, prema opisnim statističkim varijablama predstavljenim u bazama podataka. Potrebno je provjeriti uspješnost dobivenih modela te interpretirati postupak razvoja, operativnost i rezultate provjere uspješnosti. Usporediti rezultate klasifikacije za dva navedena razdoblja te komentirati učinke nacionalne regulacije uvedene od 2017. godine, u smislu poboljšanja korištenja zračnog prostora i povećanja sigurnosti. Zaključke potkrijepiti statističkim pokazateljima. Cjelokupno istraživanje provesti vlastitom programskom podrškom u programskom okruženju za statističko računarstvo R.

Izjava o samostalnoj izradi rada

Izjavljujem da sam samostalno izradio ovaj rad.

Rijeka, rujan 2023. _____

Toni Cestar

Zahvala

Izražavam iskrenu zahvalnost osobama koje su značajno doprinijele mom akademskom putu i izradi diplomskog rada. Posebna zahvala ide mentoru, Renatu Filjaru, koji me vodio kroz izradu diplomskog rada, te za njegove savjete, mišljenja i rasprave, kako u vezi s radom tako i šire. Nadalje, pružio mi je obilje kvalitetnih materijala i knjiga, koji su odigrali ključnu ulogu u poboljšanju mog razumijevanja predmeta i rješavanju izazova koje postavlja ovo istraživanje. Također želim izraziti zahvalnost kolegama Danijelu Bonašinu i Tinu Trupeljaku. Zajednički napori ne samo da su učinili studije zanimljivijima, već su i značajno poboljšali njihovu učinkovitost. Uz to, olakšali su mi priviknuti se novoj životnoj okolini u Rijeci. Na kraju, želio bih zahvaliti na podršci roditelja koji su stajali uz mene kroz cjeloživotno obrazovanje te također i kroz trenutni studij.

SADRŽAJ

1. UVOD 1
2. PRETHODNA ISTRAŽIVANJA 2
3. TEZA RADA 3
4. METODOLOGIJA 4
4.1 Klasifikacijski model k-najbližih susjeda 5
4.2 Klasifikacijski model višečlana logička regresija 5
4.3 Klasifikacijski model stroj potpornih vektora 6
4.4 Indikatori uspješnosti klasifikacijskih modela 7
5. REZULTATI ISTRAŽIVANJA 10
5.1 Struktura podataka 11
5.2 Rezultati opisne statističke analize podataka 13
5.3 Rezultati klasifikacijskog modela k-najbližih susjeda 28
5.4 Rezultati klasifikacijskog modela višečlane logističke regresije 34
5.5 Rezultati klasifikacijskog modela stroja potpornih vektora 38
6. INTERPRETACIJA I RASPRAVA 42
7. ZAKLJUČAK 44
8. LITERATURA 46
9. POPIS OZNAKA I KRATICA 47
10. SAŽETAK 48
ABSTRACT 48
11. DODATAK A – INSTALACIJA PROGRAMSKOG OKRUŽENJA R 49
DODATAK B – PREUZIMANJE I POKRETANJE KODA 49

1. UVOD

Sve veći napreci u tehnologiji omogućili su jeftiniju i kvalitetniju izradu bespilotnih letjelica. Budući da su bespilotne letjelice postale lakše i šire dostupne, potaknut je veliki porast privatne upotrebe bespilotnih letjelica koja je dovela do značajnog porasta prisutnosti bespilotnih letjelica u zračnom prostoru. Veća prisutnost bespilotnih letjelica je rezultirala odgovarajućem porastu incidenata povezanih s bespilotnim letjelicama. Prikazano istraživanje se usredotočuje na sveobuhvatnu analizu podataka, pruženih od strane Odbora za situacije u kojima je narušena sigurnost letjelica Ujedinjenog Kraljevstva (en. UK Airprox Board). Pruženi podaci sadrže zapise o incidentima s bespilotnim letjelicama unutar zračnog prostora Ujedinjenog Kraljevstva u razdoblju od 2000. godine do kraja 2016. godine. Poticaj za prikazano istraživanje proizlazi iz potrebe za rješavanjem problema rasta rizika povezanog s operacijama bespilotnih letjelica unutar zračnog prostora Ujedinjenog Kraljevstva. Temeljna motivacija proizlazi iz procjene učinkovitosti regulatornih promjena uvedenih 2017. godine u smislu poboljšavanja iskoristivosti zračnog prostora, smanjenja incidenata, jačanja sigurnosti i konačnog doprinosa poboljšanoj energetskoj učinkovitosti i održivosti u području prometa bespilotnih letjelica. Prikazano istraživanje odvija u više poglavlja, gdje svako poglavlje ima jedinstvenu usredotočenost i doprinos cjelokupnom istraživanju. U sljedećem poglavlju navesti će se prethodna istraživanja koja su se bavila ovim problemom. U trećem poglavlju se govori o tezi rada, načinu nastojanja rješavanja problema, potrebnim preduvjetima, varijablama koje sudjeluju u rješavanju problema te očekivanim rezultatima rješenja. Unutar četvrtog poglavlja, metodologija, prikazat će se svi postupci statističke analize, matematičke metode i modeli, opis korištenih podataka te obrazloženje kako navedeno pridonosi rješavanju problema i koje prednosti donosi usporedno s prethodnim pristupima. Peto poglavlje prikazuje rezultate istraživanja, strukturu napravljenog rješenja te uspješnost pojedinog rješenja. Šesto poglavlje se bazira na interpretaciji i raspravi rezultat istraživanja s obzirom na razmatrani problem i postavljene hipoteze. Uz to, razmatra se uspješnost i ograničenja istraživanja, te načine na koje se može proširiti istraživanje. Unutar zadnjeg poglavlja se donosi zaključak cjelokupnog istraživanja, pregled teze i problema rada te uspješnosti rezultata rada. Nadalje navode se ograničenja i nedostatci trenutnog istraživanja i predlažu se ciljevi narednih istraživanja.

2. PRETHODNA ISTRAŽIVANJA

Izvor informacija i značajnog broja uvida u incidente u zračnom prostoru koji uključuje bespilotne letjelice pruža Odbor za situacije u kojima je narušena sigurnost letjelica Ujedinjenog Kraljevstva. Glavne značajke Odbora obuhvaćaju identifikaciju, analizu i izvješćivanje o čimbenicima koji doprinose incidentima sa situacijama u kojima je narušena sigurnost letjelica. Skup podataka Odbora proteže se od 2000. godine do kraja 2016. godine te proširuje razumijevanje dinamike incidenata povezanih s bespilotnim letjelicama. Tijelo koje definira niz klasifikacija za incidente sa situacijama u kojima je narušena sigurnost letjelica naziva se Međunarodna organizacija civilnog zrakoplovstva (en. International Civil Aviation Organization). Jedna od klasifikacija je kategorija rizika koja se dodjeljuje samo na temelju stvarnoga rizika, a ne potencijalnog rizika. Što znači da se uzima u obzir samo preostali rizik nakon bilo koje radnje izbjegavanja. Istraživanje od strane Odbora [1], kod ispitivanja podataka otkriva zabrinjavajući trend rasta u kategorijama koje nose rizik, kategorije rizika A i B, posebice kod incidenata s malim bespilotnim letjelicama. Važno je uzeti u obzir da male bespilotne letjelice zbog veličine u nekim slučajevima mogu izbjeći otkrivanje što rezultira nedovoljnim prijavljivanjem. Vremenska distribucija incidenata u 2017. godini otkriva veću stopu incidenata tijekom ljetnih mjeseci, koja prati povećanu aktivnost općeg zrakoplovstva (en. General Aviation), što naglašava potrebu za prilagođenim sigurnosnim mjerama u ljetnom razdoblju. Zanimljivo opažanje proizlazi iz analize kategorije rizika za pojedini sektor, dok komercijalni zračni promet (en. Commercial Air Transport) pokazuje iznimno nisku stopu incidenata sa situacijama u kojima je narušena sigurnost letjelica, s trendom pada, sektori vojnog zrakoplovstva (en. Military Aviation) i općeg zrakoplovstva imaju veću učestalost letova koji nose rizik. Opažanje naglašava sektorsku prirodu problema, sugerirajući potrebu za ciljanim intervencijama. Dodatno, Odbor 2017. godine uvodi kampanju „5 sekundi do sudara“ naglašavajući važnost situacijske svijesti pilota za izbjegavanje situacije u kojima je narušena sigurnost letjelica. Dok se prethodno istraživanje bavilo razlogom dolaska do incidenata sa situacijama u kojima je narušena sigurnost letjelica, njihovim brojem i obrazovanjem pilota, istraživanje rada fokusira se na komparativnu analizu podataka od 2000. do kraja 2016. godine i ključnu 2017. godinu. Uvođenje novih regulatornih zakona u 2017. zahtijeva pažljivo ispitivanje njihovog utjecaja na incidente u zračnom prostoru. Osnovni cilj je procijeniti jesu li nove zakonske reforme dovele do smanjenja rizičnih kategorija i ukupnog broja incidenata, pridonoseći time poboljšanoj sigurnosti zračnog prostora.

3. TEZA RADA

Problem trenutnog istraživanja je pružiti uvid u utjecaj novousvojenih propisa na sigurnost zračnog prostora, situacije u kojima je narušena sigurnost letjelica te pripadajuće kategorije rizika. Način rješavanja problema je komparativna analiza podataka do kraja 2016. godine s podacima 2017. godine pruženih od strane Odbora. Za početak istraživanja, uz dobivene podatke, potrebni su određeni preduvjeti. Prije svega pripremiti navedene podatke u oblik spreman za analizu. Jedan od preduvjeta je poznavanje statističkog računanja unutar programskog okruženja R za provođenje opsežne analize podataka. Nadalje potrebno je poznavanje strojnog učenja i razvijanja klasifikacijskih modela. Trenutno istraživanje koristi jedanaest varijabli kao prediktore. Prediktori obuhvaćaju visinu prijavljenog incidenta, klasifikaciju leta oba zrakoplova, zračni prostor oba zrakoplova, lokaciju oba zrakoplova, službu zračnog prometa oba zrakoplova, pravila leta oba zrakoplova. Varijabla ishoda je kategorija rizika pripisana svakom incidentu. Uz rješavanje problema netaknutog prethodnim istraživanjem, dodatna funkcionalnost trenutnog istraživanja je sposobnost klasifikacije incidenata korištenjem klasifikacijskih modela zasnovanim na statičkom učenju. Klasifikacijski modeli se razvijaju za oba vremenska perioda. Kategorije razvijenih klasifikacijskih modela su k-najbližih susjeda, višečlana regresija i stroj potpornih vektora. Svaki klasifikacijski model doprinosi novi pogled na podatke istraživanja. Usporedbom uspješnosti i rezultata klasifikacijskih modela donosit će se zaključci u učincima nacionalne regulacije uvedene 2017. godine. Istraživanje se bavi obradom i proučavanjem podataka kao što je navedeno i temeljna pretpostavka istraživanja je povećana sigurnost zračnog prometa Ujedinjenog Kraljevstva nakon unošenja nacionalne regulacije 2017. godine.

4. METODOLOGIJA

Sve metode trenutnog istraživanja izvedene su u programskom okruženju za statističko računarstvo R. R je programski jezik i okruženje otvorenog koda za statističko računanje i grafičke prikaze. R je GNU projekt i radi na Windows, Linux i MacOS operativnim sustavima. R pruža tehnike linearnih i nelinearnih modeliranja, klasičnih statističkih testova, analiza vremenskih serija, klasifikacija i grupiranja. Za probleme s opsežnim računanjem programsko okruženje R podržava korištenje C, C++ i Fortran koda. Prednosti programskog jezika R-a su lakoća korištenja kvalitetnih grafikona koji mogu uključivati matematičke formule i simbole i širok izbor knjižnica za statističke metode i strojno učenje.

Podaci preuzeti od strane Odbora su pripremljeni makivanjem nepotrebnih stupaca (prediktora). Većina preostalih prediktora sadržala je veliki broj faktora koji su grupirani u logičke cjeline. Uz to, maknuti su svi reci s praznim i nepravilnim unosima.

Opisna statistička analiza podataka navedenog istraživanja koristi knjižnicu DataExplorer [2]. Knjižnica DataExplorer pojednostavljuje početne korake analize i istraživanja podataka te omogućuje razne funkcije i vizualizacije koje doprinose lakšem razumjevanju podataka.

Za razvoj klasifikacijskih modela korištene su caret [3], tidymodels [4] i e1071 [5] knjižnica. Knjižnice se koriste za pojednostavljenje procesa obuke i evaluacije modela strojnog učenja [6] nad zadacima klasifikacije i regresije. Klasifikacijski modeli su kategorija modela strojnog učenja korištenih za probleme kategorizacije i klasifikacije podataka u unaprijed definirane kategorije ili klase. Klasifikacijski modeli su vrsta nadziranog učenja kod kojega algoritam za donošenje odluka uči iz označenih podataka obuke i probe, nakon čega može donositi odluke za klasifikaciju novih, neoznačenih podataka. Kod klasifikacije ciljana varijabla je kategorička, varijabla koja nije numerička i unutar koje elementi nisu numerički povezani. Cilj klasifikacijskog modela je pridjeliti pojedini element pripadajućoj predefiniranoj klasi. Druga vrsta nadziranog učenja su regresijski modeli koji koristi numeričke prediktore za donošenje odluke o vrijednosti numeričke ciljane varijable. Podaci obuke i probe se dobivaju nasumičnim razdvajanjem podataka. Podaci obuke generalno sadrže nekoliko puta više opažanja od podataka probe.

4.1 Klasifikacijski model k-najbližih susjeda

Algoritam nadgledanog strojnog učenja k-najbližih susjeda [7] se može koristiti za probleme klasifikacije i regresije. Algoritam radi na principu računanja udaljenosti između opažanja koje želimo klasificirati i određenog broja označenih opažanja iz seta obuke. Kod problema klasifikacije udaljenost između navedenih opažanja računa se preko euklidske udaljenosti prema formuli 4.1 ili manhattanske udaljenosti prema formuli 4.2. Kod formula 4.1 i 4.2 x predstavlja neoznačeno opažanje, a y predstavlja označeno opažanje.

$$d(x,y) = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2} \tag{4.1}$$

$$d(x,y) = \sum_{i=1}^{m}|x_i - y_i| \tag{4.2}$$

Nakon izračuna udaljenosti određeni broj označenih susjeda se sortira po udaljenosti. Kod klasifikacije klasa za neoznačeno opažanje se odredi prebrojavanjem klasa svih promatranih označenih susjeda, najbrojnija klasa predstavlja klasu neoznačenog opažanja. Kod regresije računa se prosjek promatranih označenih opažanja te neoznačeno opažanje poprima tu vrijednost. Hiperparametar k određuje koliki broj označenih opažanja utječe na računanje neoznačenog opažanja. Generalno pravilo za određivanje hiperparametra k je $k = \sqrt{broj_opažanja}$.

U programskom okruženju R klasifikacijski model k-najbližih susjeda se računa kodom:

knn(setObuke, setProbe, ispravnaCiljanaVrijednost, k=i)

gdje *i* predstavlja broj promatranih označenih opažanja.

4.2 Klasifikacijski model višečlana logička regresija

Klasifikacijski model višečlane logističke regresije [8][9] se koristi kod kategorički nezavisni varijabli s više od dvije kategorije. Višečlana logistička regresija je proširenje logističke regresije koja se koristi za klasifikaciju kategorija s binarnim ishodom. Model višečlane logističke regresije računa se formulom 4.3, gdje $P(Y = i)$ predstavlja vjerojatnost da zavisna varijabla spada u kategoriju *j*, $P(Y = k)$ predstavlja referentnu kategoriju za usporedbu, X_1, X_2, X_p predstavljaju prediktore, β_{0i} predstavlja presretni pojam za kategoriju *j*, β_1, β_2, β_p predstavljaju koeficijente povezane s nezavisnim varijablama. Presretni pojam i koeficijente povezane s nezavisnim varijablama računamo preko log-vjerojatnosti funkcije, formula 4.4, te nakon dobivene *LL* vrijednosti uvrstimo u formulu tipa 4.5. Gdje $I(C_j = i)$ predstavlja funkciju koja iznosi 1 ako opažanje *j* spada u

kategoriju *i* i nula inače, $P(C_j = i)$ predstavlja vjerojatnost da opažanje *j* spada u kategoriju *i*, $P(C_j = J)$ predstavlja vjerojatnost da opažanje *j* spada u referentnu kategoriju *J*.

$$\log(\frac{P(Y=i)}{P(Y=k)}) = \beta_{oi} + \beta_1 X_1 + \ldots + \beta_p X_p \tag{4.3}$$

$$LL(\beta_{oi}, \beta_1, \beta_2) = \sum_{j=1}^{N} \sum_{i=1}^{k} I(C_j = i) * [\log \frac{P(C_j=i)}{P(C_j=J)}] \tag{4.4}$$

$$\frac{dLL}{d\beta_1} = 0 \tag{4.5}$$

U programskom okruženju R klasifikacijski model višečlane logističke regresija se razvija kodom:

multinom_reg() |> fit(ciljaVarijabla ~ prediktor1 + prediktor2 + prediktor3 + prediktor4, data = setObuke

4.3 Klasifikacijski model stroj potpornih vektora

Algoritam stroja potpornih vektora [8] spada pod nadgledano strojno učenje i može se koristiti za probleme klasifikacije i regresije. Stroj potpornih vektora radi klasifikaciju na principu izrade hiperravnine (en. hyperplane) koja je točka u 1d prostoru, pravac u 2d prostoru ili ravnina u 3d prostoru na način da sva opažanja klase jedan su odvojena hiperravninom od opažanja klase dva. Hiperravnina je na način da maksimizira udaljenost do opažanja svake klase i računa se za 2d prostor preko formule 4.6, gdje β_0 predstavlja poziciju hiperravnine u prostoru, β_1 i β_2 orijentaciju hiperravnine i *X1 i X2* su vrijednosti opažanja. Maksimizirana udaljenost hiperravnina se naziva margina. Opažanja koja padaju direktno na marginu se zovu potporni vektori. Za određivanje klase neoznačenih opažanja koristi se formula 4.7, gdje je *w* vektor koji sadrži koeficijente β_1 i β_2 (*w = [*β_1 i β_2*]*), *x* vektor neoznačenog opažanja (*x = [X1, X2]*) i *b* predstavlja koeficijent β_0. Formulom 4.6 se utvrđuje da li je neoznačeno opažanje veće ili jednako jedan, ili manje ili jednako minus jedan. Ovisno o vrijednosti izračuna neoznačeno opažanje svrsta se u klasu s jedne ili druge strane hiperravnine.

$$\beta_0 + \beta_1 + X1 + \beta_2 X2 = 0 \tag{4.6}$$

$$w^T x - b = 0 \tag{4.7}$$

U slučaju da podatke nije moguće odvojiti hiperravninom moguće je primijeniti nelinearnu transformaciju na sva opažanja i nakon primjene nelinearne transformacije pronaći hiperravninu za novi višedimenzionalan prostor. Naziv navedene metode je trik jezgre (en. kernel trick).

Kod klasifikacijskog modela s više od dvije klase, stroj potpornih vektora razvije $\binom{K}{2}$ strojeva potpornih vektora, od kojih svaki uspoređuje par klasa. Klasificiramo neoznačeno promatranje za svaki od $\binom{K}{2}$ stroja potpornih vektora i pridodijelimo neoznačeno promatranje najčešće izračunatoj klasi. Naziv navedene klasifikacije je jedan-protiv-jednog (en. one-versus-one).

U programskom okruženju R, klasifikacijski model stroja potpornih vektora računa se kodom:

train(ciljanaVarijabla~., data=setObuke, method="svmPoly", metric=metrika, trControl=kontrola)

gdje *svmPoly* predstavlja polinomsku jezgru, *metrika* predstavlja korištenu evaluacijsku metriku koja je postavljena na točnost i *kontrola* predstavlja kontrolne parametre koji su postavljeni na 10-struku unakrsnu provjeru.

4.4 Indikatori uspješnosti klasifikacijskih modela

Jedna od najčešće korištenih indikatora uspješnosti klasifikacijskog modela je točnost (en. accuracy). Točnost mjeri omjer točno klasificiranih opažanja, prikazano formulom 4.8, gdje su *TP* točno klasificirana pozitivna opažanja, *TN* točno klasificirana negativna opažanja, *FP* krivo klasificirana pozitivna opažanja i *FN* krivo klasificirana negativna opažanja.

$$t = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{4.8}$$

ROC krivulja prikazuje omjer između istinske pozitivne stope (en. true positive rate) i lažne pozitivne stop (en. false positive rate). ROC krivulja je grafički prikaz sposobnosti klasifikacijskog modela u razlikovanju klasa. Os x ROC krivulje prikazuje specifičnost, dok os y prikazuje osjetljivost. Izračun osi ROC krivulje prikazan je formulama 4.9 i 4.10.

$$x = \frac{FP}{(FP+TN)} \tag{4.9}$$

$$y = \frac{TP}{(TP+FN)} \tag{4.10}$$

Vrijednost AUC, prostora ispod ROC krivulje, numerički prikazuje uspješnost klasifikacijskog modela. Vrijednost AUC dobijemo izračunom površine ispod ROC krivulje. AUC vrijednost može biti broj između nula i jedan, gdje jedan predstavlja savršeni klasifikacijski model. U slučaju da je vrijednost AUC jednaka 0.5 klasifikacijski model ne daje bolje rezultate od nasumičnog pogađanja.

