

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 7106

**APLIKACIJA ZA ANALIZU PODATAKA O NATJECANJIMA  
U ORIJENTACIJSKOM SPORTU**

Antun Tišljar

Zagreb, lipanj 2021.

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 7106

**APLIKACIJA ZA ANALIZU PODATAKA O NATJECANJIMA  
U ORIJENTACIJSKOM SPORTU**

Antun Tišljar

Zagreb, lipanj 2021.

## ZAVRŠNI ZADATAK br. 7106

Pristupnik: **Antun Tišljar (0036508971)**

Studij: Računarstvo

Modul: Računarska znanost

Mentor: izv. prof. dr. sc. Alan Jović

Zadatak: **Aplikacija za analizu podataka o natjecanjima u orijentacijskom sportu**

Opis zadatka:

Cilj ovog završnog rada je razvoj aplikacije za analizu podataka za orijentacijski sport. Aplikacija treba omogućiti unos podataka, obradu i analizu te izvještavanje o rezultatima natjecanja u orijentacijskom sportu. Podatke koji će se analizirati student treba prikupiti sam ili ih pronaći slobodno dostupne na webu. Skup podataka treba sadržavati najmanje: podatke o natjecatelju, podatke o stazi i vremenu na svakoj kontrolnoj točki. Pri analizi podataka trebaju se koristiti algoritmi strojnog učenja za obradu vremenskih nizova, a osim dubinske analize s ciljem opisa ponašanja natjecatelja u orijentacijskom sportu, potrebno je provesti i deskriptivnu statističku analizu rezultata. Izbor programskog jezika za implementaciju je proizvoljan.

Rok za predaju rada: 11. lipnja 2021.

*Zahvaljujem se svom mentoru izv. prof. dr. sc. Alanu Joviću  
na pomoći, strpljenju i razumijevanju koje mi je pružio pri izradi ovog rada.*

## Sadržaj

1. Uvod .....	2
2. Orijentacijski sport .....	3
2.1. O orijentacijskom sportu .....	3
2.1.1. Orijentacija u Hrvatskoj.....	3
2.2. Orijentacijsko trčanje.....	4
3. Podatci .....	5
3.1. Obrada podataka .....	5
3.1.1. Priprema podataka .....	5
3.1.2. Deskriptivna statistička analiza .....	6
3.2. Opis podataka .....	6
4. Algoritmi klasifikacije .....	7
4.1. Naivni Bayesov klasifikator .....	8
4.2. K-najbližih susjeda .....	9
4.3. Stabla odluke .....	9
5. Implementacija .....	11
5.1. Poslužiteljska strana .....	11
5.1.1. Weka.....	11
5.2. Klijentska strana i izgled aplikacije.....	13
Zaključak .....	17
Literatura .....	18
Sažetak.....	19
Summary.....	20
Skraćenice.....	21
Privitak .....	22

# 1. Uvod

Orijentacijsko trčanje i orijentacijski sport su pojmovi za koji je malo ljudi čulo, a još manje ih se time bavi. Ta nepoznatost utjecala je na slabu modernizaciju tog sporta što se tiče najnovijih tehnologija. Neke aplikacije i platforme postoje, ali one se u pravilu bave obradom rezultata natjecanja i analizom GPS tragova natjecatelja. Ova aplikacija provjerava prikladnost težine staze za neku kategoriju, a takvih sličnih aplikacija nema.

Cilj ovog završnog rada je stvoriti novu aplikaciju za obradu i analizu podataka pojedinih natjecanja u orijentacijskom sportu. Aplikacija će omogućiti unos i obradu podataka s natjecanja te će pružiti deskriptivnu statističku analizu koja će služiti kao pomoć za organiziranje i planiranje budućih natjecanja. Svojim korisnicima aplikacija će omogućiti bolje razumijevanje u to kako razni faktori utječu na težinu staze. Pomoću algoritama strojnog učenja korisnici aplikacije moći će vidjeti koliko je pojedino natjecanje teže ili lakše u odnosu na sva prethodna. Iz klasifikacije kategorija vidjet će se koliko je koja kategorija pogodna za stazu koju je dobila na natjecanju.