Efektivan stupanj slobode (en. effective degree of freedom) prikazuje kompleksnost klasifikacijskog modela. Generalno se računa prema formuli 4.11, gdje je P broj koeficijenata klasifikacijskog modela i N broj opažanja.

$$DF = N - P \tag{4.11}$$

Odstupanje (en. deviance) mjeri koliko dobro klasifikacijski model pristaje podacima. Manja vrijednost odstupanja predstavlja bolje pristajanje. Odstupanje se računa kao razlika vjerojatnosti pristajanog modela (VPM) i zasićenog modela (VZM), prikazano formulom 4.12.

$$D = -2\log VPM + 2\log VZM \tag{4.12}$$

Akaike informacijski kriterij (Akaike information criterion) predstavlja pristajanje klasifikacijskog modela u omjeru na kompleksnost modela. Manja vrijednost Akaike informacijskog kriterija predstavlja bolji model. Način računanja navedenog indikatora prikazan je formulom 4.13, gdje k predstavlja broj procijenjenih parametara klasifikacijskog modela i L predstavlja maksimiziranu vrijednost funkcije vjerojatnosti za model.

$$AIC = 2k - 2\ln L \tag{4.13}$$

Interval pouzdanosti (en. Confidence interval) je raspon vrijednosti varijable s određenim postotkom pouzdanosti, a računa se prema formuli 4.14, gdje *SM* predstavlja aritmetički sredinu uzoraka i *MoE* predstavlja marginu greške. *MoE* računamo preko formule (4.15), gdje Z je kritična vrijednost odabrane distribucije željene razine pouzdanosti, SE predstavlja standardnu grešku i n predstavlja broj uzoraka.

$$CI = S \pm MoE \tag{4.14}$$

$$MoE = Z * \frac{SE}{\sqrt{n}} \tag{4.15}$$

Stopa bez informacija (en. No-information rate) predstavlja točnost klasifikacijskog modela ako bi model uvijek predviđao najzastupljeniju klasu i računa se formulom 4.16, gdje I predstavlja opažanja najzastupljenije klase i T predstavlja sveukupni broj opažanja.

$$NIR = \frac{I}{T} \tag{4.16}$$

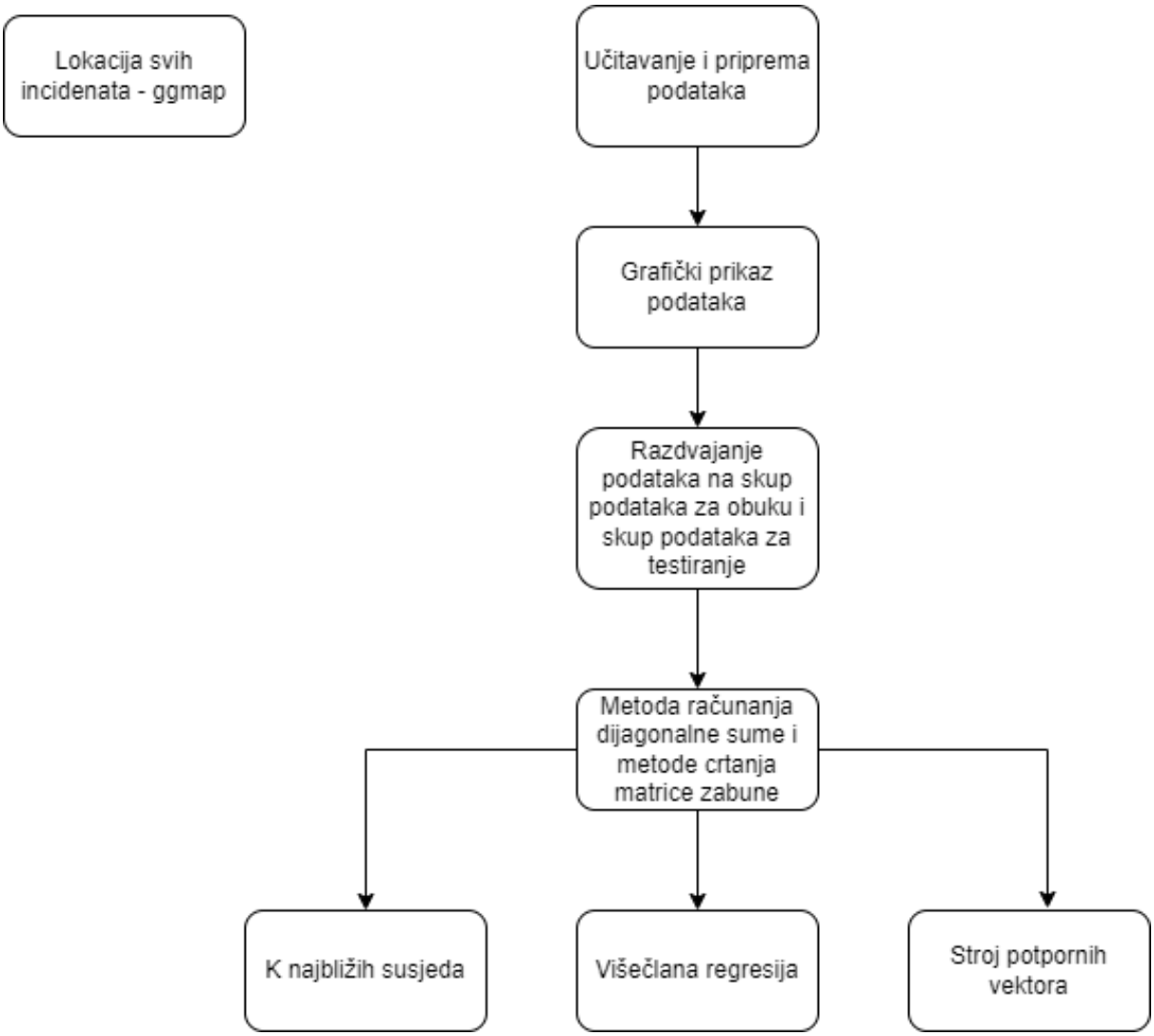
Matrica zabune (en. confusion matrix) grafički prikazuje rezultate klasifikacije svih opažanja seta obuke. Prikazuje sva točno klasificirana opažanja na glavnoj dijagonali, dok sve druge vrijednosti predstavljaju krivo klasificirana opažanja.

Cohenova kappa (en. Cohen's kappa) se koristi za mjerenje stupnja sporazuma između dva promatrača. Vrijednost kappe iznad nule predstavlja veći sporazum od sporazuma kojega bi dobili nasumično. Računa se prema formuli 4.17.

$$k = \frac{2*(TP*TN-FN*FP)}{(TP+FP)*(FP+TN)+(TP+FN)*(FN+TN)} \tag{4.17}$$

5. REZULTATI ISTRAŽIVANJA

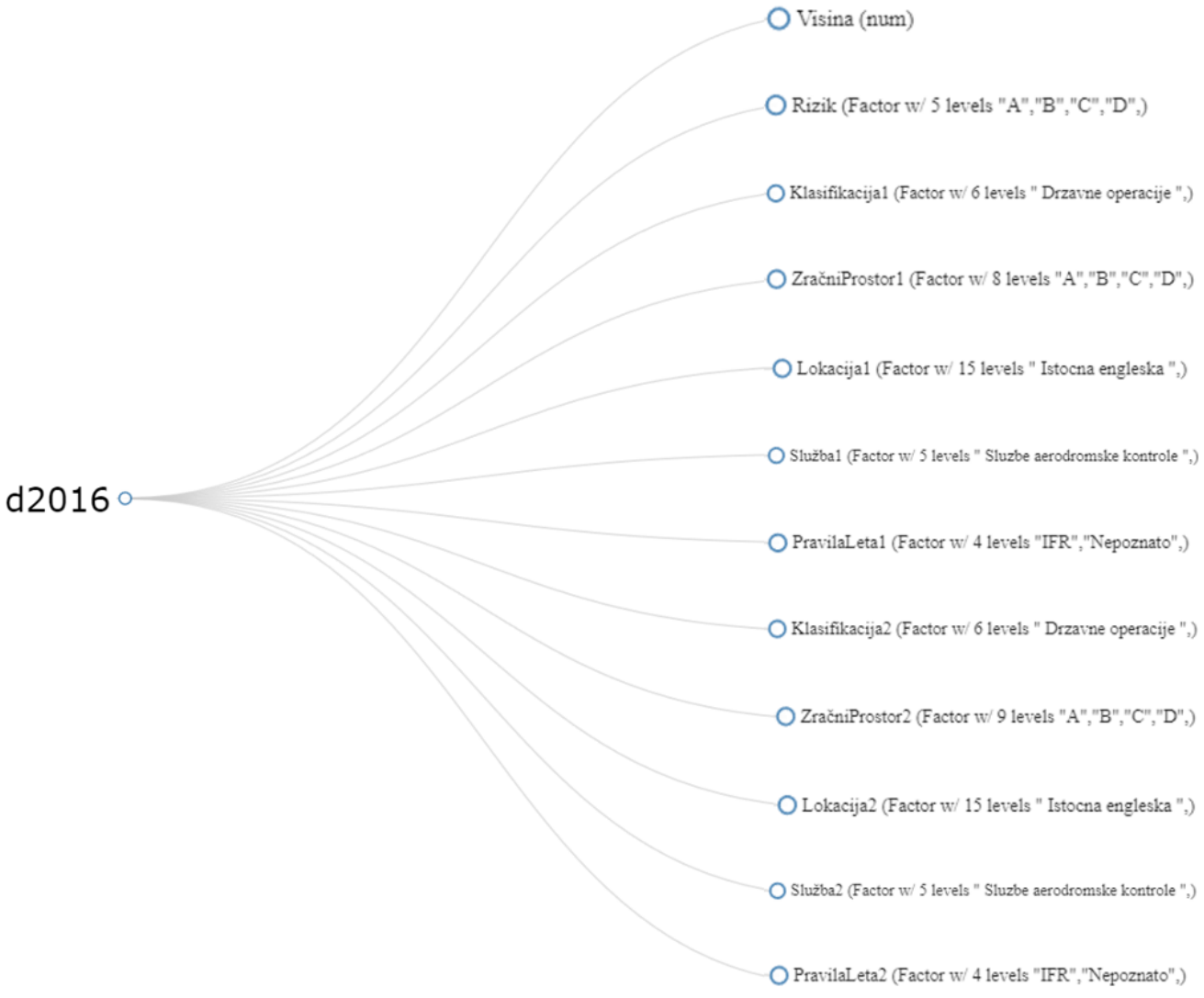
Istraživanja su izvedena u programskom okruženju R. Programski kod za iscrtavanje svih lokacija incidenata na Google mapama koristi ggmap knjižnicu i odvojen je od ostatka koda. Programski kod cijelog istraživanja te njemu potrebne datoteke mogu se naći u dodatku [B]. Struktura programskog koda istraživanja prikazana je pomoću dijagrama toka, slika 5.1.



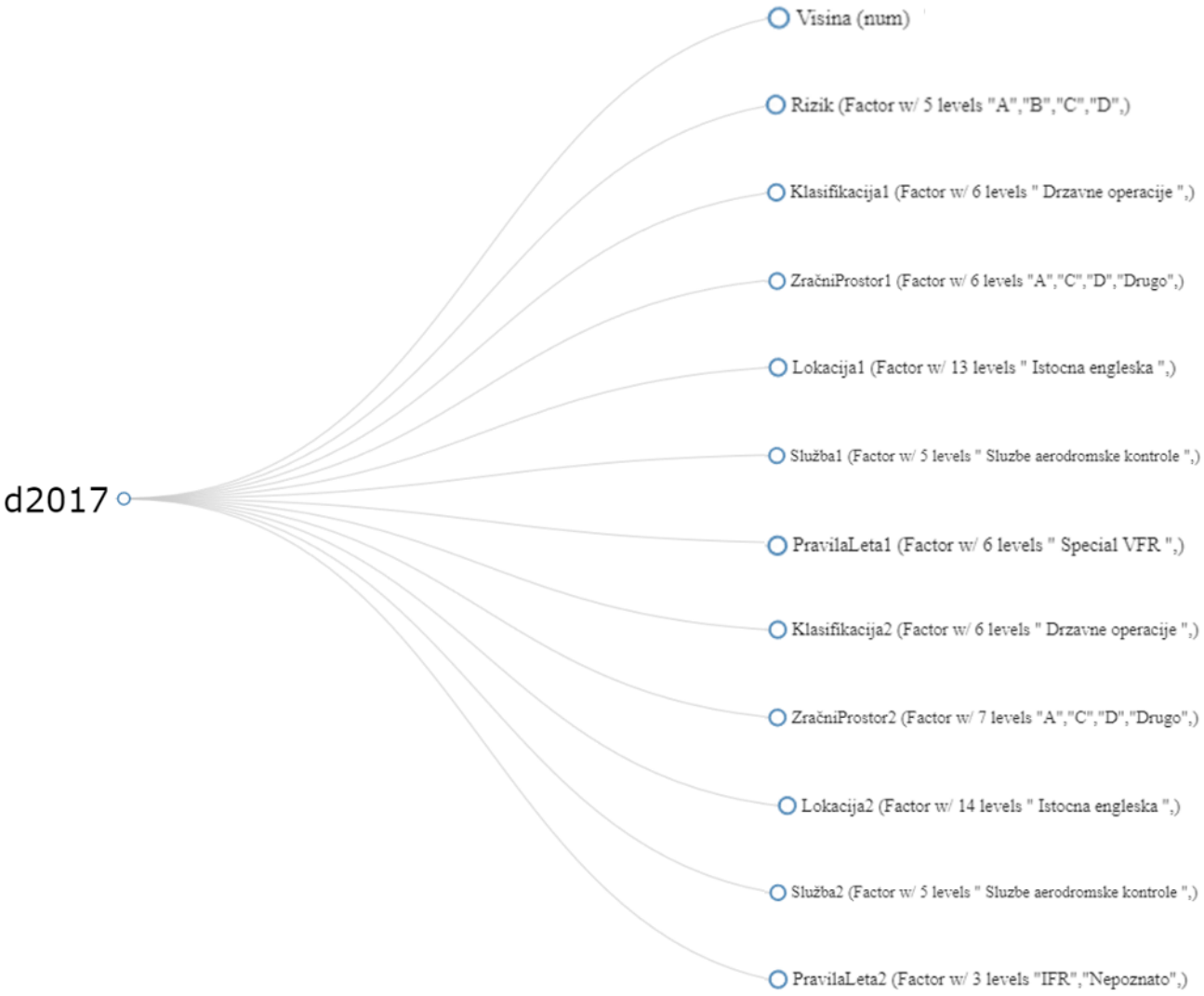
Slika 5.1 Dijagram toka programskog rješenja

5.1 Struktura podataka

Struktura podataka nakon navedene pripreme, sadrži 3094 opažanja za podatke do kraja 2016. godine i 336 opažanja za podatke 2017. godine. Struktura se sastoji od jedanaest prediktora i jedne varijable ishoda. Prediktor visina je jedini kontinuirani prediktor dok su svi ostali diskretni. Kod incidenata sa situacijama u kojima je narušena sigurnost letjelica sudjeluju dvije letjelice, tako da preostalih deset prediktora dolaze u paru: Klasifikacija1 Klasifikacija 2, Lokacija1 Lokacija2, Služba 1 Služba 2, ZračniProstor1 Zračni Prostor 2, PravilaLeta1 PravilaLeta2, slika 5.2 i 5.3.



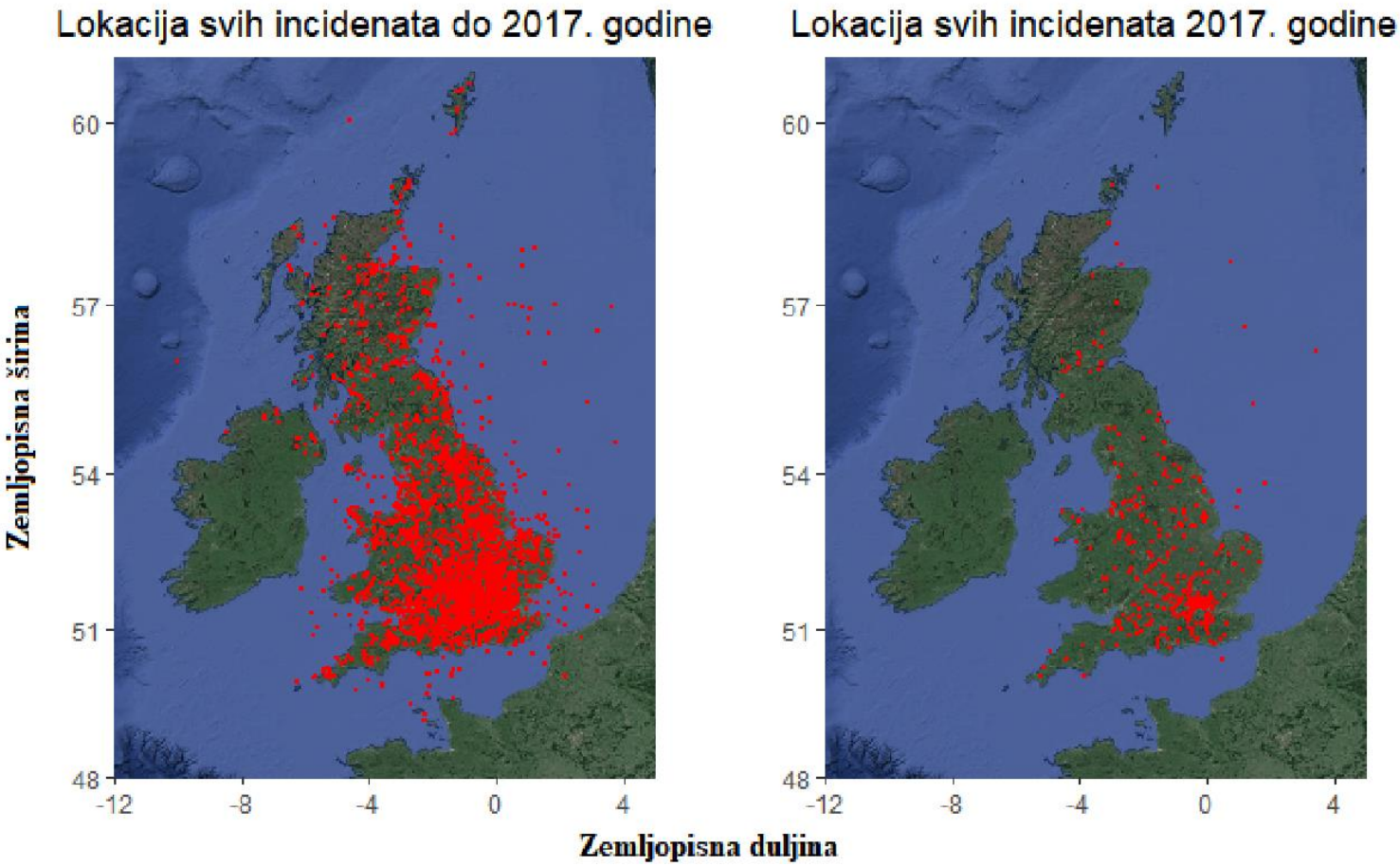
Slika 5.2 Struktura podataka do kraja 2016. godine



Slika 5.3 Struktura podataka za 2017. godinu

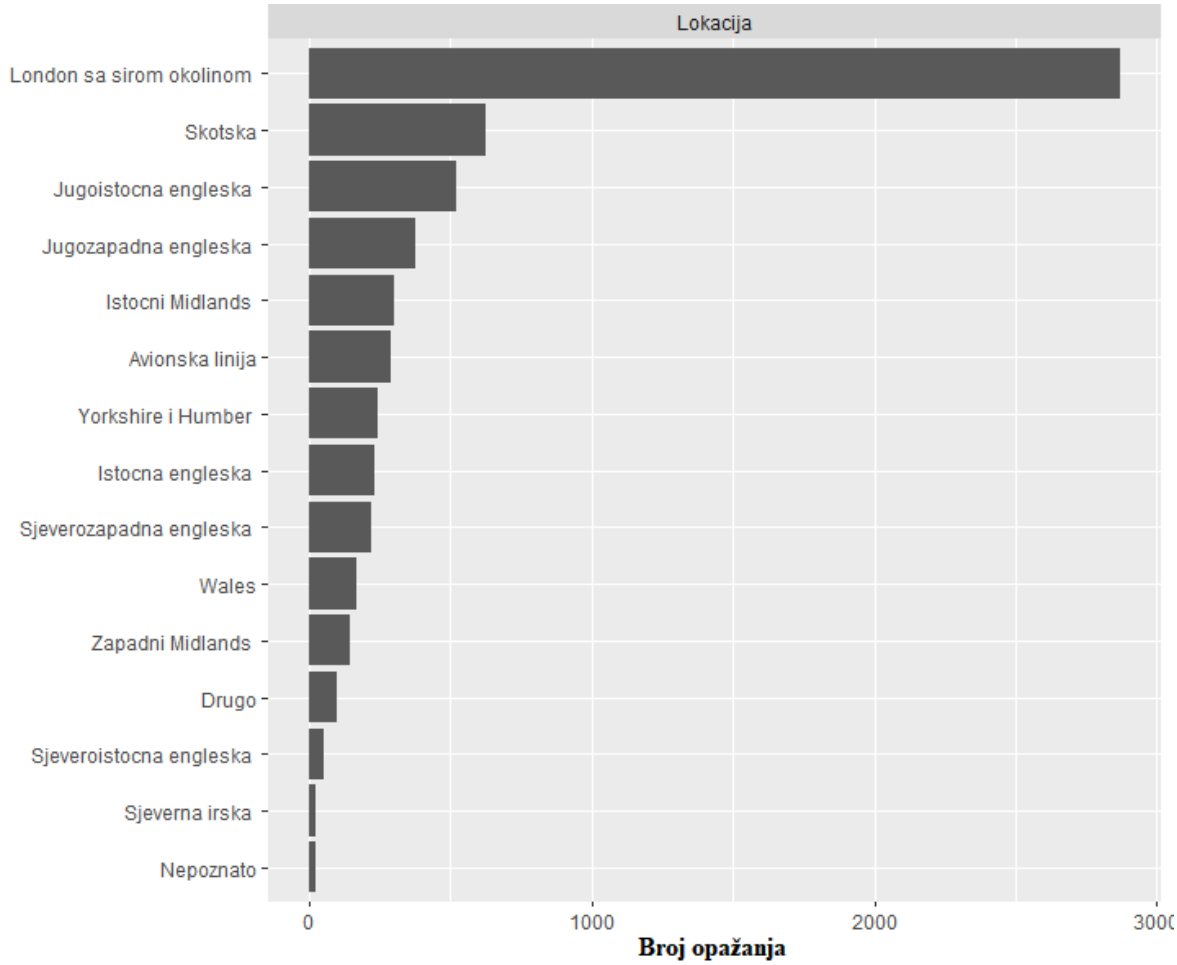
5.2 Rezultati opisne statističke analize podataka

Rezultat programskog koda za iscrtavanje svih incidenata prikazan je na slici 5.4. Rezultat je dobiven iscrtavanjem svih parova zemljopisnih širina i zemljopisnih dužina za prijavljene incidente. Iz rezultata vidi se znatna razlika između podataka do kraja 2016. godine i podatak 2017 godine. Uz to što sadrže vidljivo više opažanja, lokacije svih prijavljenih incidenata do kraja 2016. godine su mnogo raspršenije po teritoriju Ujedinjenog kraljevstva za razliku od lokacija svih incidenata 2017. godine, dok su London i šira okolica daleko najzastupljeniji kod prijave incidenata u oba vremenska perioda

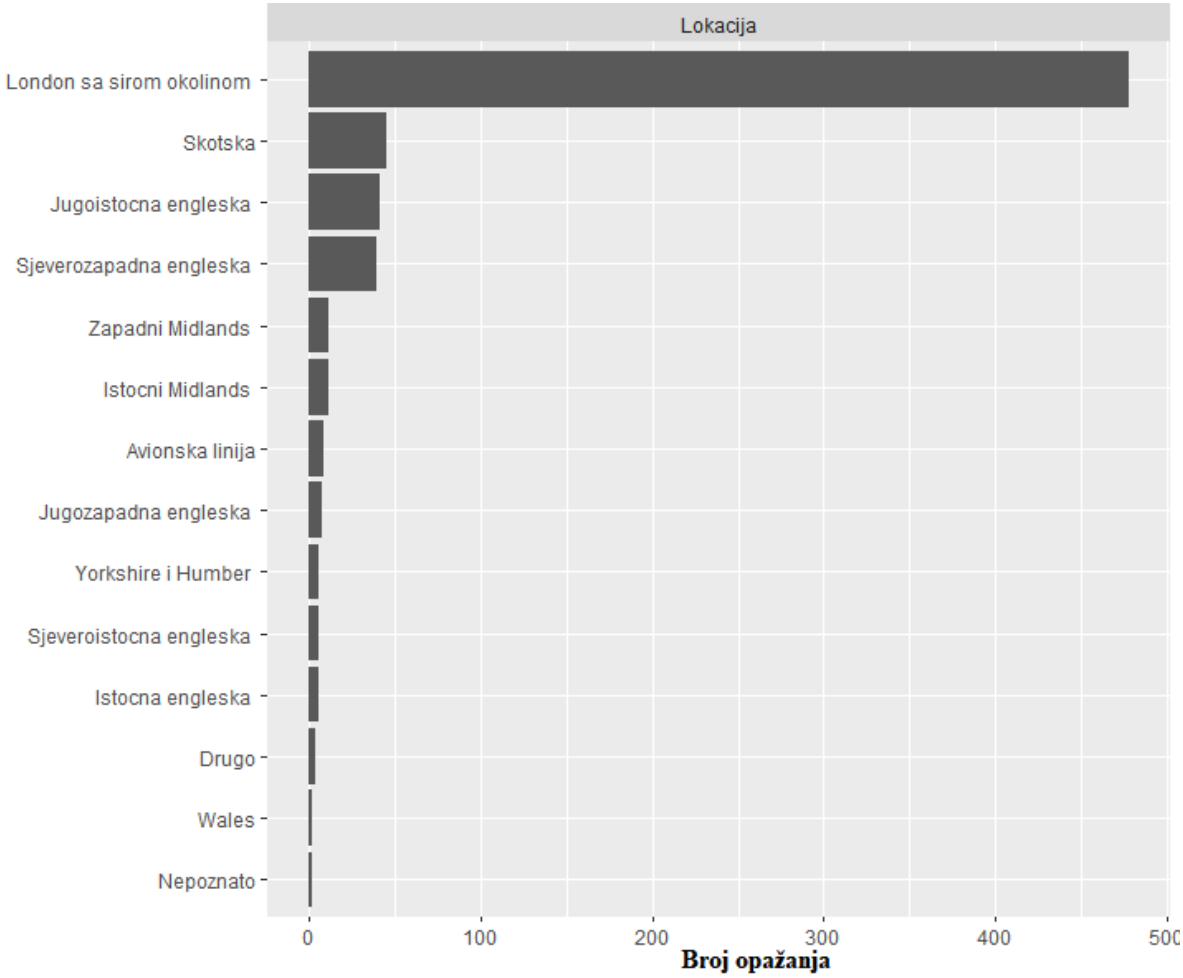


Slika 5.4 Lokacija svih incidenata - ggmap knjižnica

Razlika između podataka za lokaciju svih incidenata do kraja 2016. godine i podataka 2017. godine lakše je zamijetiva iz rezultata prikaza lokacija svih incidenata knjižnice DataExplorer, slika 5.5 i 5.6. Vidljivo je da London sa širom okolinom dominira kao lokacija s najvećem brojem incidenata za oba vremenska perioda, ali za period 2017. godine sve ostale lokacije unutar Ujedinjenog Kraljevstva sadrže znatno manji omjer prijavljenih incidenata što ukazuje na znatno bolju sigurnost letjelica na lokacijama izvan Londona i šire okoline. Budući da u periodu do kraja 2016. godine i 2017. godine London sadrži veliku većinu prijavljenih incidenata možemo zaključiti da kroz London i prometuje daleko najveći broj letjelica, što ujedno i uzrokuje toliki broj incidenata.

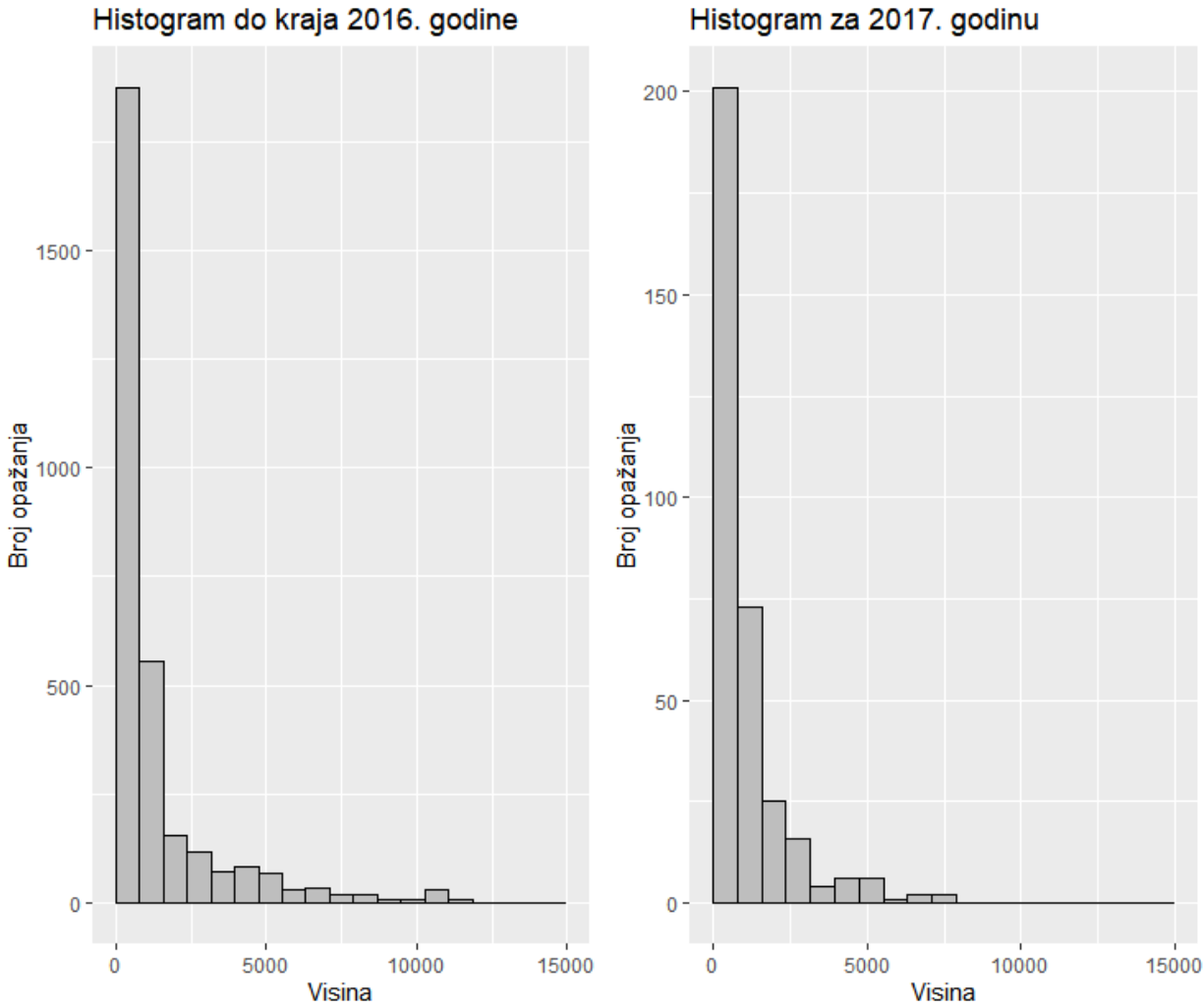


Slika 5.5 Lokacija svih incidenata do kraja 2016. godine



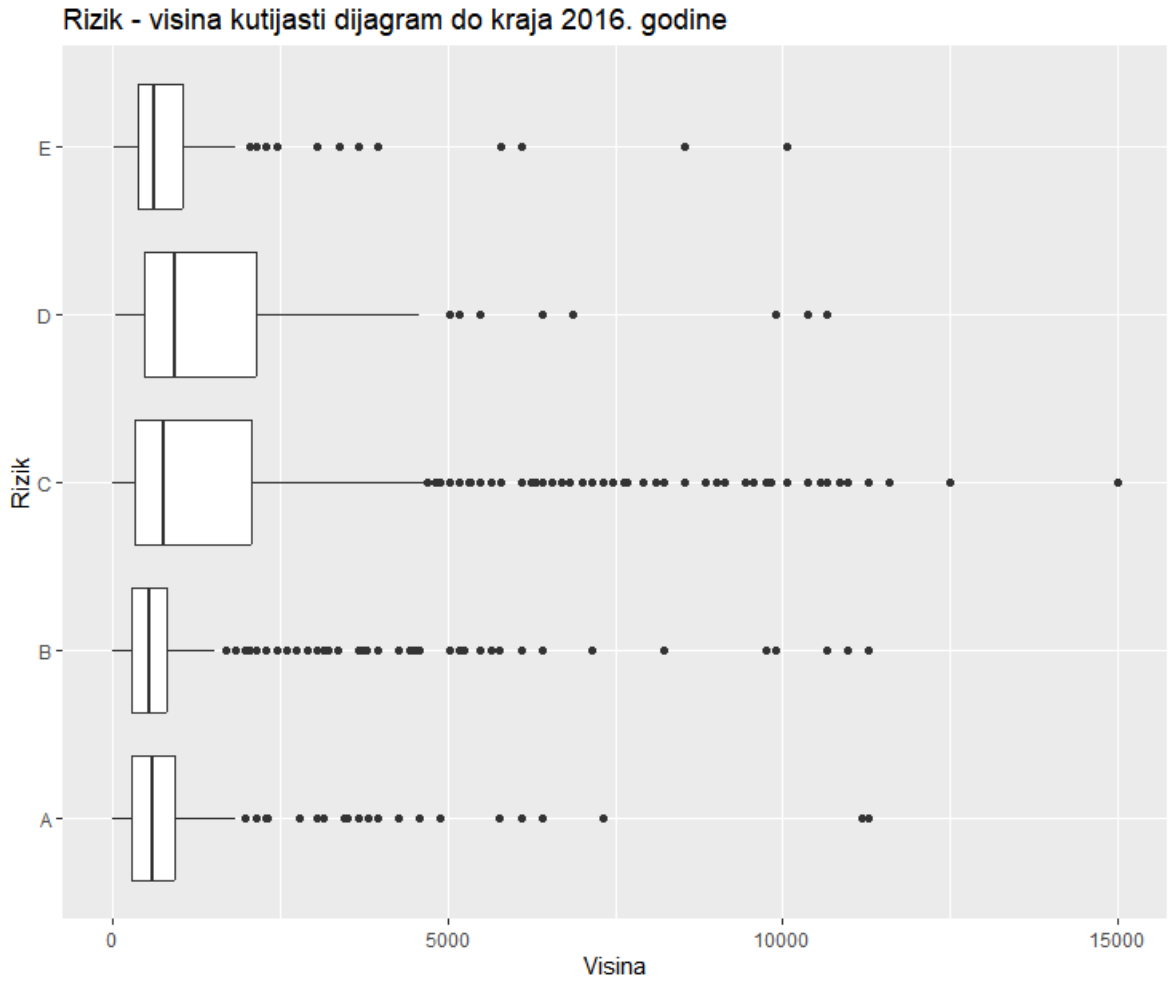
Slika 5.6 Lokacija svih incidenata za 2017. godinu

Izradom histograma visine, slika 5.7, moguće je uočiti da visina leta letjelica do kraja 2016. godine sadrži mnogo veći broj incidenata s visinom oko i iznad 5000 metara. Uz to, iz visina prijavljenih incidenata za 2017. godinu možemo vidjeti da ne postoji niti jedan prijavljeni incident iznad 25 000 stopa, od prilike 7.62 kilometra, što ukazuje da je jedan od ciljeva novih regulatornih mjera uvedenih 2017. godine smanjenje visine leta bespilotnih letjelica.

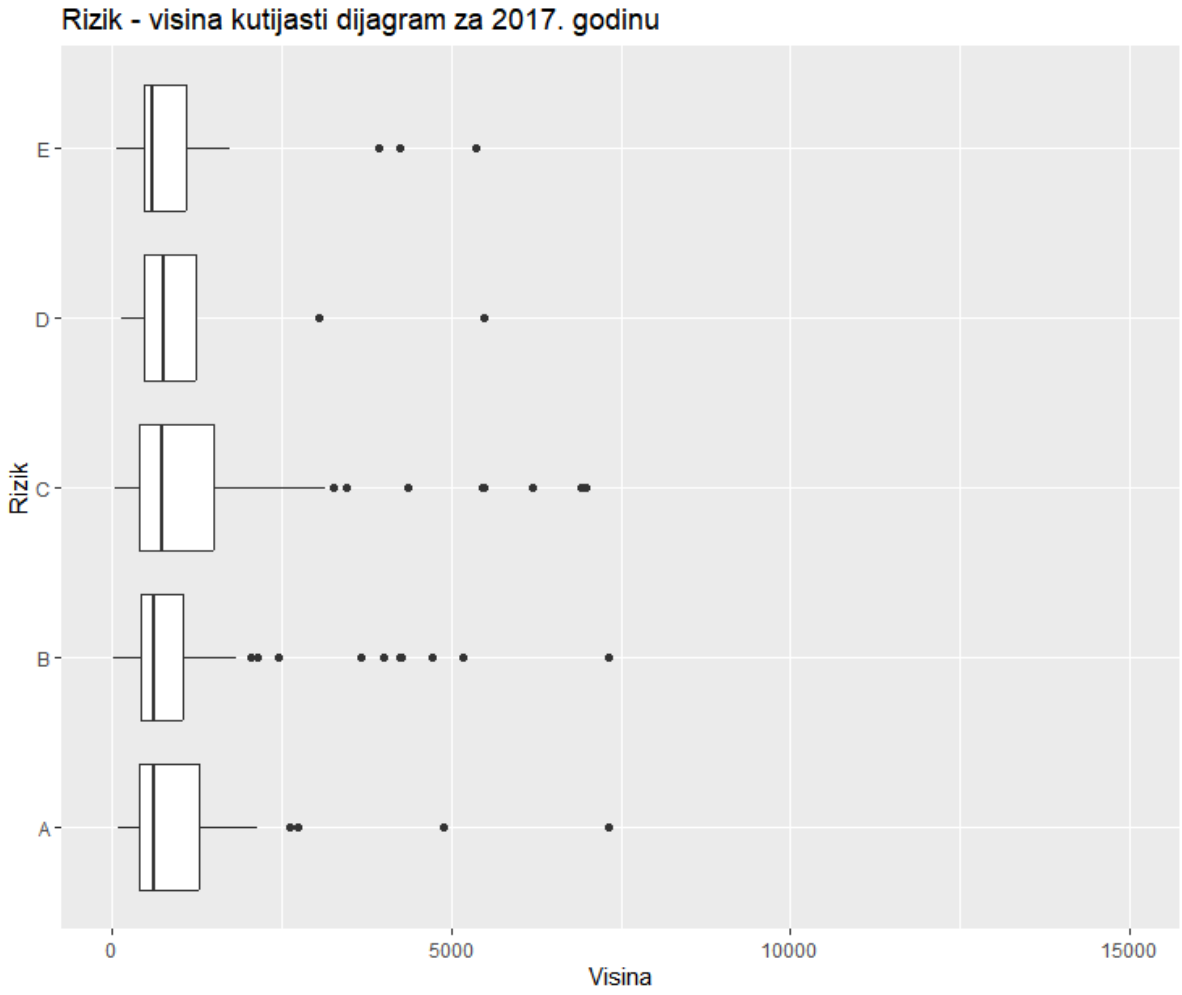


Slika 5.7 Histogrami visine

Isto opažanje možemo donijeti proučavanjem rizik-visina kutijastih dijagrama, slika 5.8 i 5.9. Iz kutijastih dijagrama je vidljivo da na podacima do kraja 2016. godine prosječna visina prijavljenih incidenata je mnogo veća od 2017. godine, te da sadrži mnogi broj prijavljenih incidenata na visinama iznad 25 000 stopa za razliku od podatak do kraja 2016.godine. Uz to, kutijasti dijagrami ukazuju da za razliku od prijavljenih incidenata do kraja 2016. godine gdje kod kategorije rizika C i kategorije rizika D prijavljeni incidenti se nalaze na prosječno mnogo većim visinama, visina prijavljenih incidenata 2017. godine je značajno više uravnotežena. Opažanjem bi se moglo zaključiti da porastom korištenja malih bespilotnih letjelica proporcionalno raste i prijava incidenata na nižim nadmorskim visinama.

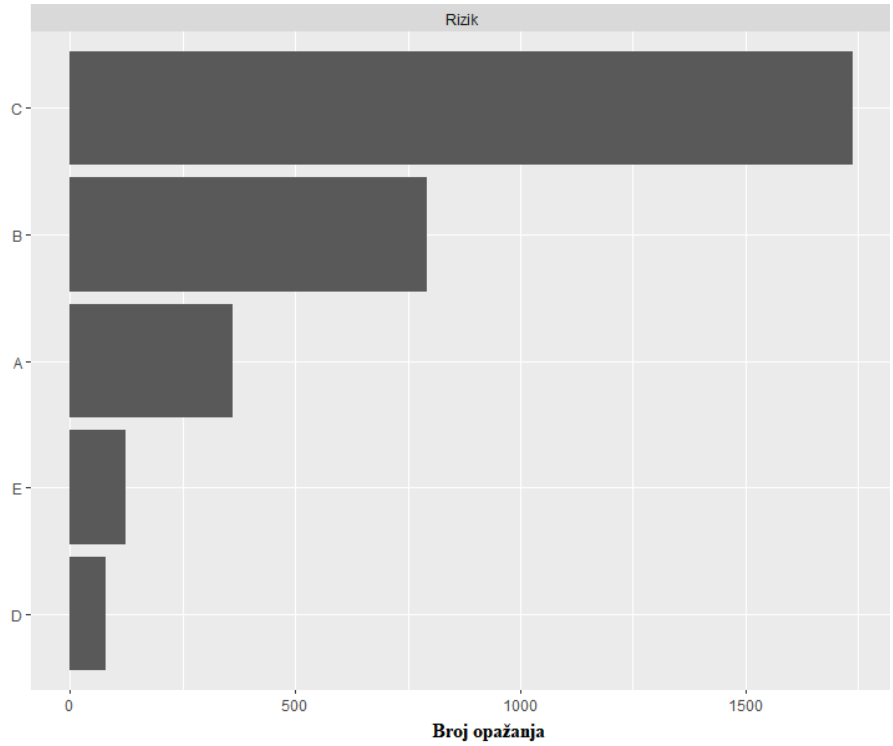


Slika 5.8 Rizik – visina kutijasti dijagram za podatke do kraja 2016. godine

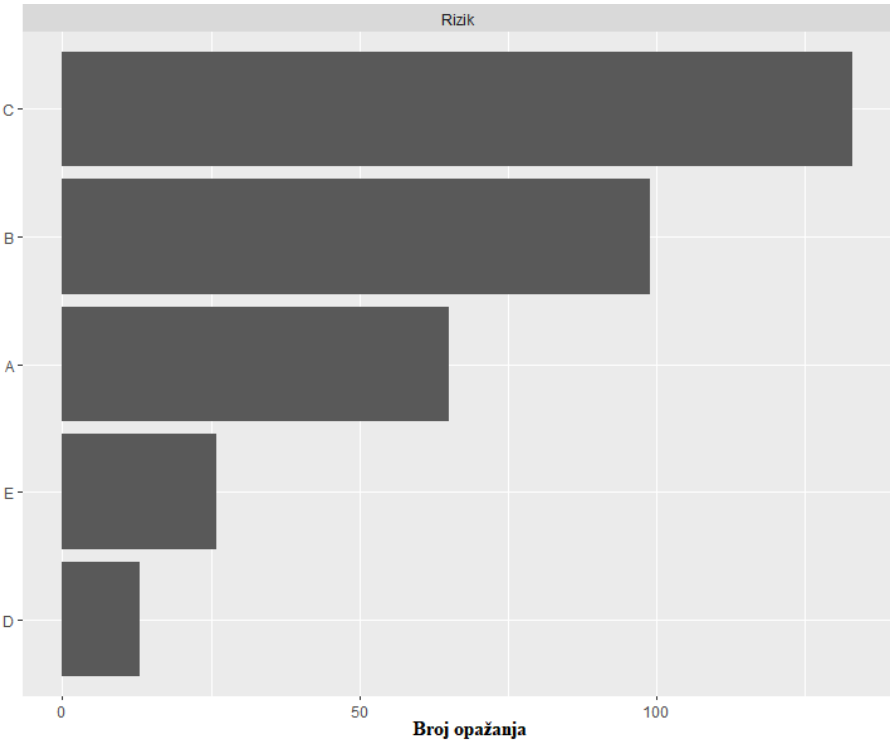


Slika 5.9 Rizik – visina kutijasti dijagram za podatke 2017. godine

Kako bi dublje razumjeli raspodjelu visine po kategorijama rizika promatramo ciljanu varijablu kategorija rizika, slika 5.10 i 5.11. Daleko najzastupljenija kategorija rizika za podatke do kraja 2016. godine je kategorija rizika C, koja predstavlja da nema opasnosti od sudara. Za razliku, kod 2017. godine u omjeru na kategoriju rizika C sve su ostale kategorije rizika učestalije. Kategorija rizika A predstavlja opasnost od sudara, dok kategorija rizika B predstavlja da sigurnost nije zajamčena. Kategorije rizika A i B koje su ujedno i rizične kategorije su znatno učestalije od perioda do kraja 2016. godine. Budući da se omjer rizičnih kategorija povećao 2017. godine moguće je zaključiti da je sigurnost letjelica manja za navedeni period. Kategorija rizika E, koja je jedinstvena za Ujedinjeno Kraljevstvo, znači da su za prijavljeni incident važili normalni sigurnosni standardi.