## 2. Orijentacijski sport

### 2.1. O orijentacijskom sportu

Orijentacijski sport zahtijeva navigacijsku vještinu prilikom kretanja po nepoznatom terenu. Natjecatelji na startu dobiju kartu na kojoj je zadan niz kontrolnih točaka koje oni trebaju redom obići u što kraćem vremenu. Pri tome put između svake dvije kontrole svaki natjecatelj bira sam, nastojeći minimizirati ukupno vrijeme provedeno na stazi. Najpopularnija grana orijentacijskog sporta je orijentacijsko trčanje (engl. *FootO*), a još postoji i orijentacija brdskim biciklima (engl. *MTBO*), orijentacijsko skijanje (engl. *SkiO*) i precizna orijentacija (engl. *TrailO*). Prve tri varijacije zahtijevaju i psihičku i fizičku spremu dok je precizna orijentacija otvorena i za ljude s invaliditetom jer se u njoj mjeri točnost čitanja karte i brzina odlučivanja, a ne brzina kretanja.

#### 2.1.1. Orijentacija u Hrvatskoj

Orijentacija se u Hrvatskoj počela razvijati sredinom prošlog stoljeća u okviru Hrvatskog planinarskog saveza. Početkom 1980-ih godina na natjecanjima u Hrvatskoj je počeo prevladavati moderni (nordijski) oblik orijentacijskih natjecanja uz korištenje preciznih petobojnih karata. 2005. godine osnovan je Hrvatski orijentacijski savez.

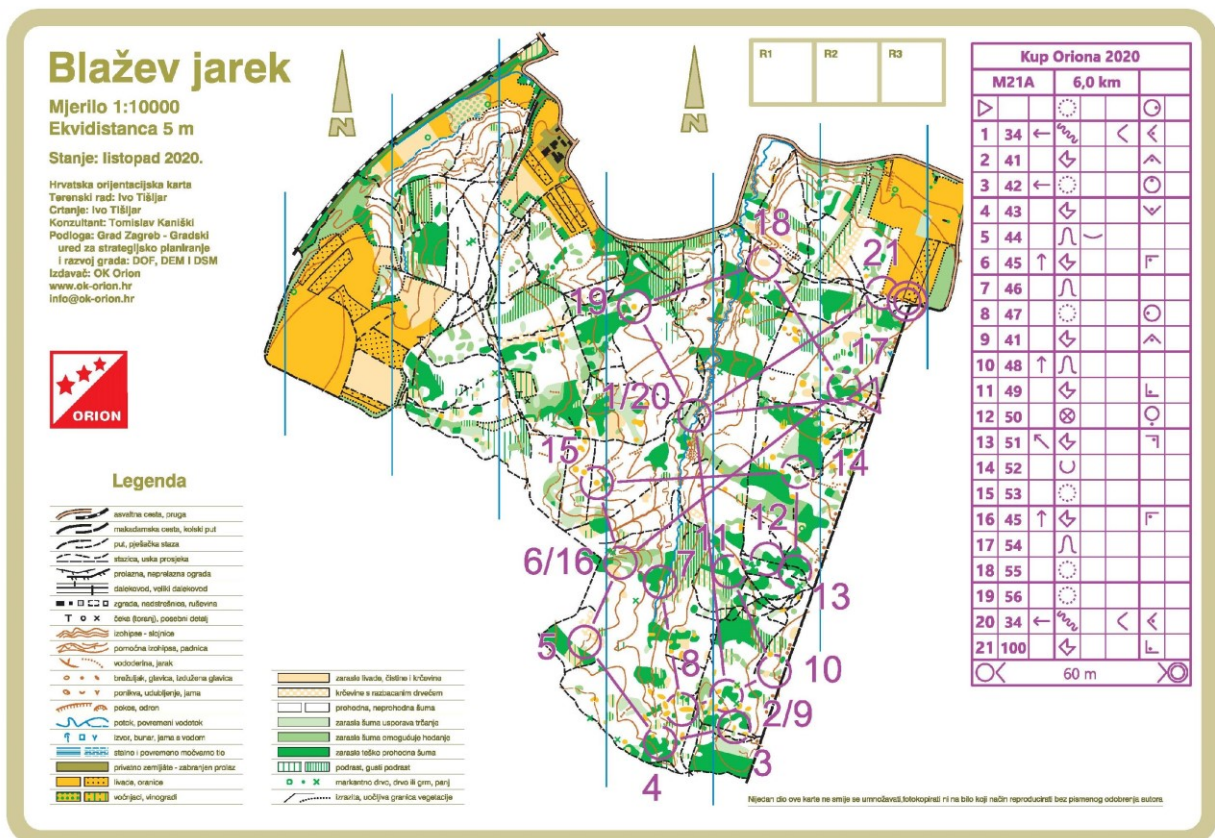
U Hrvatskoj se natjecanja u orijentacijskom trčanju organiziraju po šumama, gradskim parkovima i naseljima, te zgradama škola i fakulteta, pa je tako nekoliko natjecanja u zatvorenom održano i u zgradi FER-a.