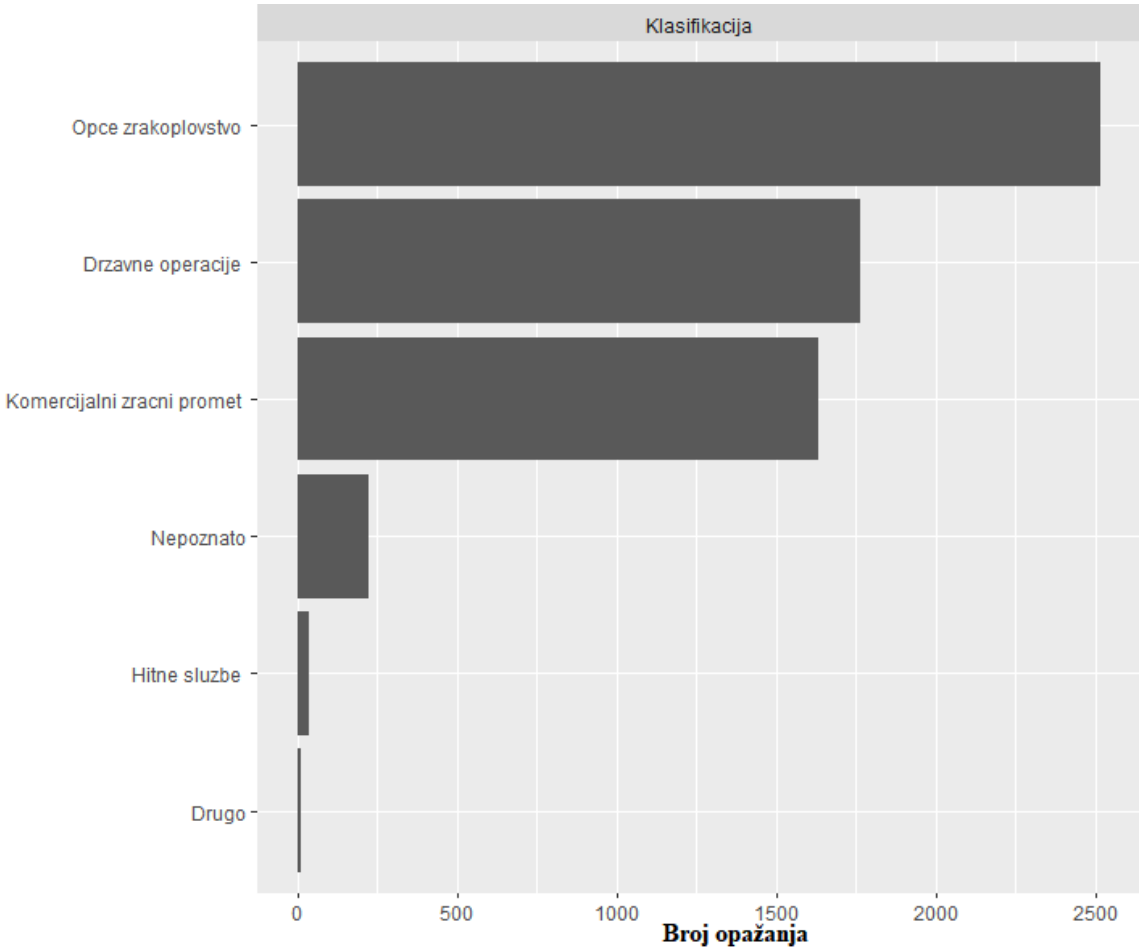


Slika 5.10 Kategorija rizika za podatke do kraja 2016. godine

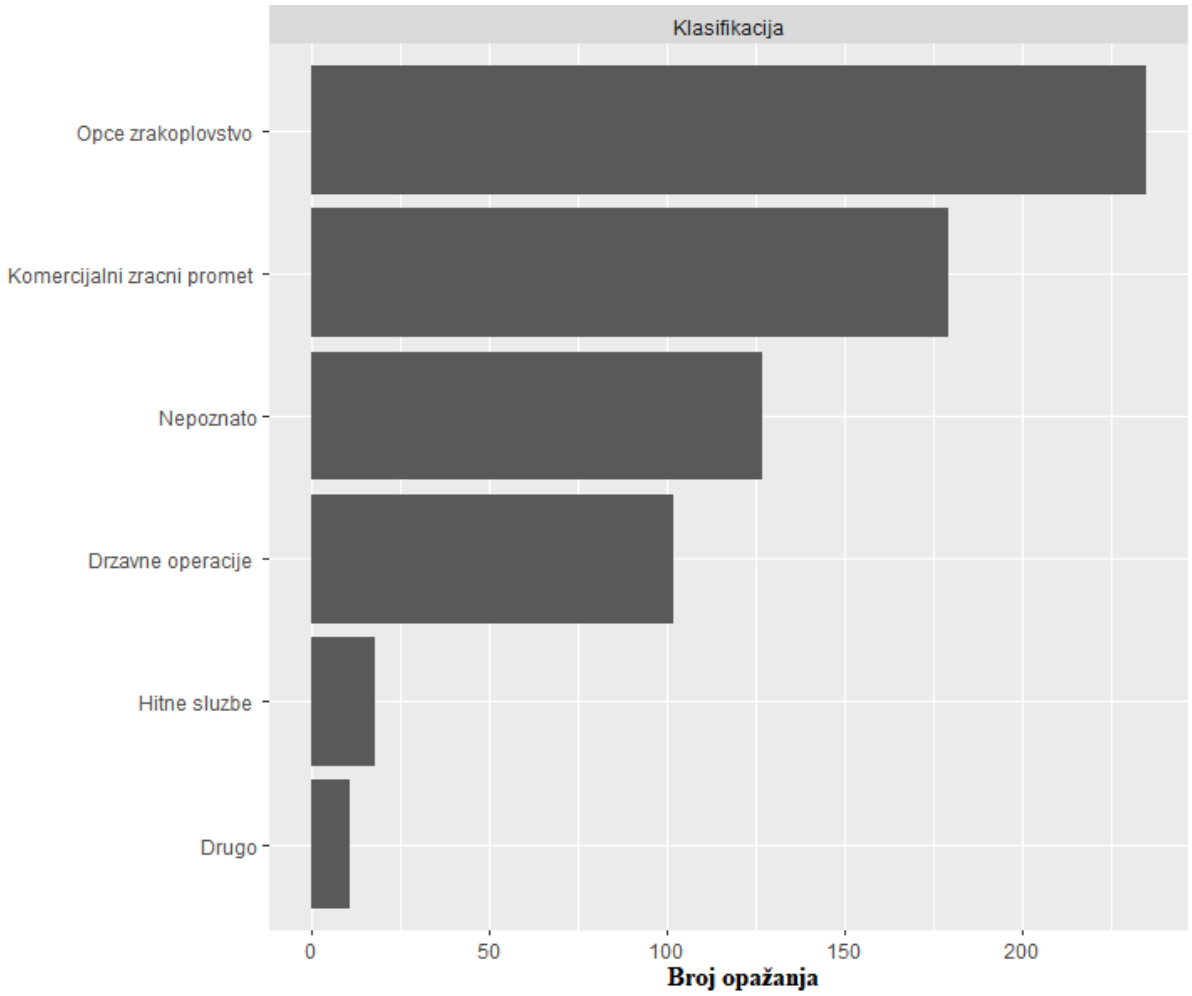


Slika 5.11 Kategorija rizika za podatke 2017. godine

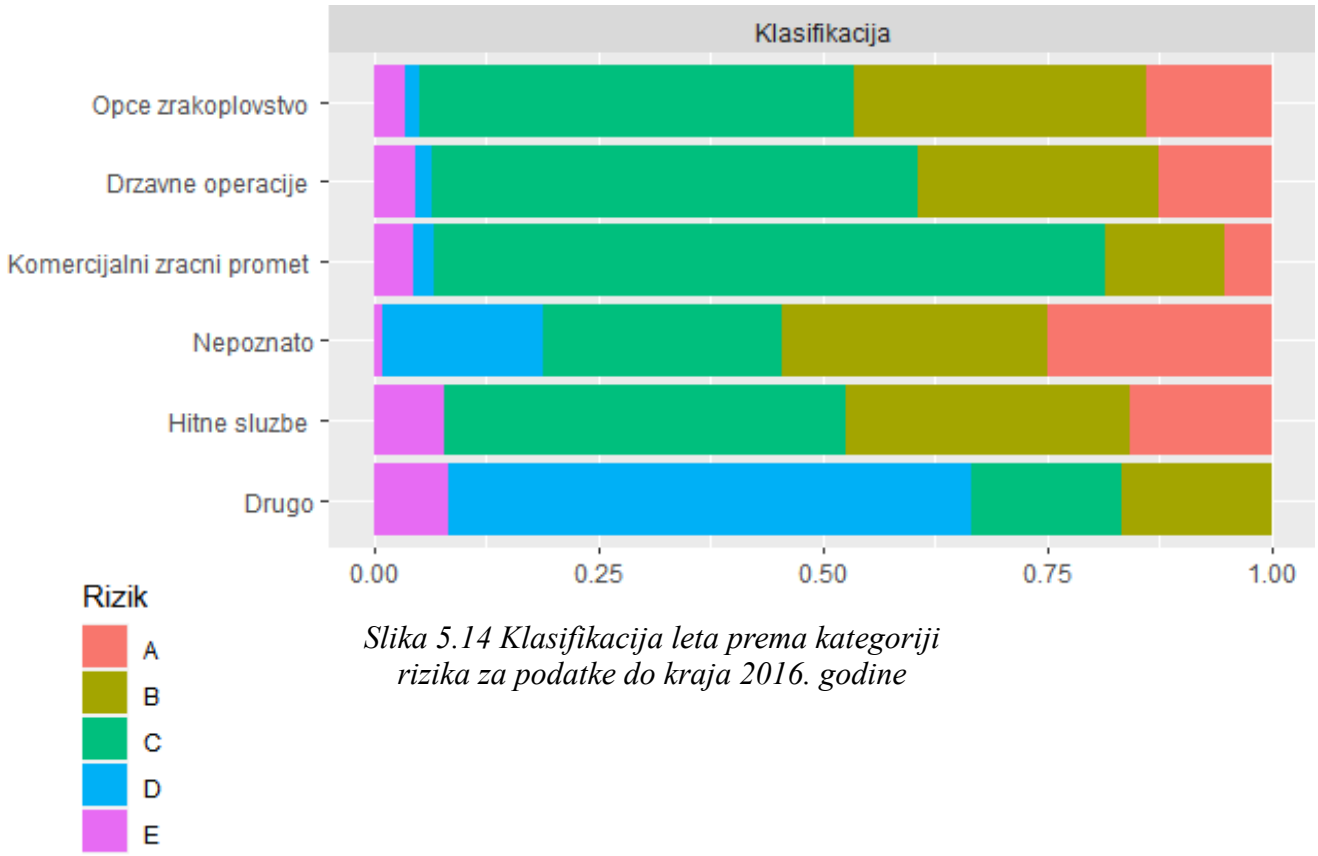
komercijalni zračni promet i državne operacije. Opće zrakoplovstvo se odnosi na svu civilnu uporabu letjelica, sve dok nisu i komercijalne svrhe. Komercijalni zračni promet su dakle svi komercijalni letovi, dok u državne operacije spadaju svi letovi vojnih i policijskih službi. Letovi služba spašavanja i medicinskih intervencija spadaju pod hitne službe. Raspodjela klasifikacije letova za podatke do kraja 2016. godine prikazana je na slici 5.12, te za 2017. godinu na slici 5.13. Iz navedenih slika vide se omjeri svake klasifikacije leta iz kojih se može vidjeti da je opće zrakoplovstvo najčešće prijavljeni incident. Nadalje, vidi se rast prijava incidenata s klasifikacijom leta komercijalni zračni promet is prvog vremenskog perioda u drugi, što ukazuje na manju sigurnost komercijalnih letova. Daljnjom obradom prikazana je klasifikacija leta prema kategoriji rizika, slika 5.14 i 5.15. Iz navedenih slika može se uočiti da za period do kraja 2016. godine rizične kategorije, kategorije rizika A i B, su uočljivo manje nego za 2017. godinu, osim za klasifikaciju leta hitne službe. Najistaknutiji porast u kategoriji rizika je kod klasifikacije leta komercijalnog zračnog prometa koji je do kraja 2016. godine imao postotak rizičnih kategorija manji od 25%, koji se 2017. godine popeo na 50%.



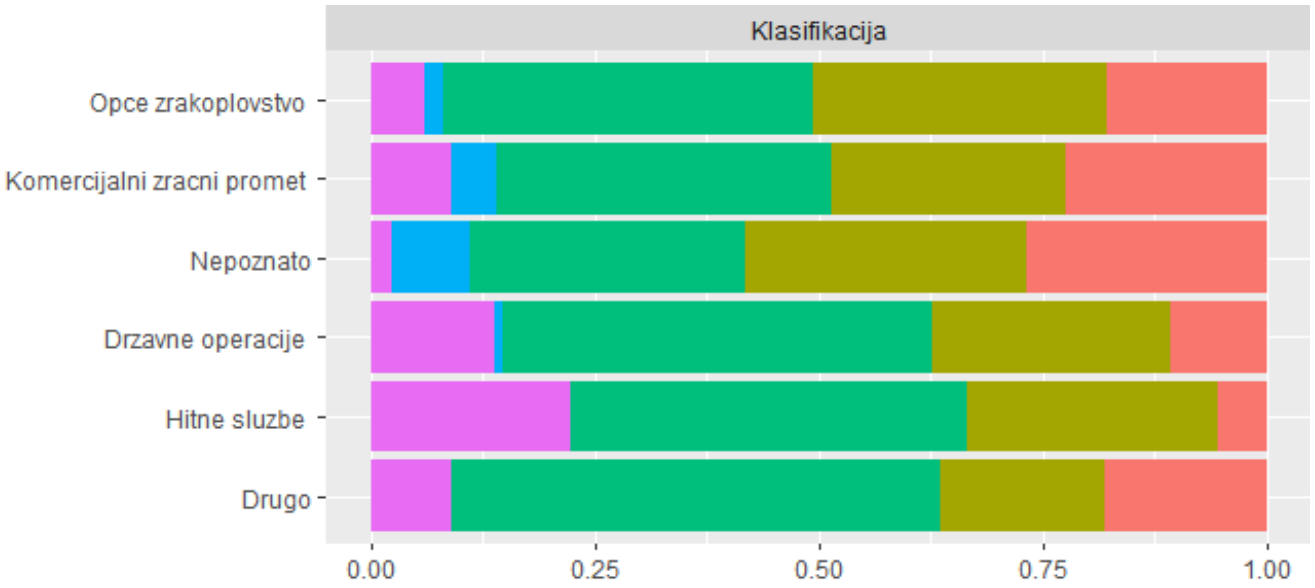
Slika 5.12 Klasifikacija leta za podatke do kraja 2016. godine



Slika 5.13 Klasifikacija leta za podatke 2017. godine

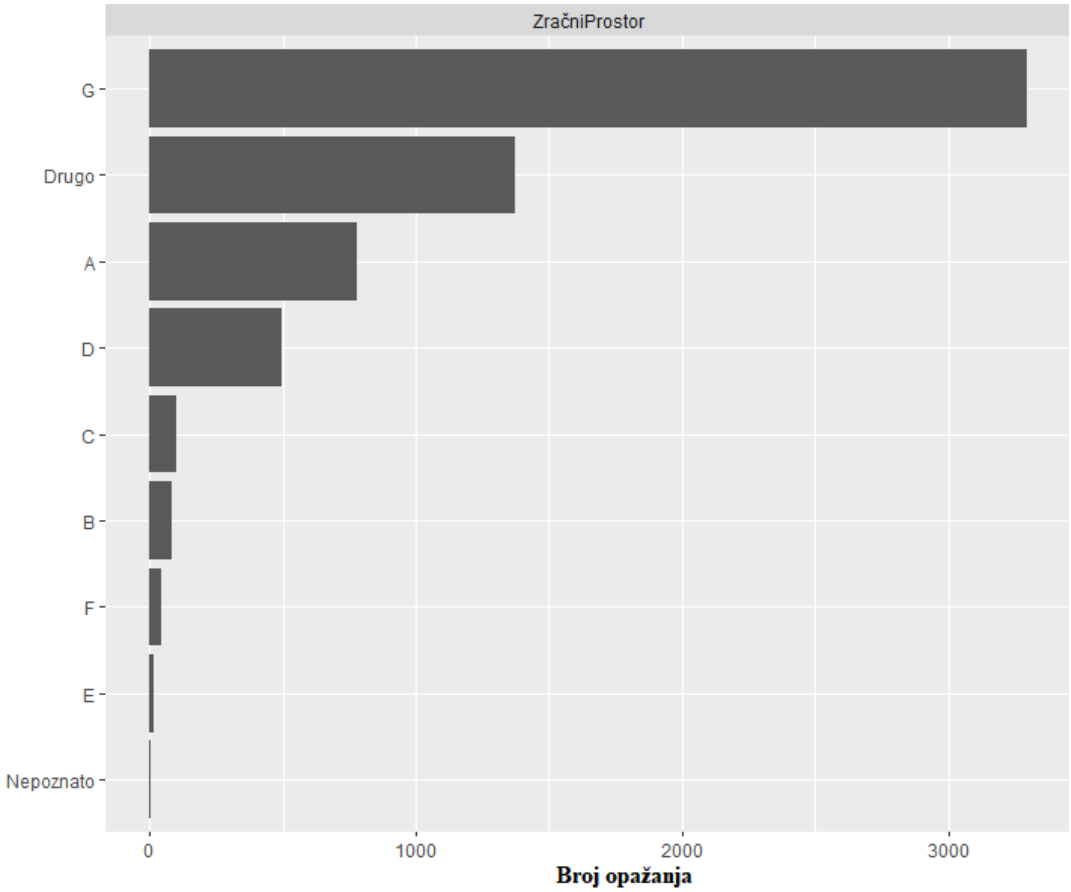


Slika 5.14 Klasifikacija leta prema kategoriji rizika za podatke do kraja 2016. godine

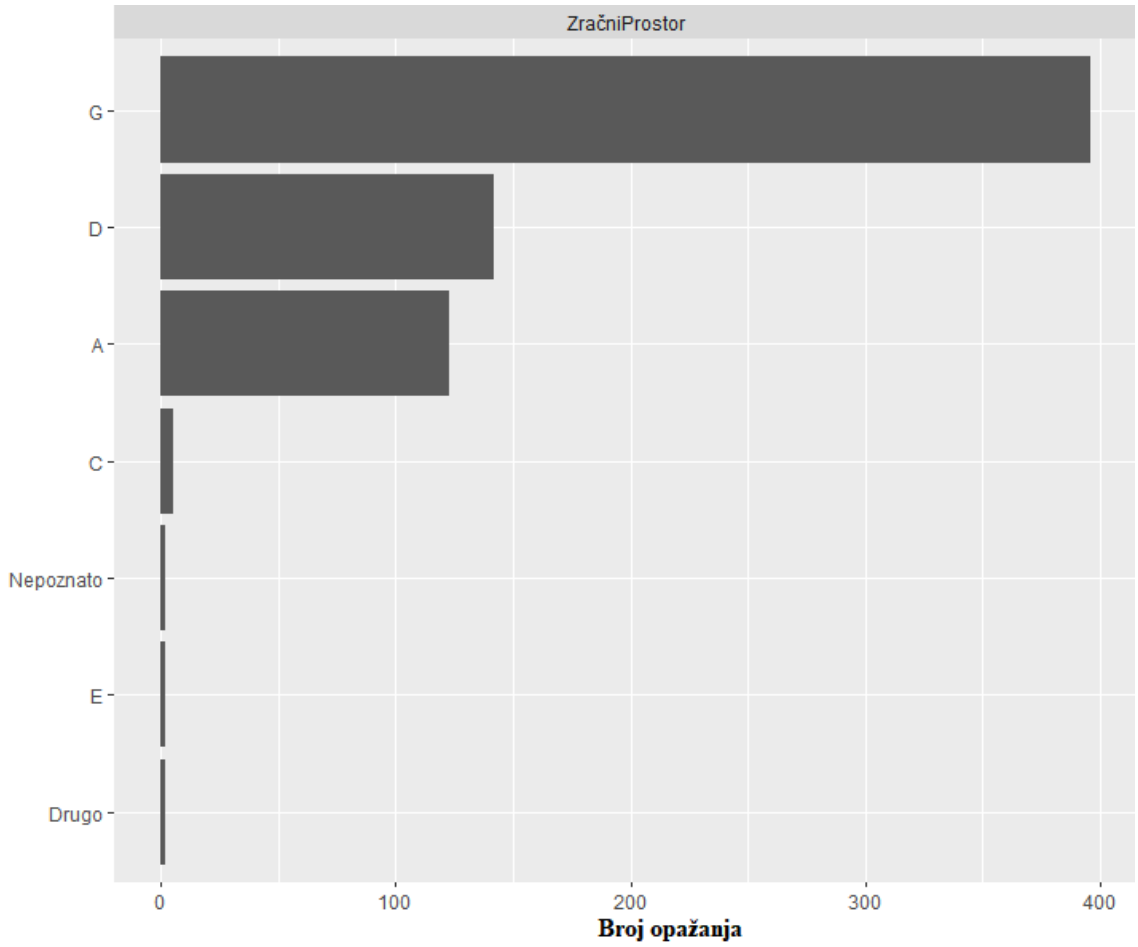


Slika 5.15 Klasifikacija leta prema kategoriji rizika za podatke 2017. godine

Posljednja zanimljiva opažanja su opažanja zračnog prostora, slika 5.16 i 5.17. Zračni prostor do kraja 2016. godine raspršen je u više kategorija usporedno s 2017. godinom, a u oba vremenska perioda kategorija zračnog prostora G je daleko najzastupljenija. Za period 2017. godine gotovo svi prijavljeni incidenti spadaju unutar jednu od tri kategorije zračnog prostora: G, D ili A. Kategorija zračnog prostora G, koja je najzastupljenija, je ujedno i slabije kontrolirana kategorija kod koje su dozvoljeni letovi u skladu s pravilima instrumentalnog letenja (en. Instrument Flight Rules) i letovi prema pravilima vizualnog letenja (en. Visual Flight Rules) te je navedenim letovima pružena služba letnih informacija (en. Flight Information Service) samo na zahtjev. Za razliku od kategorije zračnog prostora G, kategorije A i D imaju razdvajanje letova u skladu s pravilima instrumentalnog letenja, svi letovi primaju informacije od strane službe letnih informacija te kategorija A dozvoljava samo letove u skladu s pravilima instrumentalnog letenja, dok kategorija D dozvoljava letove oba tipa, ali za dodatnu sigurnost letovi u skladu s pravilima instrumentalnog letenja primaju informacije o svim letovima prema pravilima vizualnog letenja, dok letovi prema pravilima vizualnog letenja primaju informacije o svim drugim letovima.



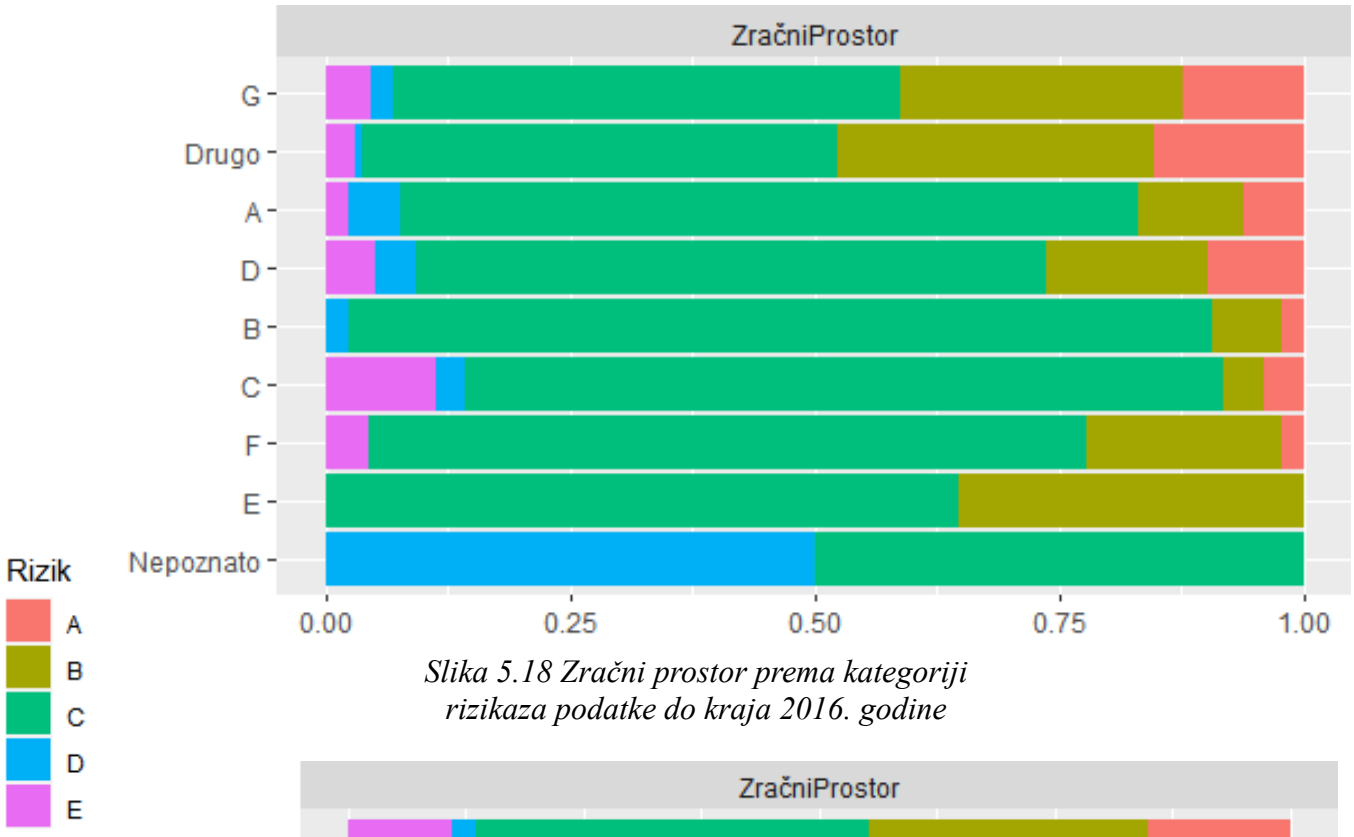
Slika 5.16 Zračni prostor za podatke do kraja 2016. godine



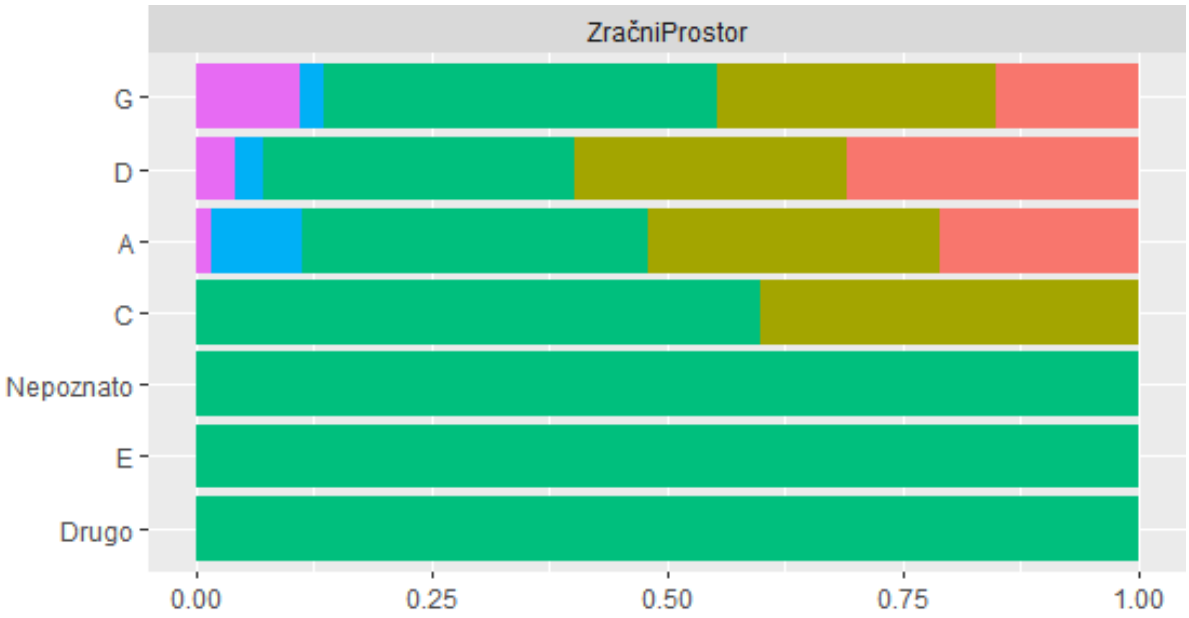
Slika 5.17 Zračni prostor za podatke 2017. godine

Kategorije zračnog prostora A i D su poprilično kontrolirane kategorije i očekivano je manji broj prijave incidenata rizičnih kategorija do kraja 2016. godine, kao što se može vidjeti na slici 5.18. Očekivano bi bilo i za 2017. godinu da kontroliranije kategorije zračnog prometa pružaju veću sigurnost letjelicama, ali kao što je vidljivo na slici 5.19. to nije slučaj. Kategorije zračnog prometa A i D 2017. godine imaju dvostruko veću količinu rizičnih kategorija. Iako su kategorije zračnog prometa A i D 2017. godine imale porast rizičnih kategorija, kategorija G koja je najmanje kontroliran zračni prostor je ostala na približno istom omjeru rizičnih kategorija. Promatranjem slika 5.20 i 5.21 možemo bolje razumjeti navedeni fenomen. Slike prikazuju kategoriju zračnog prostora prema klasifikaciji leta i možemo vidjeti da kategorija zračnog prostora G koja se minimalno promijenila iz perioda do kraja 2016. godine i perioda 2017. godine posjeduje malen broj klasifikacija leta kategorije komercijalni zračni promet. Kategorije zračnog prometa A i D posjeduju velik

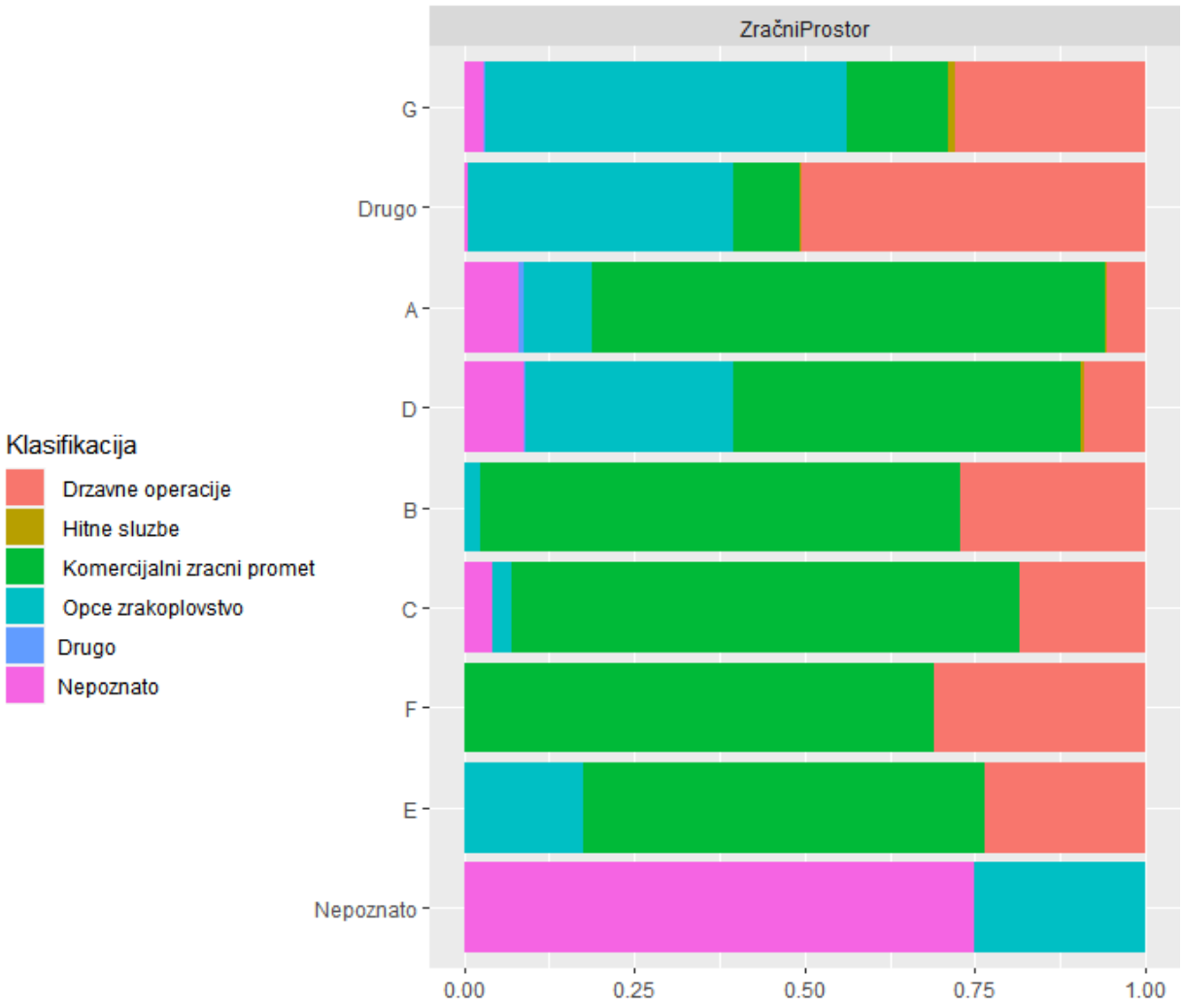
broj prijavljenih incidenata s klasifikacijom leta kategorije komercijalnog zračnog prometa i time možemo zaključiti da letjelice komercijalnog zračnog prometa su glavne žrtve povećanja korištenja bespilotnih letjelica.



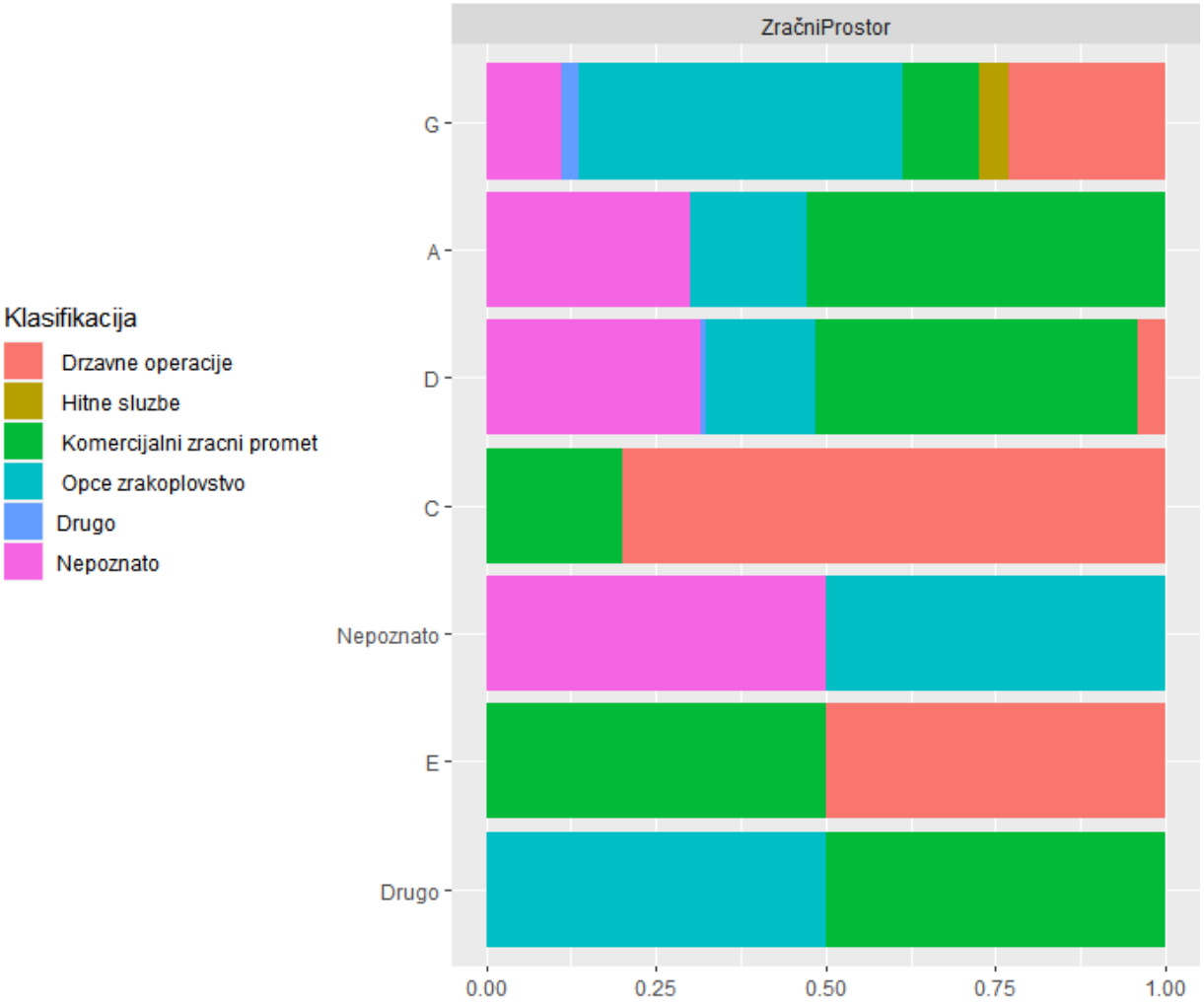
Slika 5.18 Zračni prostor prema kategoriji rizikaza podatke do kraja 2016. godine



Slika 5.19 Zračni prostor prema kategoriji rizika za podatke 2017. godine



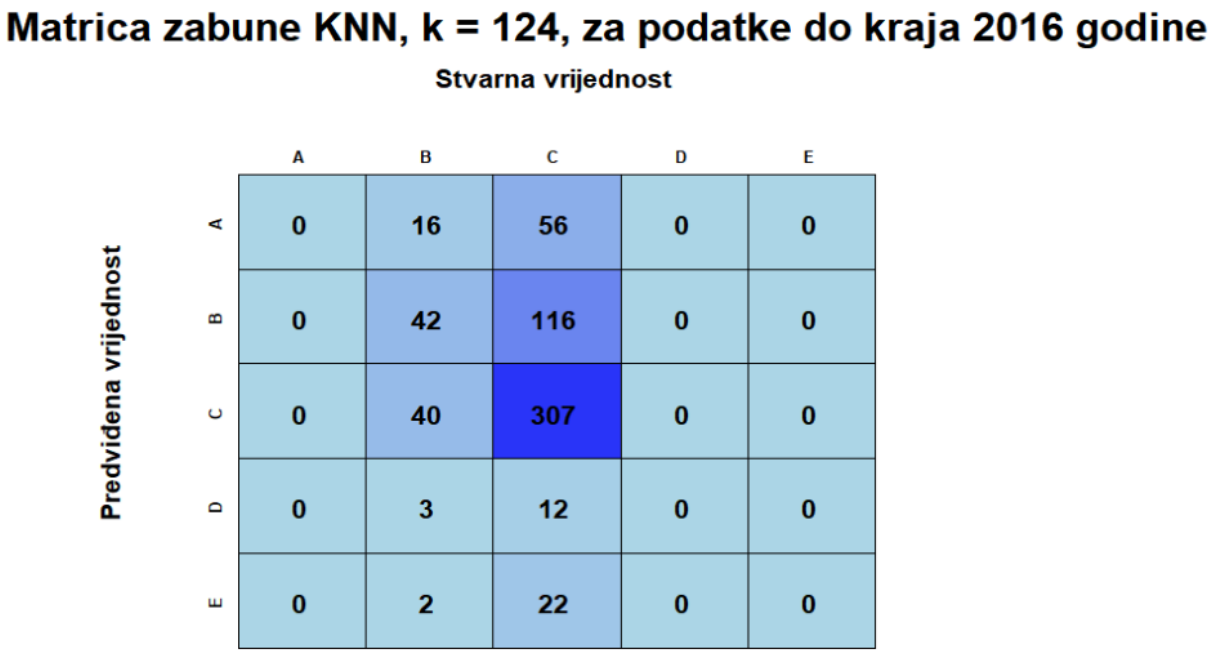
Slika 5.20 Zračni prostor prema klasifikaciji leta za podatke do kraja 2016. godine



Slika 5.21 Zračni prostor prema klasifikaciji leta za podatke 2017. godine

5.3 Rezultati klasifikacijskog modela k-najbližih susjeda

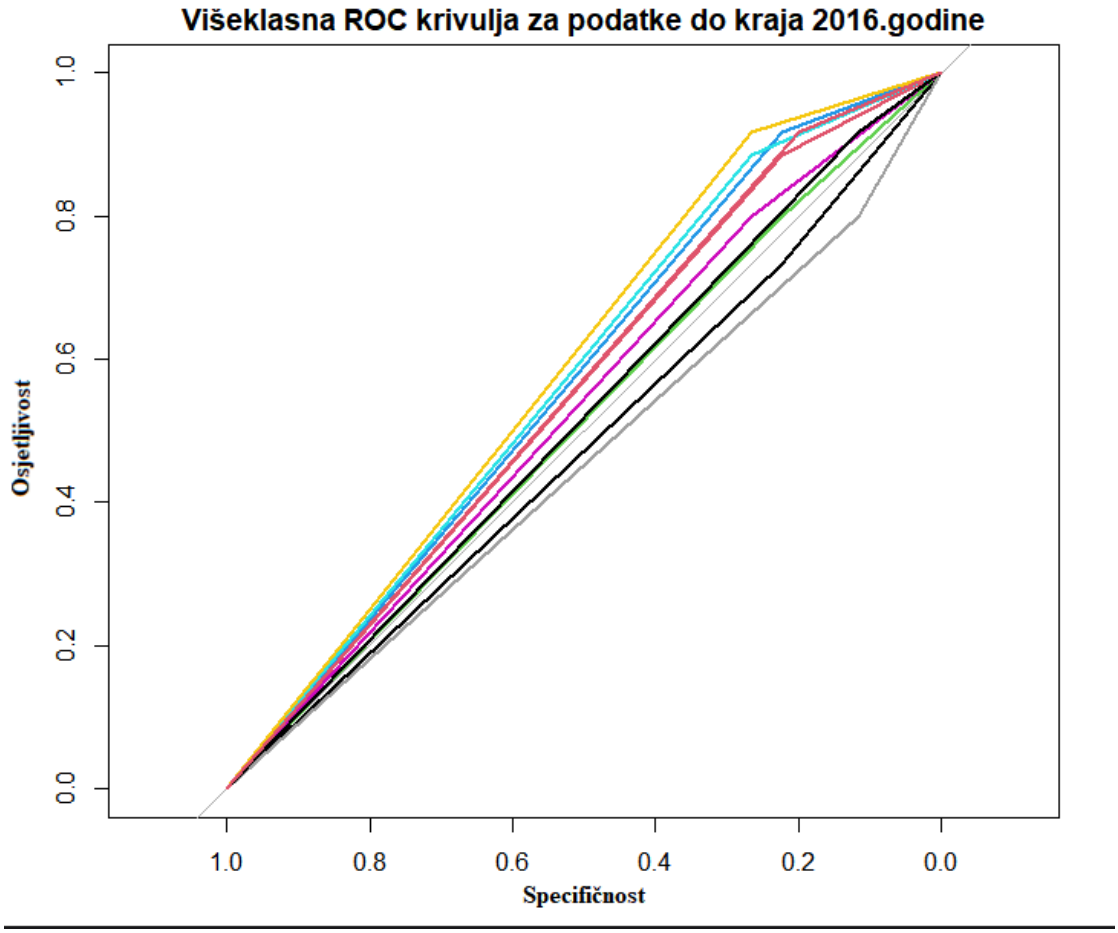
Rezultati klasifikacije koristeći k-najbližih susjeda nad podacima do kraja 2016. godine prikazani su na slici 5.22. Iz slike matrice zabune možemo vidjeti da model klasifikacije vrijednosti kategorije rizika koja se sastoji od pet razina svrstava u samo dvije razine, B i C. Točnost klasifikacijskog modela iznosi 0.5666, s intervalom pouzdanosti između 0.5264 i 0.6061. P-vrijednost iznosi 1 što ukazuje da model nije bolje od nasumičnog pogađanja. Stopa bez informacija iznosi 0.8328 čime vidimo omjer najčešće razine. Niska kappa vrijednost ukazuje na loše slaganje između modela i točnih razina.