Hrvati su najuspješniji u preciznoj orijentaciji u kojoj smo do sada osvojili desetak medalja na Svjetskim i Europskim prvenstvima [1].

U srednjoj Europi na natjecanja dođe i po nekoliko tisuća natjecatelja, a najveća orijentacijska natjecanja na svijetu se organiziraju u Skandinaviji na kojima se okupi i više od 20.000 natjecatelja, dok na natjecanja u orijentacijskom trčanju u Hrvatskoj danas dođe oko 200-300 natjecatelja. Oni se po spolu i dobi dijele u 20-tak kategorija. Ova tvrdnja je važna jer ukazuje na manjak podataka s kojim baratamo.

## 2.2. Orijentacijsko trčanje

Orijentacijsko trčanje je vremenska utrka tijekom koje natjecatelj mora proći sve kontrolne točke u zadanom redoslijedu. Pobjednik je natjecatelj koji to napravi u najkraćem vremenu. Natjecatelj mora brzo i točno pročitati topografske oznake na karti te se odlučiti za najbrži put između kontrolnih točaka. No, najkraća udaljenost nije uvijek i najbrža zbog raznih prepreka u krajoliku. Natjecatelj također mora biti snažan i brz kako bi završio s pobjedničkim vremenom.



Slika 1.1 Karta za kategoriju M21A u Horvatima [2]

Na slici 1.1 vidimo tipičnu kartu za orijentacijsko trčanje. Važni dio karte je opis staze i kontrolnih točaka (na slici desni dio karte obojan je u ružičasto). Iz njega možemo iščitati kategoriju, duljinu staze, uspon, redni broj kontrole te opis svake kontrole.



## **3. Podatci**

Svi podatci skupljeni su u suradnji s Hrvatskim orijentacijskim savezom i orijentacijskim klubom OK Orionom. Hrvatska norma zapisa podataka definirana je u pravilniku za orijentacijska natjecanja. [3]

Zbog slabe popularnosti orijentacijskog sporta dostupni skupovi podataka su maleni. Nadalje, Hrvatski orijentacijski savez je tek prije dvije godine službeno stvorio normu. Prije standarda zapis podataka većine natjecanja ovisio je o tome koji orijentacijski klub organizira natjecanje te su se razlikovali u mnogo čemu, što je uvelike sprječavalo razvoj projekata poput ovoga. Ovakva aplikacija ne bi imala smisla da je se može koristiti za svako treće natjecanje zbog nekompatibilnosti podataka.

Ova aplikacija razvijena je primarno za orijentacijsko trčanje, zato što ne postoji dovoljan broj podataka niti standard za ostale orijentacijske discipline. Iznimka je precizna orijentacija za koju postoje podatci, ali su zbog prirode sporta često nekompatibilni i nenormirani.