Točnost	Interval pouzdanosti	Stopa bez informacija	p-vrijednost	Kappa
0.5666	(0.5264, 0.6061)	0.8328	1	0.1118

Slika 5.22 Matrica zabune k-najbližih susjeda za podatke do kraja 2016. godine

Višeklasna ROC krivulja ukazuje na modelovu mogućnost razlikovanja više klasa. Iz slike 5.23 se vidi da za sve klase navedeni model nije mnogo bolji od nasumičnog pogađanja, a da je za neke i gori. Iz slike 5.24 vidimo rezultat izračuna višeklasne AUC vrijednosti koji iznosi 0.5344 što znači da je model malo bolji u klasifikaciji od nasumičnog pogađanja.

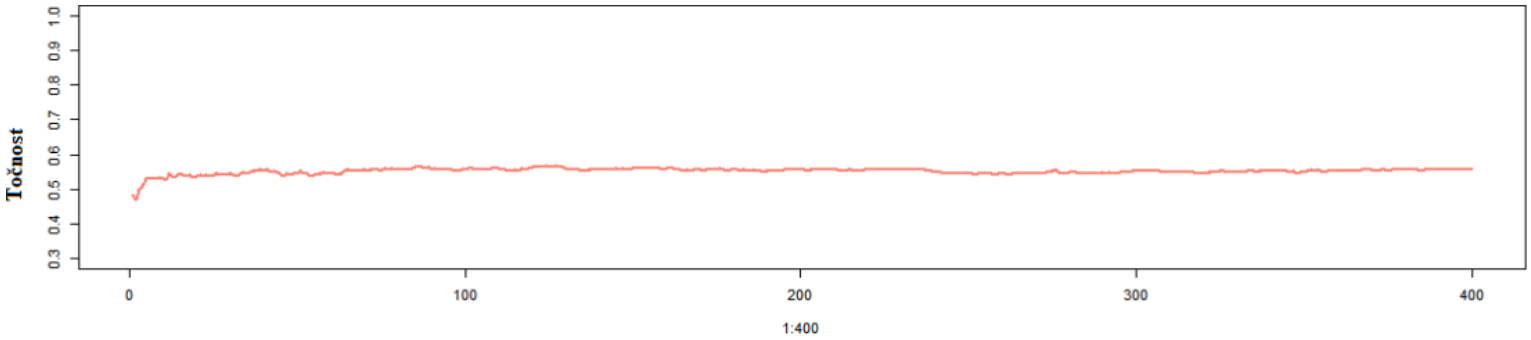


Slika 5.23 ROC krivulja za podatke do kraja 2016. godine

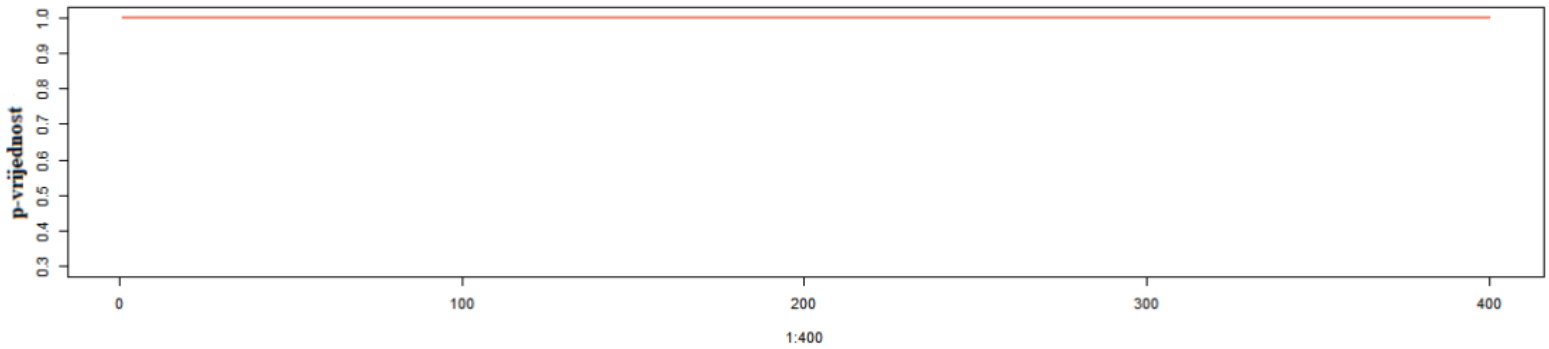
```
Call:
multiclass.roc.default(response = ValuesTest2016, predictor = testPredicted2016)

Data: testPredicted2016 with 5 levels of ValuesTest2016: 1, 2, 3, 4, 5.
Multi-class area under the curve: 0.5344
```

Slika 5.24 AUC vrijednost za podatke do kraja 2016. godine

Za pronalaženje hiperparametra za k-najbližih susjeda korištena je petlja od vrijednosti 1-400. Opće pravilo je koristiti korijen od ukupnog broja promatranja no zbog slabijih rezultat on je proširen. Kao što se može vidjeti iz slika 5.25 i 5.26, vrijednost točnosti kreće ispod 0.5 za niske k-eve i naraste iznad 0.5 gdje se zadržava za sve ostale vrijednosti hiperparametra k. P-vrijednost je gotovo konstantna za sve vrijednost hiperparametra i iznosi jedan ili vrijednost vrlo blizu jedan. Kriterij odabira hiperparametra u ovom slučaju je točnost.

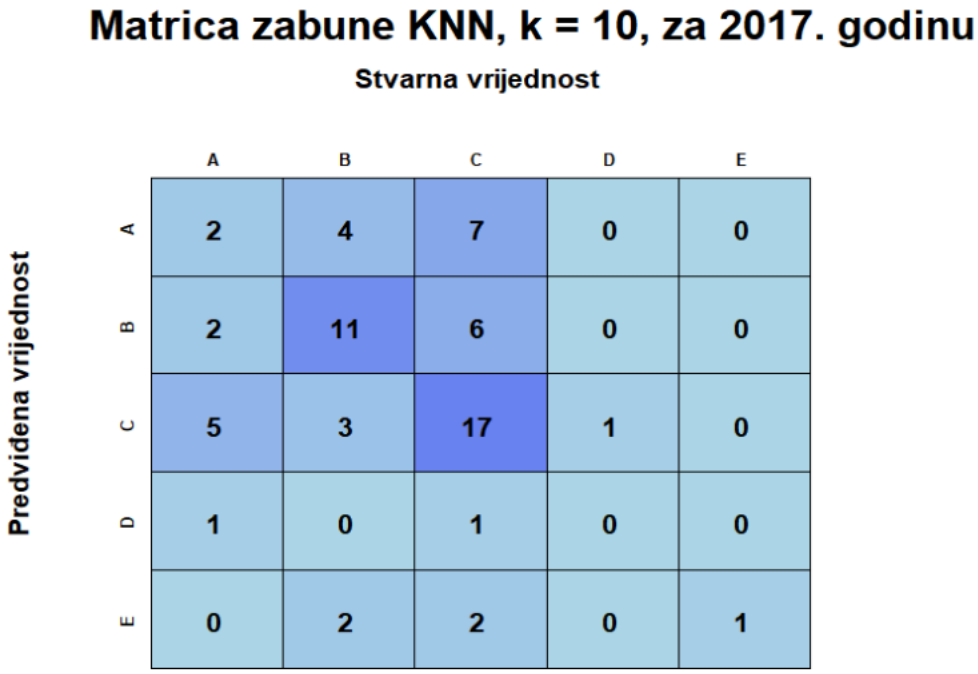


Slika 5.25 Točnost za k=1-400 za podatke do kraja 2016. godine



Slika 5.26 P-vrijedonst za k=1-400 za podatke do kraja 2016. godine

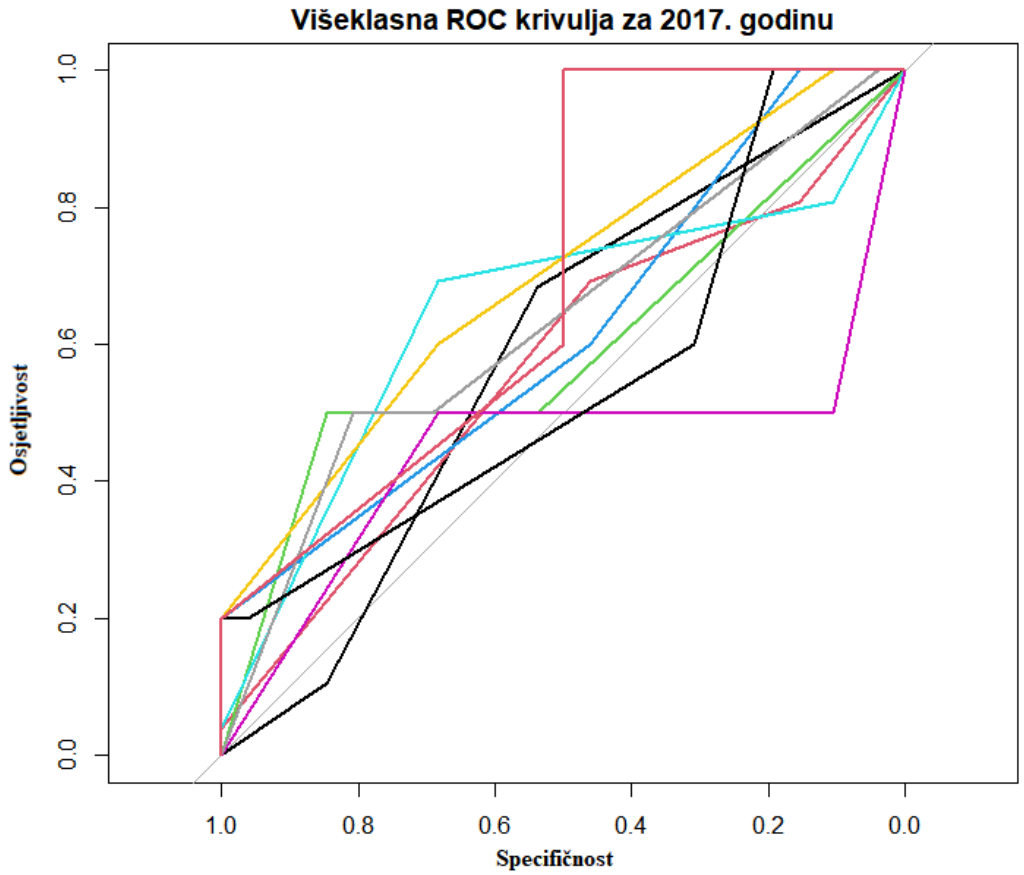
Rezultati iste klasifikacije koristeći podatke za 2017. godinu prikazani su na slici 5.27. Prikazani model klasifikacije koristi sve razine kod svrstavanja. Točnost klasifikacijskog modela iznosi 0.4769 s intervalom pouzdanosti između 0.3515 i 0.6046. P-vrijednost je 0.7324 što je i dalje daleko previsoka vrijednost. Stopa bez informacija iznosi 0.5077 i kappa vrijednost je ponovo niska, 0.2246.



Točnost	Interval pouzdanosti	Stopa bez informacija	p-vrijednost	Kappa
0.4769	(0.3515, 0.6046)	0.5077	0.7324	0.2246

Slika 5.27 Matrica zabune k-najbližih susjeda za podatke 2017. godine

Višeklasna ROC krivulja na slici 5.28 ukazuje da navedeni model nije mnogo bolji od nasumičnog pogađanja. Iz slike 5.29 vidimo rezultat izračuna višeklasne AUC vrijednosti koji iznosi 0.6036 što ga čini boljim u usporedbi s modelom nad podacima do kraja 2016. godine, ali su vrijednosti i dalje niske.

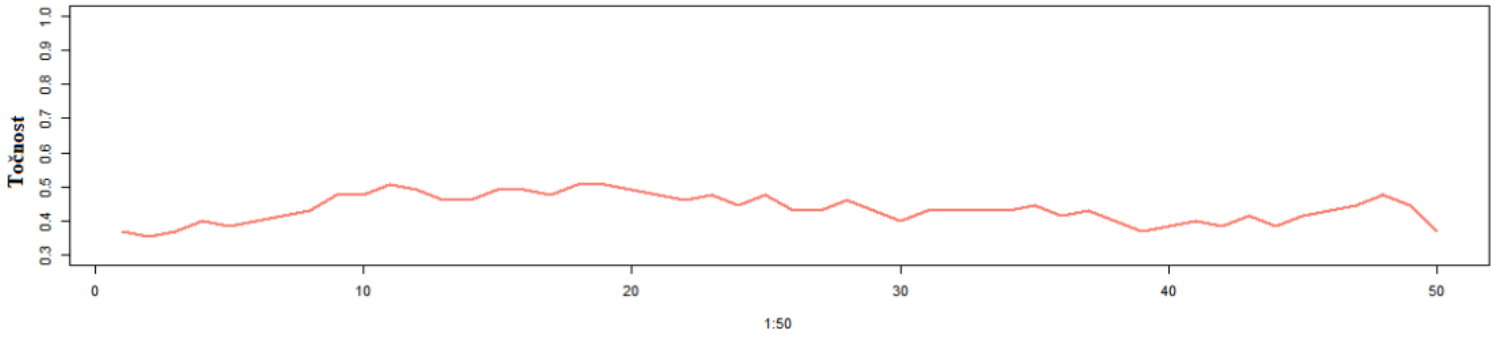


Slika 5.28 ROC krivulja za podatke 2017. godine

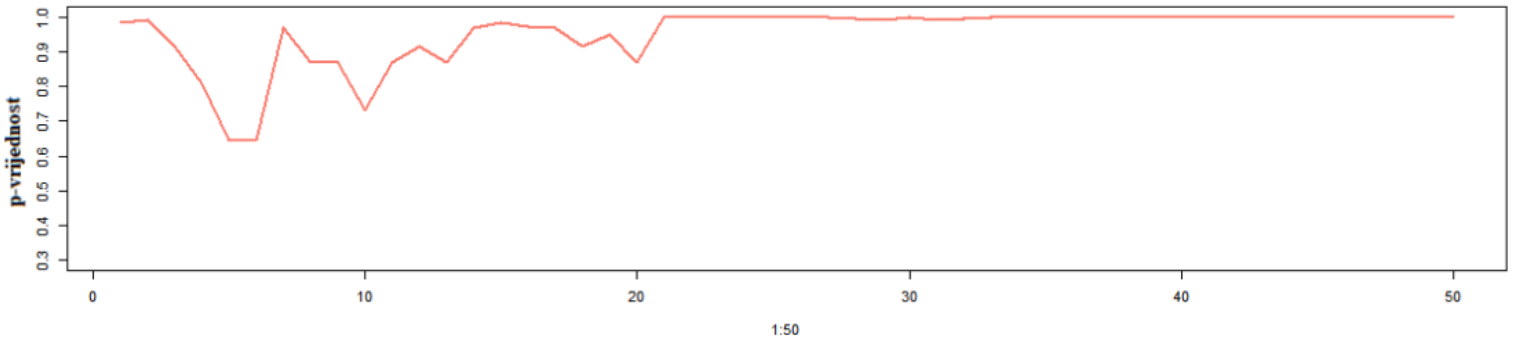
```
Call:
multiclass.roc.default(response = ValuesTest2017, predictor = testPredicted2017)

Data: testPredicted2017 with 5 levels of ValuesTest2017: 1, 2, 3, 4, 5.
Multi-class area under the curve: 0.6036
```

Slika 5.29 AUC vrijednost za podatke 2017. godine

Pronalaženje hiperparametra k koristi manje vrijednosti za podatke 2017. godine budući da testni set sadrži samo 372 opažanja. Proučavanjem slike 5.30 vidi se da točnost u nekim slučajevima prolazi 0.5, ali kao što je vidljivo na slici 5.31 paralelno raste i p-vrijednost. Kriterij korišten za biranje hiperparametra je najveća točnost za najmanju p-vrijednost.



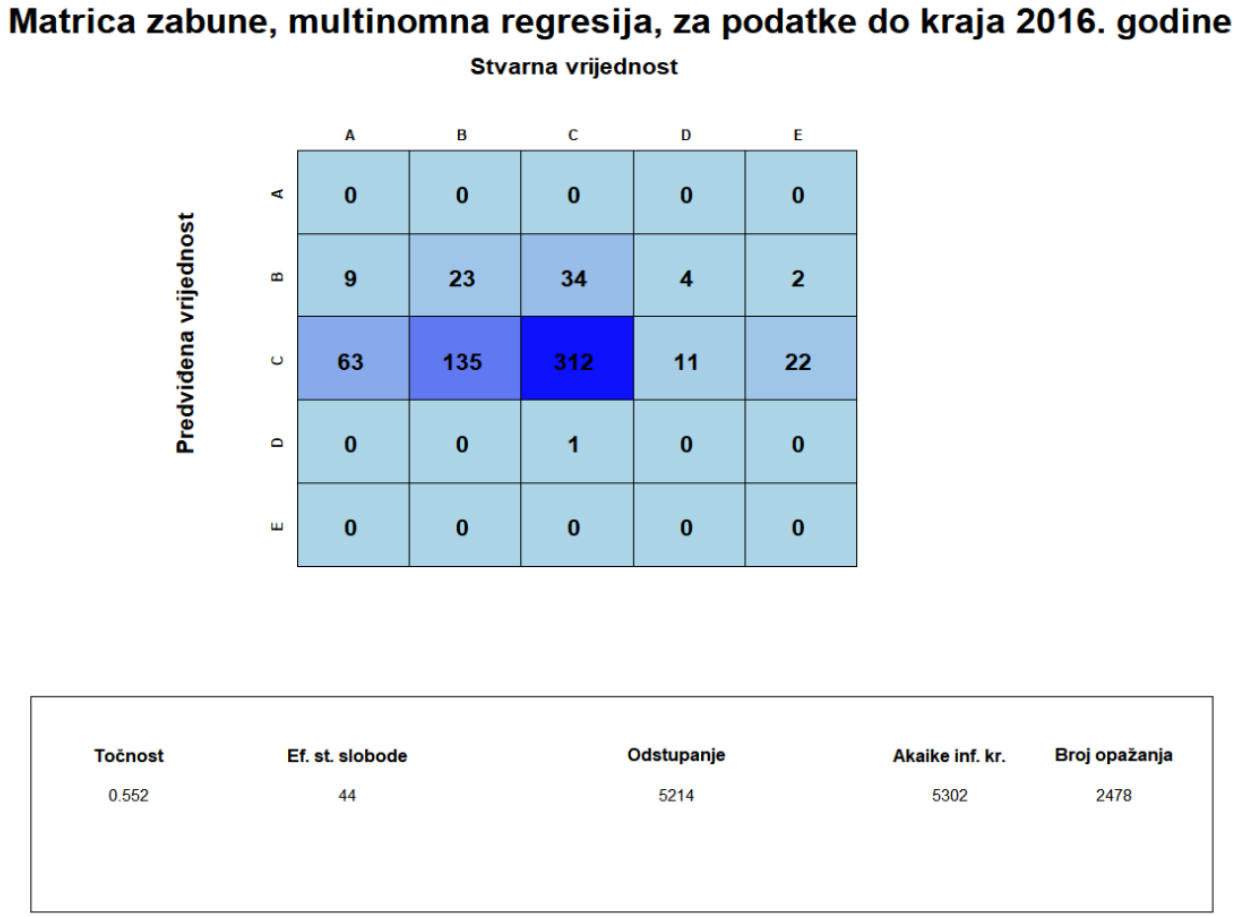
Slika 5.30 Točnost za k=1-50 za podatke 2017. godine



Slika 5.31 P-vrijednost za k=1-50 za podatke 2017. godine

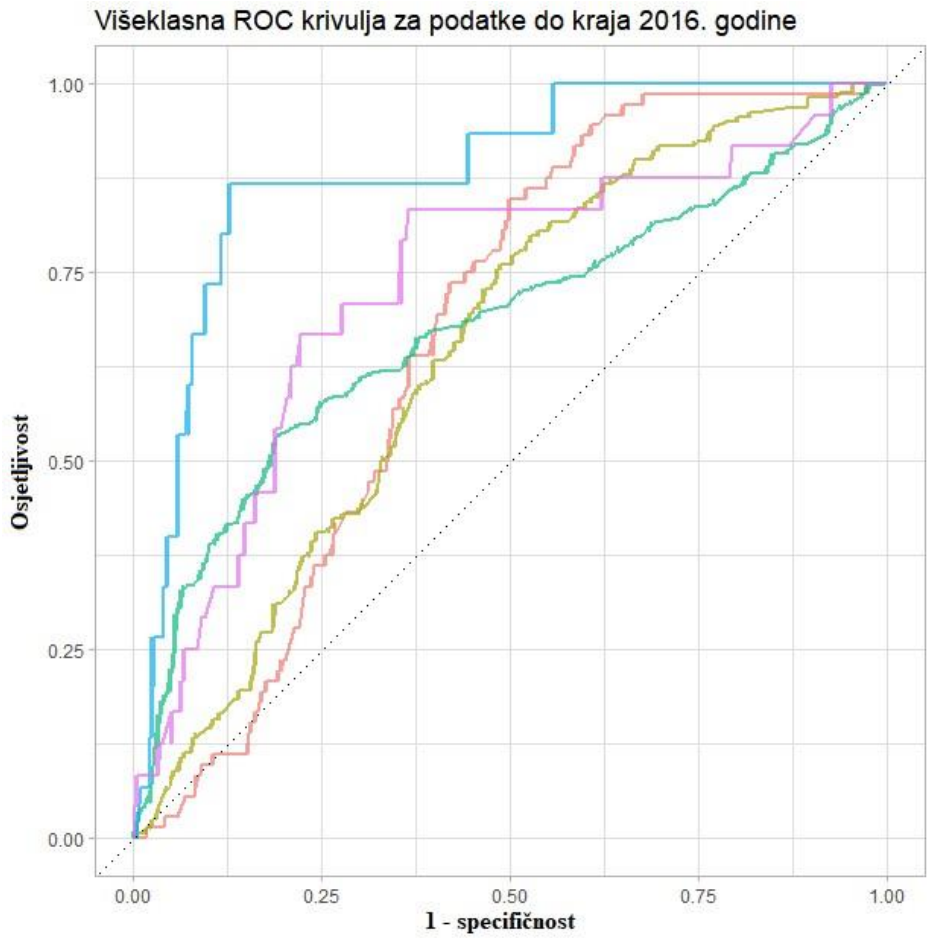
5.4 Rezultati klasifikacijskog modela višečlane logističke regresije

Slika 5.32 prikazuje rezultate klasifikacijskog modela višečlane logističke regresije na podacima do kraja 2016. godine. Klasifikacijski model koristi sve razine kategorija rizika kod svrstavanja. Točnost navedenog modela iznosi 0.532. Vrijednosti efektivnog stupnja slobode, odstupanja i Akalike informacijskog kriterija su apstraktne u vakuumu pa će biti uspoređene s navedenim vrijednostima istog klasifikacijskog modela nad podacima iris skupa podataka koji se često koristi u edukacijske svrhe te je mnogo manje kompleksnosti od skupa podataka pruženog od strane Odbora. Efektivni stupanj slobode 44 ukazuje na mnogo veću složenost klasifikacijskog modela usporedno sa efektivnim stupnjem slobode iris skupa podataka koji iznosi 6. Odstupanje trenutnog klasifikacijskog modela iznosi 5214 što ukazuje na lošiju usklađenost modela s podacima usporedno s odstupanjem iris skupa podataka koje iznosi 4.45. Akaike informacijski kriterij iznosi 5302 i ukazuje na loš kompromis između složenosti modela i usklađenosti modela. Vrijednost Akaike informacijskog kriterija iris skupa podataka iznosi 16.5.