Još jedan veliki razlog nedostatka podataka je trenutna epidemiološka situacija zbog koje se kroz skoro čitavu godinu nisu mogla održavati natjecanja.

### **3.1. Obrada podataka**

#### **3.1.1. Priprema podataka**

Najveći problem u pripremi podataka predstavlja često kršenje ili ignoriranje standarda.

Ova aplikacija radi isključivo s hrvatskim orijentacijskim standardom te je prije korištenja nužno sve podatke prebaciti u taj oblik.

### 3.1.2. Deskriptivna statistička analiza

Statistička analiza prikuplja i tumači podatke kako bi se otkrili trendovi. U ovom radu ona se koristi radi procjene težine staza. Promatra se koliko natjecatelja nije završilo stazu ili je bilo diskvalificirano u pojedinoj kategoriji jer je to dobar indikator da sa stazom nešto ne valja.

Nadalje, prikazuju se zanimljive informacije poput koliko je iznosilo najbrže utrčavanje na cilj, kolikom prosječnom brzinom trče pobjednici u svim kategorijama, koji klub je osvojio najviše medalja, itd.

### 3.2. Opis podataka

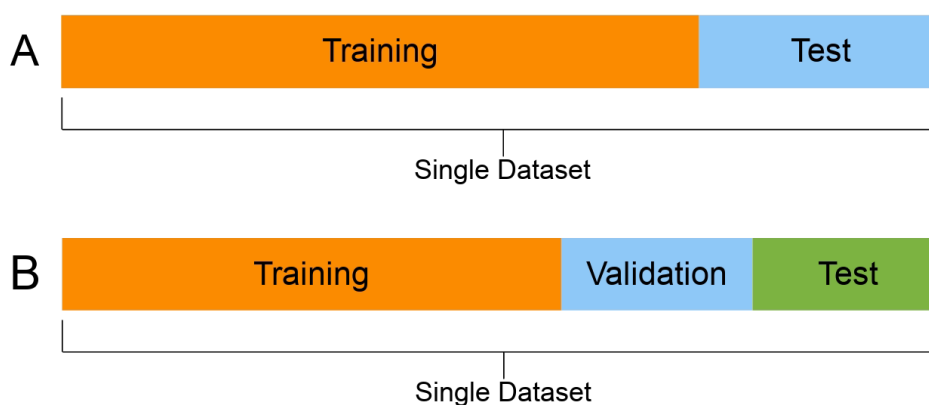
U tablici 2.1 nalaze se podatci koje ćemo koristiti u algoritmima strojnog učenja. Duljina staze, uspon, broj kontrola i kategorija usko su vezani uz pojedinu stazu. Spol je važan jer se kategorije dijele na muške i ženske te su muške kategorije za isto godište u pravilu malo teže od ženskih kategorija. Vrijeme, tip i plasman govore nam kako se pojedinac nosio sa zadanom stazom.

Spol	Muško ili žensko
Vrijeme	Vrijeme na stazi u sekundama
Tip	0 – natjecatelj je završio stazu 1 – natjecatelj nije nastupio 2 – natjecatelj nije završio stazu 3 – natjecatelj je diskvalificiran
Duljina staze	Duljina staze u kilometrima
Uspon	Ukupan uspon na stazi u metrima
Broj kontrola	Broj kontrola na stazi
Plasman	Položaj u poretku
Kategorija	Kategorija staze

Tablica 2.1 Opis najvažnijih podataka [4]

## 4. Algoritmi klasifikacije

Klasifikacija je postupak razumijevanja, prepoznavanja i grupiranja predmeta u unaprijed zadane kategorije ili klase. Algoritmi klasifikacije u strojnom učenju koriste ulazne podatke za predviđanje vjerojatnosti da će sljedeći podaci spadati u jednu od određenih kategorija. Njihov posao jest prepoznavanje uzoraka kako bi se pronašla pravila i zakonitosti koja će se moći primijeniti na budućim skupovima podataka.