Slika 5.32 Matrica zabune višečlane regresije za podake do kraja 2016. godine

Slika 5.33 prikazuje višeklasnu ROC krivulju. Iz navedene slike vidi se da je svaka klasa ROC krivulje znatno bolje klasificirana modelom nego nasumičnim pogađanjem. Izračunati prostor ispod krivulje iznosi 0.714, prikazano na slici 5.34. Vrijednost AUC ukazuje da klasifikacijski model može razlikovati između razina kategorija rizika.

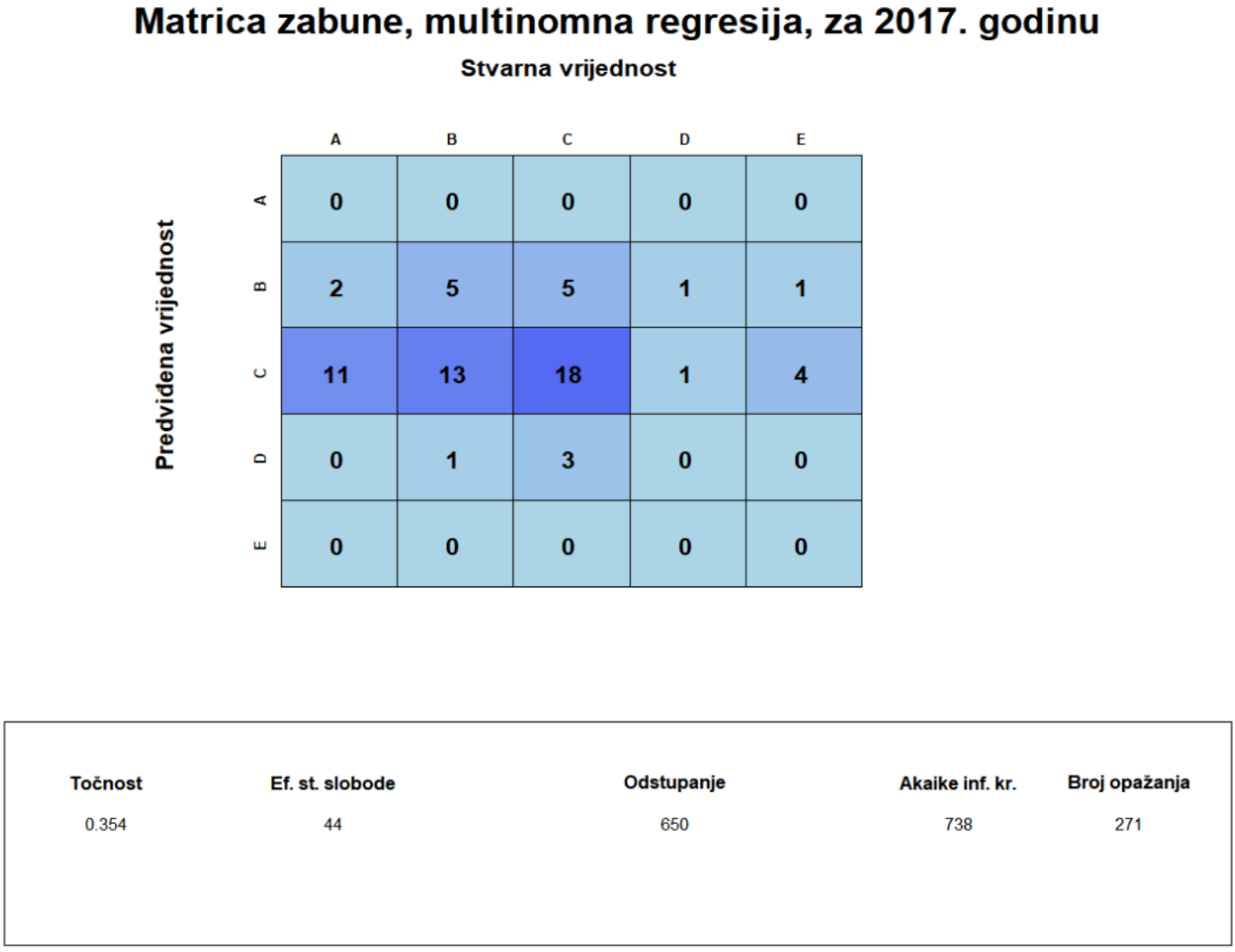


Slika 5.33 ROC krivulja za podake do kraja 2016. godine

```
# A tibble: 1 x 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 roc_auc hand_till      0.714
```

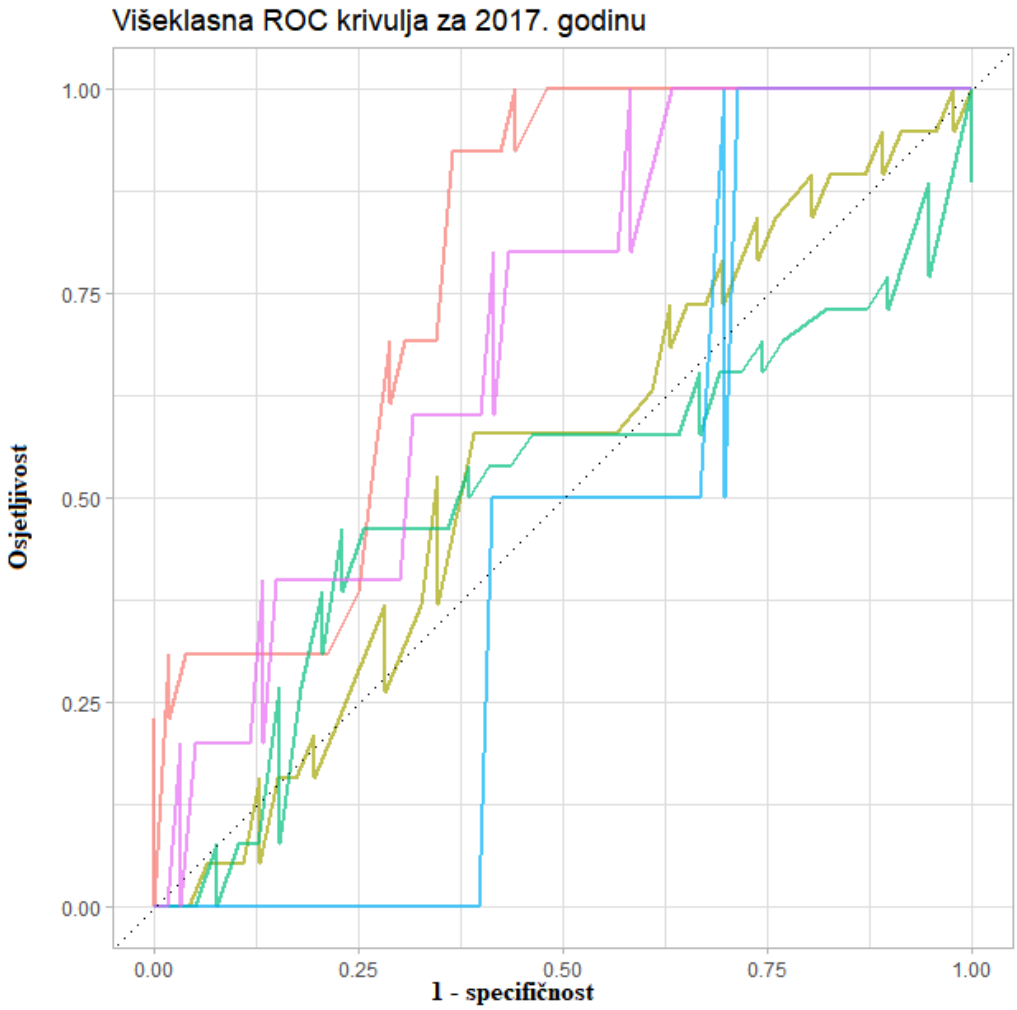
Slika 5.34 AUC vrijednost za podake do kraja 2016. godine

Slika 5.35 prikazuje rezultate klasifikacijskog modela višečlane logističke regresije na podacima za 2017. godinu. Klasifikacijski model koristi sve razine kategorija rizika kod svrstavanja. Točnost navedenog modela iznosi 0.354. Efektivni stupanj slobode iznosi isto kao i u prethodnom periodu, 44. Vrijednost efektivnog stupnja slobode ukazuje na mnogo veću složenost klasifikacijskog modela uspoređujući ga s klasifikacijskim modelom iris skupa podataka. Odstupanje iznosi 650 i Akaike informacijski kriterij iznosi 738, što je mnogo puta manje od klasifikacijskog modela za prethodni vremenski interval, ali i dalje vrlo visok u usporedbi s rezultatima klasifikacijskog modela iris skupa podataka.



Slika 5.35 Matrica zabune višečlane regresije za podake 2017. godine

Višeklasna ROC krivulja prikazana slikom 5.36 ukazuje da klasifikacijski model dobro razlikuje neke klase, no isto take loše razlikuje druge klase što daje konačni AUC rezultat 0.586. Rezultat nije mnogo bolji od nasumičnog pogađanja.



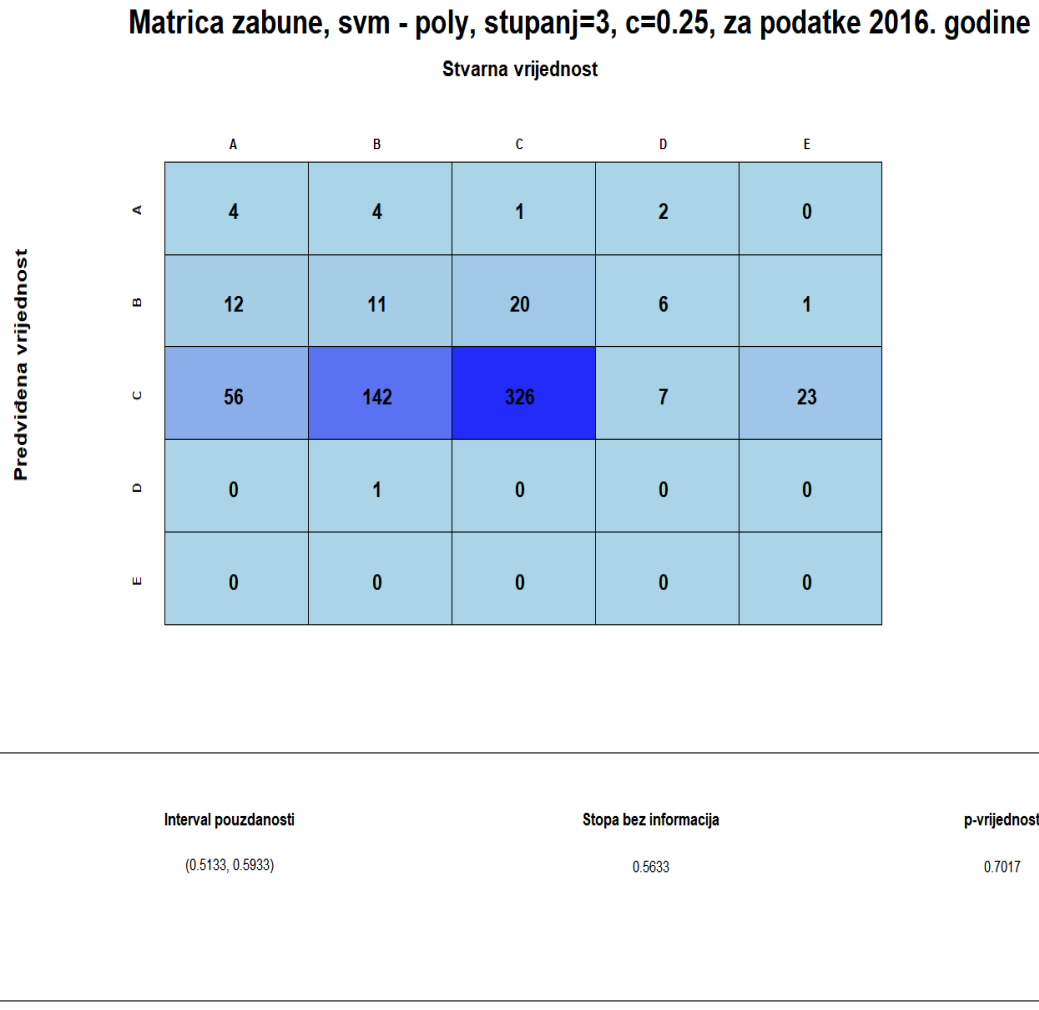
Slika 5.36 ROC krivulja za podake 2017. godine

```
# A tibble: 1 × 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 roc_auc hand_till      0.586
```

Slika 5.37 AUC vrijednost za podake 2017. godine

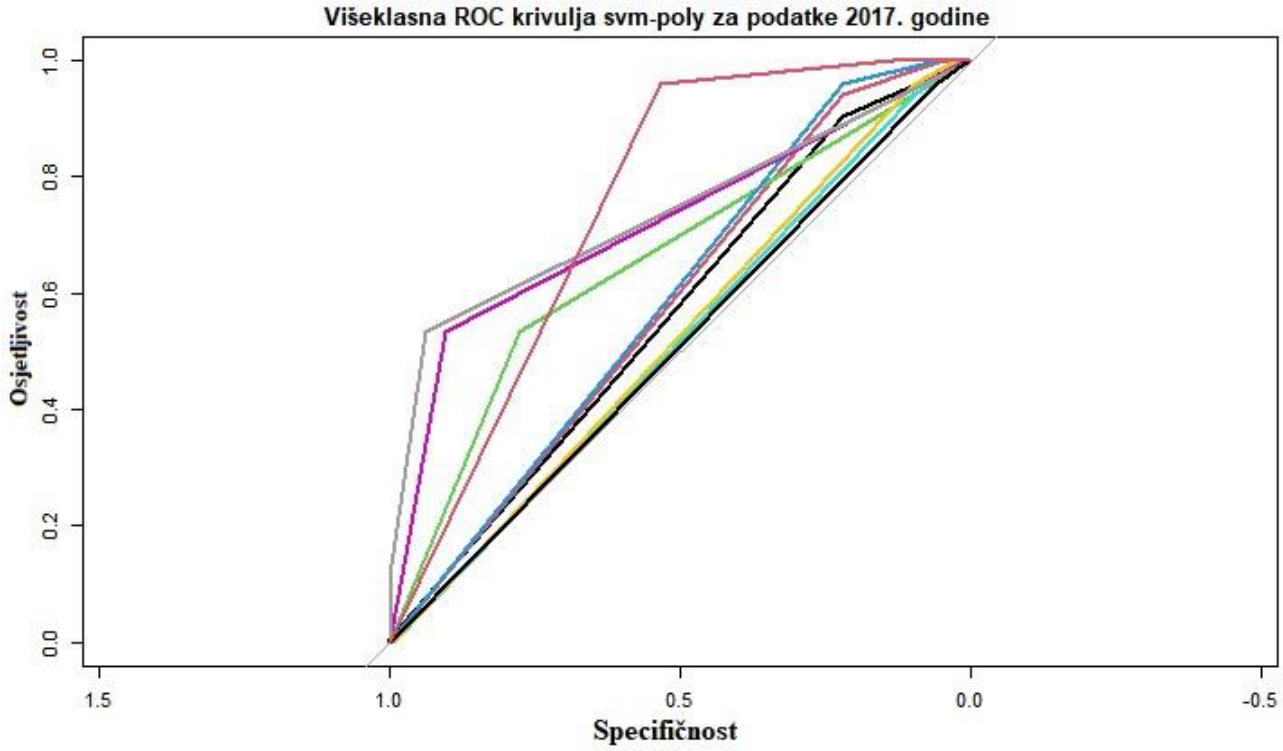
5.5 Rezultati klasifikacijskog modela stroja potpornih vektora

Rezultati klasifikacije koristeći stroj potpornih vektora nad podacima do kraja 2016. godine prikazani su na slici 5.38. Matrica zabune pokazuje da su korištene sve razine kategorije rizika kod svrstavanja. Stupanj=3 predstavlja polinomni stupanj klasifikacijskog modela što znači da kod određivanja prostora pojedine razine kategorije rizika model može koristiti kubične funkcije. Hiperparametar c je odabran unakrsnom provjerom valjanosti i daje najveću točnost za vrijednost 0.25. Točnost navedenog klasifikacijskog modela iznosi 0.5536 s intervalom pouzdanosti između 0.5133 i 0.5933. P-vrijednost iznosi 0.7017. Vrijednost stope bez informacija je 0.5633. Kappa vrijednost je vrlo blizu nule.



Slika 5.38 Matrica zabune stroja potpornih vektora za podake do kraja 2016. godine

Višeklasna ROC krivulja na slici 5.39 prikazuje da je model malo bolji od nasumičnog pogađanja i to za svaku razinu. To nadalje potvrđuje slika 4.50 s rezultatom AUC od 0.6152. Iako je klasifikacijski model bolji od nasumičnog pogađanja to je i dalje vrlo niska AUC vrijednost.