Slika 3.1 Primjeri načina za podjelu skupa podataka [5]

Skupovi podataka dijele se na dva (slika 3.1 skup A) ili na tri dijela (slika 3.1 skup B).

Skup za učenje (engl. *training set*) koristi se tijekom procesa učenja, tijekom kojeg se iterativno podešavaju težine parametara. Prati se točnost obrade te se uvodi mjera pogreške.

Skup za provjeru (engl. *validation set*) treba slijediti istu raspodjelu vjerojatnosti uzoraka kao i skup podataka za učenje. Također se prati točnost, no nad ovim skupom se ne uči. Kao što mu ime kaže skup za provjeru se koristi za validaciju skupa za učenje kako ne bi došlo do gubljenja generalizacije. Ukratko on, služi za provjeru generalizacijskih svojstva modela tijekom njegove optimizacije

Skup za testiranje također treba slijediti istu raspodjelu vjerojatnosti uzoraka kao i skup podataka za učenje. On služi za konačnu provjeru te za usporedbu naučenog modela s

drugim modelima. Kako ne bi došlo do problema pretreniranosti u kojem model gubi svojstvo generalizacije, učenje se prekida kada pogreška počne rasti na skupu za provjeru.

## 4.1. Naivni Bayesov klasifikator

Bayesovo pravilo računa vjerojatnost da se podatak nalazi u nekoj kategoriji ili da se ne nalazi u njoj.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Formula 3.4 Bayesova formula

Formula sa slike 3.4 može se i ovako izreći:

Vjerojatnost A, ako je B istinita, jednaka je vjerojatnosti B, ako je A istinita, puta vjerojatnost da je A istinita, podijeljena s vjerojatnošću da je B istina.

Prednost naivnog Bayesovog klasifikatoraje to što je potrebnarelativno malena količina skupa podataka za procjenu značajki važnih za klasifikaciju.

Nedostatak naivnog Bayesovog klasifikatora je pretpostavljanje neovisnosti između svih značajki zbog čega se i naziva „naivnim“. Ta pretpostavka u stvarnosti može biti narušena, ali će model svejedno dobro raditi.

Model mora razlučiti najvjerojatniju hipotezu iz prostora hipoteza za predočene podatke D (engl. *MAP*). Korištenjem Bayesove formule MAP hipoteza računa se na sljedeći način:

$$h_{MAP} = \arg \max_{h_i \in H} P(D | h_i) P(h_i)$$

Formula 3.5: MAP hipoteza

## 4.2. K-najbližih susjeda

K-najbližih susjeda spada pod domenu nadziranog učenja te je jedan od osnovnih algoritama klasifikacije. To je algoritam prepoznavanja koji koristi skup za učenje kako bi u budućim podacima našao  $k$ -najbližih susjeda.

Ovaj algoritama podrazumijeva da se svi primjeri mogu prikazati u Euklidskom prostoru. Najbliži susjedi definiraju se prema Euklidskoj udaljenosti.

## 4.3. Stabla odluke

Stablo odluke funkcionira poput dijagrama toka jer razdvaja podatke u dvije slične klase odjednom i tako od korijena sve do listova gdje klase postaju veoma slične.

Konstrukcija stabla počinje računanjem početne vrijednosti entropije skupa podataka  $D$ . Pripadnost razredu  $C_j$  poznata je za sve primjere te se početna entropija računa prema:

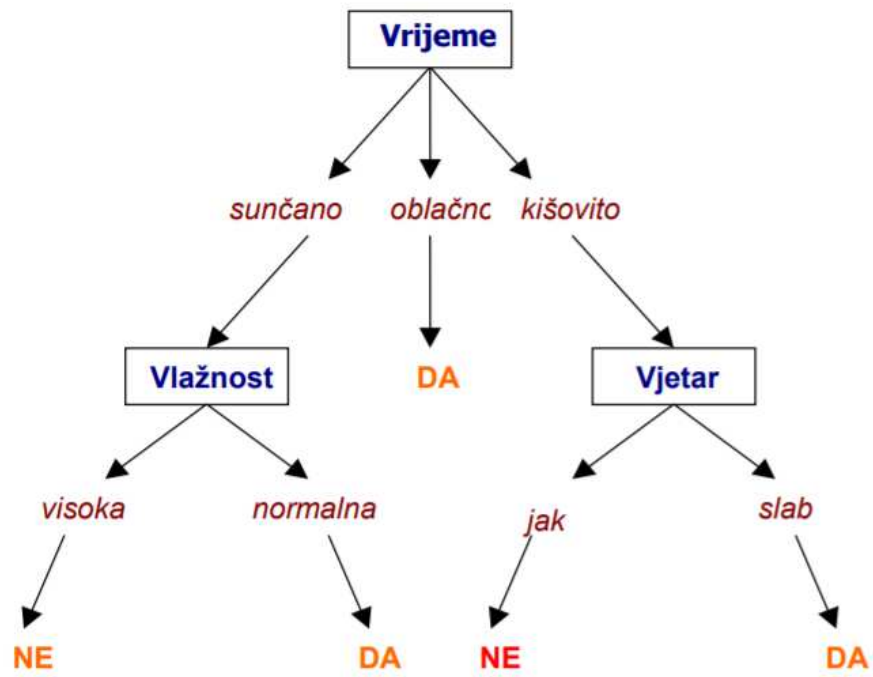
$$\text{Entropija}(D) = \sum_{j=1}^K - \left( \frac{N_j}{N} \right) \log_2 \left( \frac{N_j}{N} \right) = \sum_{j=1}^K - P(C_j) \log_2 P(C_j)$$

Formula 3.6 Računanje početne entropije

Za odabir korijenskog stabla bira se atribut koji ima najveću informacijsku dobit. Informacijska dobit atributa je očekivano smanjenje entropije uzrokovano podjelom primjera za učenje prema tom atributu.

Dalje se gradi sljedeća razina stabla tako da se bira atribut koji će biti čvor u sljedećoj razini. Atribut se bira tako da za sve grane dobijemo maksimalnu informacijsku dobit. Gradnja se nastavlja sve dok svi podskupovi ne budu iz jednog razreda jer je entropija sustava tada jednaka nuli.

Kao zaštitu od prenaučivosti koristi se podrezivanje stabla. Podrezivanje stabla uklanja čvor i pripadno podstablo zamjenjujući ga s listom koji sadrži najčešću vrijednost ciljnog atributa u tom čvoru.



Slika 3.6 Stablo odluke – jutarnja šetnja? DA/NE [6]

## 5. Implementacija

### 5.1. Poslužiteljska strana

Poslužiteljska strana bazira se na Spring Frameworku te na algoritmima strojnog učenja koristi ugrađene Wekine funkcije.

#### 5.1.1. Weka

Weka je alat pisan u Javi koji sadrži razne algoritme strojnog učenja za analizu podataka poput regresije, klasifikacije, grupiranja i slično. U sklopu ove aplikacije radi se klasifikacija.

Weka umjesto CSV datoteka koristi ARFF format. Datoteke u ARFF formatu imaju dva različita dijela. Prvi dio je zaglavlje u kojem se nalaze atributi i njihovi tipovi dok se u drugom dijelu nalaze sami podatci.

```
@RELATION iris

@ATTRIBUTE sepallength NUMERIC
@ATTRIBUTE sepalwidth NUMERIC
@ATTRIBUTE petallength NUMERIC
@ATTRIBUTE petalwidth NUMERIC
@ATTRIBUTE class       {Iris-setosa,Iris-versicolor,Iris-virginica}

1

@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
```

Slika 4.1 Primjer ARFF datoteke [7]

Weka nam nudi laganu pretvorbu datoteka iz CSV formata u ARFF format.

```
private void convertCsvToArff() throws IOException {
    CSVLoader loader = new CSVLoader();
    loader.setSource(file.getInputStream());
    Instances data = loader.getDataSet();

    ArffSaver saver = new ArffSaver();
    saver.setInstances(data);
    saver.setFile(arfffile);
    saver.writeBatch();
}
```

Isječak koda 4.2 Korištenje ugrađenih Wekinih funkcija za pretvorbu .csv u .arff

Prije stvaranja modela potrebno je izbaciti suvišne primjerke i attribute koji nisu važni za klasifikaciju. Nakon učenja i stvaranja modela na skupu za učenje, model se vrednuje na dosad neviđenom skupu za testiranje.