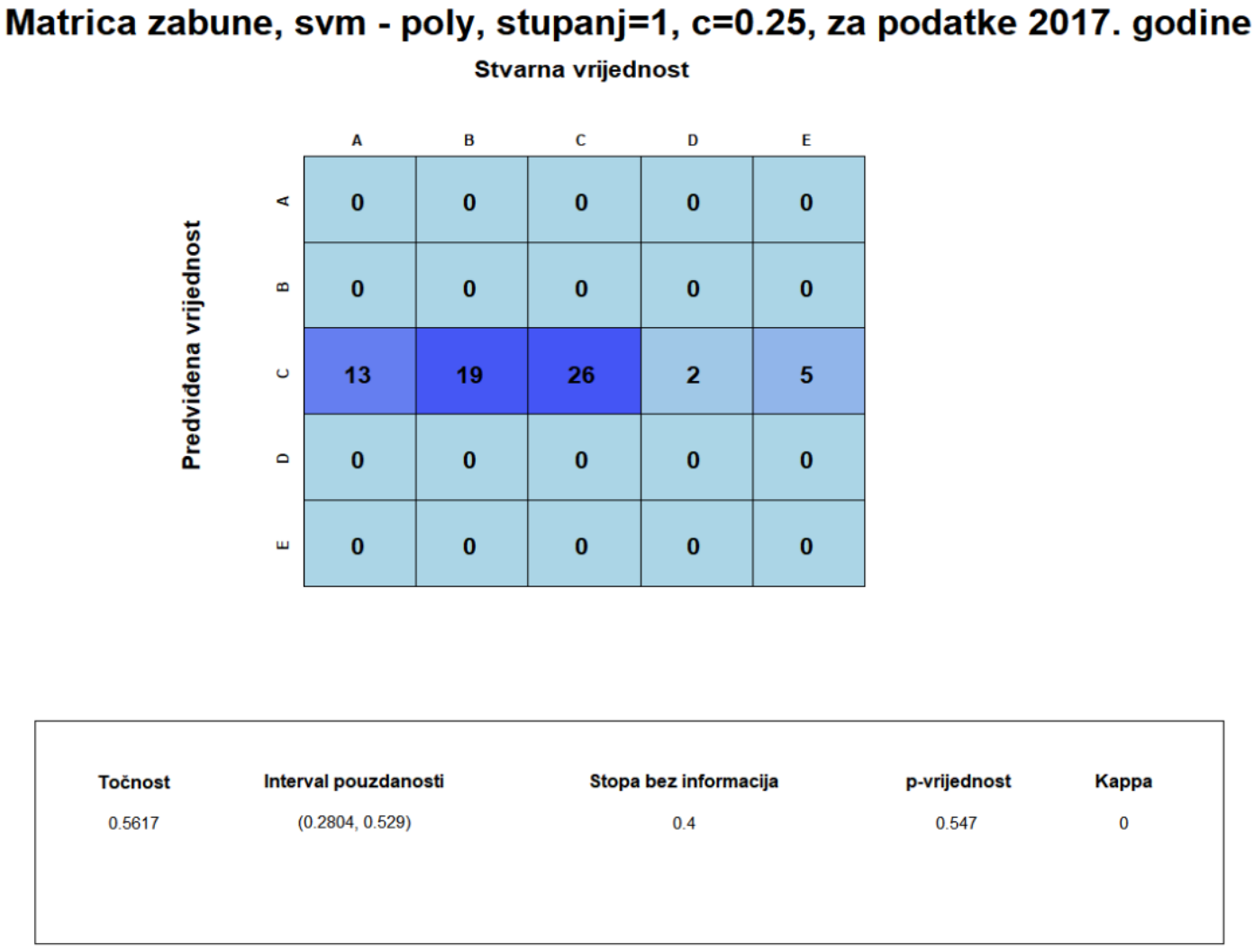


Slika 5.39 ROC krivulja za podake do kraja 2016. godine

```
Call:
multiclass.roc.default(response = test2016$Rizik, predictor = as.numeric(predictions))

Data: as.numeric(predictions) with 5 levels of test2016$Rizik: 1, 2, 3, 4, 5.
Multi-class area under the curve: 0.6152
```

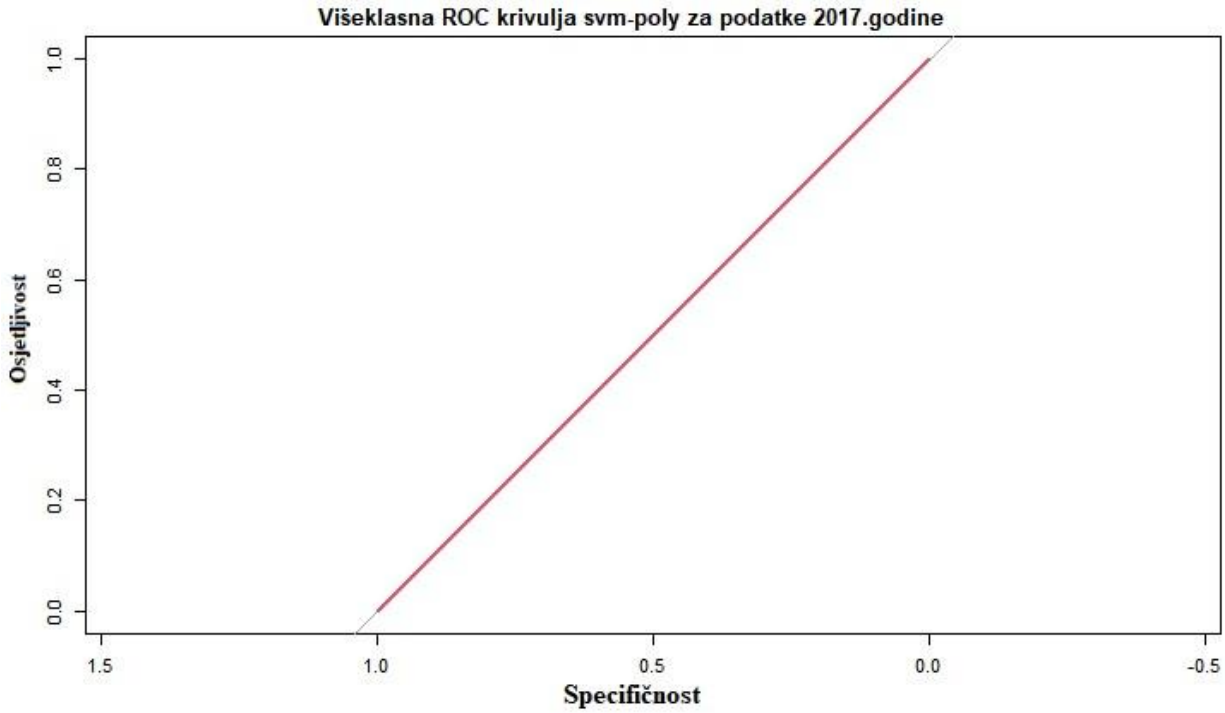
Slika 5.40 AUC vrijednost za do kraja 2016. godine

Za podatke 2017. godine, klasifikacijski model stroja potpornih vektora koristi samo jednu razinu kategorije rizika kod svrstavanja, razinu C, slika 5.41. Za razliku od klasifikacijskog modela za podatke do kraja 2016. godine koji je koristio stupanj=3, prikazani model koristi stupanj=1 koji ukazuje da se koristi originalan prostor kod klasifikacije. Hiperparametar c je odabran unakrsnom provjerom valjanosti za najveću točnost. Točnost modela iznosi 0.5617 s intervalom pouzdanosti između 0.2804 i 0.529. P-vrijednost je niža nego kod podataka prethodnog vremenskog perioda i iznosi 0.547. Stopa bez informacija je 0.4 i kappa vrijednost je 0.



Slika 5.41 Matrica zabune stroja potpornih vektora za podake 2017. godine

Višeklasna ROC krivulja sa slike 5.42 prikazuje da klasifikacijski model ne primjećuje razlike između razina kategorije rizika. Slika 5.34 prikazuje vrijednost AUC koja je točno 0.5 što znači da klasifikacijski model pruža iste rezultate kao i nasumično pogađanje.



Slika 5.42 ROC krivulja za podake 2017. godine

```
Call:
multiclass.roc.default(response = test2017$Rizik, predictor = as.numeric(predictions))

Data: as.numeric(predictions) with 5 levels of test2017$Rizik: 1, 2, 3, 4, 5.
Multi-class area under the curve: 0.5
```

Slika 5.43 AUC vrijednost za podake 2017. godine

6. INTERPRETACIJA I RASPRAVA

Opisnom statističkom analizom podataka može se uočiti više zanimljivosti. Usporedbom podataka lokacija letjelica prijavljenih incidenata, slika 5.5 i slika 5.6, može se uočiti veliko sniženje omjera u prijavama incidenata u drugom promatranom vremenskom periodu za sve lokacije osim Londona sa širom okolinom. Sniženje omjera ukazuje na sigurnije letove letjelica kroz Ujedinjeno Kraljevstvo osim Londona sa širom okolinom, no navedeni omjer se može objasniti i u situaciji da se 2017. godine prijavilo mnogo više incidenata u Londonu sa širom okolinom usporedno s prvim promatranim vremenskim periodom što bi umanjilo omjere prijavljenih incidenata s drugih lokacija. Na slici 5.4 vidi se da u drugom promatranom periodu nije prijavljen niti jedna incident za cijelu irsku te isto tako nije bilo prijava incidenata za cijelo područje sjeverozapadne škotske. Opažanjem rezultata može se zaključiti da se sigurnost letova kroz Ujedinjeno Kraljevstvo, bez Londona sa širom okolinom, povećala za vrijeme drugog promatranog perioda dok se sigurnost letjelica unutar Londona sa širom okolinom umanjila.

Usporedbom visine leta u oba promatrana vremenska perioda, slika 5.7, može se primijetiti da za vrijeme drugog promatranog perioda nije prijavljena niti jedan incident iznad 25 000 stopa (7.62 km), za razliku od prvog promatranog vremenskog perioda gdje je bio dovoljan broj navedenih prijava incidenata. Nadalje, usporedbom kutijastih dijagrama visina na slikama 5.8 i 5.9 vidi se da se u prosjeku visina prijavljenih incidenata smanjila što ukazuje da ako i nije uvedena maksimalna visina leta za bespilotne letjelica, uvedena je regulacija kojom su smanjene visine leta bespilotnih letjelica.

Slike 5.10 i 5.11 prikazuju raspodjelu kategorije rizika za oba promatrana vremenska perioda. Iz slika je vidljivo da u drugom promatranom vremenskom periodu postoji mnogo veći omjer prijavljenih incidenata rizičnih kategorija. Budući da je omjer prijavljenih incidenata mnogo veći u drugom promatranom vremenskom periodu, prijašnje navedeni zaključak da se sigurnost letjelica unutar Londona sa širom okolinom umanjila i dalje stoji. Nadalje, sa slika 5.14 i 5.15 vidi se koja klasifikacija leta najviše pati od navedenog pada sigurnosti letjelica. Tri daleko najzastupljenije klasifikacije leta su opće zrakoplovstvo, komercijalni zračni prostor i državne operacije. Usporedbom oba promatrana vremenska perioda može se primijetiti da je porast rizičnih kategorija za klasifikacije leta opće zrakoplovstvo i državne operacije minimalan, dok se kod klasifikacije leta komercijalni zračni promet omjer rizičnih kategorije više nego udvostručio. Promatranjem podataka može se zaključiti da velika većina porasta u rizičnim kategorijama proizlazi iz komercijalnog zračnog prometa, koji je do drugog promatranog perioda bio daleko najsigurniji. Zaključak je dodatno potvrđen promatranjem kategorija rizika zračnog prostora na slikama 5.18 i 5.19 gdje se vidi da kategorije zračnog prostora A i D koje su kontroliranije, čime bi trebale biti sigurnije, imaju nagli porast rizičnih kategorija u drugom

promatranom vremenskom periodu. Uz to, kod navedenih kategorija A i D najzastupljeniji su letovi komercijalnog zračnog prometa, prikazano na slikama 5.20 i 5.21, dok se kategorija zračnog prometa G ,koja sadrži minimalne promjene omjera rizičnih kategorija prijavljenih incidenata, sastoji najmanjim djelom od komercijalnog zračnog prometa. Prethodno navedeni zaključci o narušavanju sigurnosti komercijalnog zračnog prometa su ujedno i doveli do potrebe za novim regulatornim mjerama.

Rezultati razvijenih klasifikacijskih modela, k-najbližih susjeda, višečlana regresija i strojnih potpornih vektora su drugačiji od očekivanih. Za prvi promatrani period vremena najuspješniji klasifikacijski model bio bi model višečlane regresije. Model ima točnost 0.552 i prostor ispod krivulje 0.714 što ukazuje da iako nije previše točan proces svrstavanja je bolji od nasumičnog pogađanja. Vrlo slične rezultate pruža klasifikacijski model stroja potpornih vektora no kod njega je prostor ispod krivulje znatno manji, dok klasifikacijski model k-najbližih susjeda ima p-vrijednost 1 i prostor ispod krivulje 0.53 što ukazuje na nasumični pogađanje. Najuspješniji klasifikacijski model za drugi promatrani period vremena bio bi klasifikacijski model k-najbližih susjeda. Točnost navedenoga je 0.4769 te mu je p-vrijednost 0.73. Kod drugog promatranih perioda višečlana logistička regresija ima točnost 0.35 što je jako nisko, a rezultati klasifikacijskog modela stroja potpornih vektora pružaju veću točnost, 0.56, ali navedena točnost dobivena je svrstavanjem svih kategorija rizika unutar najpopularnije C kategorije. Iako su višečlana logistička regresija za prvi promatrani period i k-najbližih susjeda za drugi promatrani period navedeni kao najuspješniji modeli njihovi rezultati nisu ni blizu zadovoljavajući. Svi navedeni klasifikacijski modeli isprobani su i na iris skupu podataka koji ima karakteristike pogodne za učinkovito treniranje i evaluaciju modela. Rezultati nad navedenim skupom podataka su za svaki model iznimni, na primjer, kod klasifikacijskog modela k-najbližih susjeda s hiperparametrom k=3 dobije se točnost 0.97, p-vrijednost 0 i prostor ispod krivulje 0.9847. Tražeći razloge za slabu uspješnost klasifikacijskih modela dolazimo do raspodjele podataka unutar svake kategorije prediktora i ciljane varijable. Kategorija rizika za podatke prvog promatranog vremenskog perioda se sastoji preko 80% od kategorija rizika C i B. Uz to, kategorija rizika C predstavlja više od 50% ukupne kategorije rizika. Promatrajući ostale prediktore možemo doći do istog zaključka, svaki prediktor ima jedan ili nekoliko klasa koje čine veliku većinu prediktora. Veliki razmjer oko ciljane varijable uvjetuje stvaranju klasifikacijske modele koji visoko naginju navedenim kategorijama. Nadalje, podaci za drugi promatrani period vremena imaju isti problem s tim da se klasa nepoznato, koja je bila vrlo nezastupljena u prvom periodu, sada pojavljuje u velikom broju. Dodatno, sveukupni broj opažanja je mnogo manji što uzrokuje puno manji skup podataka za treniranje klasifikacijskog modela. Rezultati bi se mogli poboljšati daljnjim istraživanjem. Promatranjem korištenih prediktora mogli bi se izbaciti prediktori koji značajno ne pridonose odluci o ciljnoj varijabli. Uz to, podaci za drugi period vremena se mogu proširiti podacima narednih godina. Također, kod ugađanja klasifikacijskih modela se mogu koristiti druge metrike uspješnosti za razliku od točnosti i p-vrijednosti.

7. ZAKLJUČAK

Rastom korištenja bespilotnih letjelica unutar zračnog prostora dolazi do nacionalne regulacije uvedene 2017. godine, u smislu poboljšanja korištenja zračnog prostora i povećanja sigurnosti. Cilj rada je bio istražiti utjecaj navedenih regulacija na zračni prostog Ujedinjenog Kraljevstva usporedbom podataka prijavljenih incidenata u vremenskom periodu od 2000. godine do kraja 2016. godine i za vremenski period 2017. godine. Interpretacijom rezultata može se vidjeti povećanje sigurnosti letjelica za većinski dio Ujedinjenog Kraljevstva s iznimkom Londona i njemu šire okoline. Dok je sigurnost letjelica u drugim predjelima povećana, za teritorij Londona i njemu šire okolice sigurnost letjelica je u daljnjem padu. Teza istraživanja je bila pretpostavka da novo uvedene nacionalne regulacije povećavaju sigurnost zračnog prometa Ujedinjenog Kraljevstva. Teza istraživanja je djelomično potvrđena prethodno navedenim opažanjima. Zračni prostor Londona sa širom okolinom je nakon uvođenja nacionalne regulacije nesigurniji. London sa širom okolinom je brojem letjelica najzastupaniji dio Ujedinjenog Kraljevstva i navedeno područje treba daljnje istražiti. Proučavanjem pruženih skupova podataka vidi se iznenadan rast rizičnih kategorija od 2014. godine, što je ujedno i početak rasta korištenja bespilotnih letjelica. Za precizniju sliku utjecaja nacionalnih regulacija uvedenih 2017. godine za naredna istraživanja trebale bi se koristiti kasnije godine prvoga promatranoga perioda vremena, budući da do 2014. godine nije bilo znatnog korištenja bespilotnih letjelica. Nadalje, za prikupljene podatke perioda 2017. godine postoji mnogo prijavljenih incidenata kod kojih je više unosa po kategorijama nepoznat. Prijave incidenata sa nepoznatim kategorijama otežavaju svaku vrstu analize nad navedenim podacima.

Izradom trenutnog istraživanja razvijena su tri klasifikacijska modela za oba promatrana vremenska perioda. Cilj razvoja je provjera učinka nacionalne regulacije uvedene 2017. godine. Najuspješniji klasifikacijski model za prvi promatrani period je klasifikacijski model višečlane logističke regresije. Za drugi promatrani period najuspješnije rezultate daje klasifikacijski model k-najbližih susjeda. Za razvijene klasifikacijske modele k-najbližih susjeda i stroj potpornih vektora pokazatelji uspješnosti: točnost, p-vrijednost, ROC krivulja i prostor ispod krivulje ukazuju da navedeni klasifikacijski modeli nisu mnogo uspješniji od nasumičnog pogađanja. Preostali klasifikacijski modeli višečlane logističke regresije koriste druge pokazatelje uspješnosti: točnost, efektivni stupanj slobode, odstupanje, Akaike informacijski kriterij, ROC krivulja i prostor ispod krivulje isto tako ukazuju na klasifikacijske modele čija uspješnost nije mnogo bolja od nasumičnog pogađanja. Dio teze istraživanja se odnosi na korištenje dobivenih klasifikacijskih modela kod opažanja učinaka regulacije uvedene 2017. godine. Razvijeni klasifikacijski modeli su neuspješni u klasifikaciji kategorija rizika tako da se neće koristiti za opažanja učinaka regulacije uvedene 2017. godine. Teza rada za dio klasifikacijskih modela je opovrgnuta budući da korišten set podataka pokazuje nedostatke koji ga čine neprikladnim za razvoj modela klasifikacije.

Razlog slabe uspješnosti klasifikacijskih modela je omjer klasa svakog prediktora i ciljane varijable. Svaki prediktor i ciljana varijabla imaju jednu ili nekoliko klasa koje čine veliku većinu cjelokupne kategorije što uvjetuje razvijanju klasifikacijskih modela koji visoko naginju navedenim klasama. Nadalje, podaci za drugi promatrani period vremena imaju problem pojavljivanja klase nepoznato u velikom broju, uz to što navedeni period sadrži i mnogo manje opažanja. Daljnjim istraživanjem mogli bi se odbaciti manje značajni prediktori, koristiti dodatne metrike uspješnosti i proširiti drugi promatrani period da uključuje naredne godine.

8. LITERATURA

[1] Director UK Airprox Board, „Analysis of Airprox in UK Airspace“, s Interneta, https://www.airproxboard.org.uk/media/ynof2s3h/bluebook33.pdf, rujan 2023.

[2] Cui, Boxuan, „Introduction to DataExplorer“ s Interneta https://cran.r-project.org/web/packages/DataExplorer/vignettes/dataexplorer-intro.html , rujan 2023.

[3] Kuhn, Max, „The caret package“ s Interneta https://topepo.github.io/caret/index.html, rujan 2023.

[4] Kuhn, Max i Wickham Hadley, „tidymodels“ s Interneta https://www.tidymodels.org/, rujan 2023.

[5] Tehnološko sveučilište Wien, „e1071“ s Interneta https://cran.rproject.org/web/packages/e1071/index.html, rujan 2023.

[6] Boehmke, B i Greemwall, B „Hands-On Machine Learning with R“ s Interneta https://bradleyboehmke.github.io/HOML/, rujan 2023.

[7] Garcia Ceja, Enrique „Behavior Analysis with Machine Learning Using R“ s Interneta https://enriquegit.github.io/behavior-free/index.html#welcome, rujan 2023.

[8] James, G, Witten, D, Hastie, T, Tibshirani, R, “An Introduction to Statistical Learning”, Springer, 2023.

[9] Garcia-Portugues, Eduardo, “Notes for Predictive Modeling” s Interneta https://bookdown.org/egarpor/PM-UC3M/, rujan 2023.

9. POPIS OZNAKA I KRATICA

Odbor - odbor za situacije u kojima je narušena sigurnost letjelica Ujedinjenog Kraljevstva

10. SAŽETAK

Rast korištenja bespilotnih letjelica dovodi do uvođenja novih nacionalnih regulacija zračnog prostora 2017. godine. Istraživanje koristi opisnu statističku obradu podataka i klasifikacijske modele zasnovane na statističkom učenju za usporedbu sigurnosti zračnoga prometa u vremenskom periodu od početka 2000. godine do kraja 2016. godine i vremenskom periodu 2017. godine. Cilj istraživanja je pružiti uvid u rezultat novih nacionalnih regulacija zračnog prostora. Interpretacijom rezultata opisne statističke analize uočeno je povećanje sigurnosti zračnog prometa za područje Ujedinjenog Kraljevstva s iznimkom Londona i njemu šire okolice. Utvrđeno je da komercijalni zračni promet snosi daleko najveći pad u sigurnosti leta rastom korištenja bespilotnih letjelica. Razvijeni klasifikacijski modeli nisu mnogo bolji od nasumičnog pogađanja. Uzrok navedene performanse klasifikacijskih modela su neuravnotežen skup podataka.

Ključne riječi: bespilotne letjelice, opisna statistička analiza, klasifikacijski model.

ABSTRACT

The growth in the use of drones leads to the introduction of new national airspace regulations in 2017. The research uses exploratory statistical data processing and classification models based on statistical learning to compare air traffic safety in the time period from the beginning of 2000 to the end of 2016 and the time period of 2017. The goal of the research is to provide insight into the results of the new national airspace regulations. By interpreting the results of the exploratory statistical analysis, an increase in air traffic safety was observed for the territory of the United Kingdom with the exception of London and its wider surroundings. Commercial aviation has been found to bear by far the greatest decline in flight safety with the rise in the use of unmanned aerial vehicles. The classification models developed are not much better than random guesses. The cause of the mentioned performance of the classification models is an unbalanced data set.

Key words: unamnned aerial vehicles, exploratory data analysis, classification model.

11. DODATAK A – INSTALACIJA PROGRAMSKOG OKRUŽENJA R

Prvo je potrebno instalirati R. Exe datoteka za instalaciju se može preuzeti s web stranice The Comprehensive R Archive Network na linku: https://cran.r-project.org/.

Nakon instalacije R-a potrebno je preuzeti i instalirati Rstudio s web stranica posit na linku: https://posit.co/.

DODATAK B – PREUZIMANJE I POKRETANJE KODA

Preuzeti sve dokumente diplomskog rada s figshare web stranice na linku: https://figshare.com/articles/software/Klasifikacija_incidenata_sa_sudjelovanjem_bespilotnih_letjelica_zasnovana_nastrojnom_u_enju_Toni_Cestar_diplomski_rad/24114666.
Pokrenuti Rstudio i otići na file -> open file... i izabrati diplomskiRad.R i Map.rmd. Unutar oba projekta staviti putanju na lokaciju Vaših datoteka. Unutar konzole Rstudija za svaku knjižnicu (library) u kodu pokrenuti: install.packages(„ime_knjižnice“). Projekt Map.rmd koristi Googleov API i potreban je ključ za spajanje na njihov API. Ključ koji je stavljen unutar projekta je isključen jer je projekt stavljen na javno dostupno mjesto i ključ se može maliciozno koristiti. Potreban ključ moguće je generirati pomoću web stranice na linku: https://console.cloud.google.com/. Za ograničenja (en. Restrictions) ključa stavite niti jedno (en. none). Nakon generiranja ključa zamijenite vrijednost register_google(key= “moj_ključ“) sa svojim ključem. Nakon toga možete pokrenuti aplikaciju.