Zbog nedostatka podataka naivni Bayesov klasifikator pokazao se kao najbolji algoritam za klasifikaciju te je redovito davao bolje rezultate nego npr. algoritam stabla odluke. Mana naivnog Bayesovog klasifikatora je što nam nije jednostavno odgovoriti na pitanje zašto je model klasificirao neki primjerak u baš tu klasu, za razliku od stabla odluke u kojima se jasno može pratiti tok klasificiranja.

```
DataSource testSource = new DataSource(arfffile);
Instances test = testSource.getDataSet();
if(test.classIndex() == -1){
    test.setClassIndex(test.numAttributes() - 1);}
test.deleteWithMissing(7);

Remove rm = new Remove();
rm.setAttributeIndices("1,2,3,5,6,7,10,11,12,13,14,15,16");
rm.setInputFormat(train);
train = Filter.useFilter(train, rm);
rm.setInputFormat(test);
test = Filter.useFilter(test, rm);
```



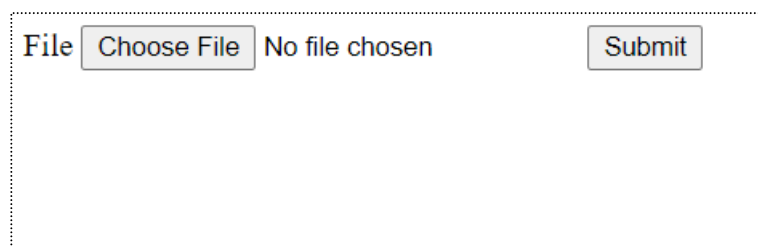
```
NaiveBayes nb = new NaiveBayes();
nb.buildClassifier(train);
Evaluation eval = new Evaluation(train);
eval.evaluateModel(nb, test);
```

Isječak koda 4.3 Filtriranje podataka i stvaranje modela

## 5.2. Klijentska strana i izgled aplikacije

Klijentska strana bazira se na radnom okviru VueJS zasnovanom na JavaScriptu koji sam odabrao zbog lakoće korištenja i generalne raširenosti.

Početni izgled aplikacije može se vidjeti na slici 4.4. Klikom na gumb „Choose File“ otvara se istraživač datoteka te korisnik odabire željeno natjecanje i učitava ga klikom na gumb „Submit“.



Slika 4.4 Početni ekran aplikacije

Nakon obrade na poslužiteljskoj strani podatci se prikazuju kao na slici 4.5. Jedno od čestih pitanja na svakom orijentacijskom natjecanju jest koji klub je bio najuspješniji tj. koji klub je osvojio najviše medalja u svim kategorijama.

No često se u obzir ne uzima broj natjecatelja koji pripada svakom klubu. Svi ti podatci prikazuju se u aplikaciji te je lagano za vidjeti i zaključiti kako je koji klub prošao na natjecanju.

Također se može vidjeti i ukupan broj natjecatelja naspram broju natjecatelja koji su bili diskvalificirani iz raznih razloga.



Slika 4.5 Dio aplikacije sa statističkim podacima

Zbog velike razlike u duljini, količini uspona te drugim svojstvima staze važno je model trenirati i testirati na stazama iz iste kategorije natjecanja što znači da natjecanja s dugim stazama ne uspoređujemo s natjecanjima s kratkim stazama. U protivnome model će dati loše rezultate. Na slikama je prikazano natjecanje na kratke staze.

Na slici 4.6 vide se razni numerički podatci rezultata klasifikacije. Na prvi pogled model nije precizan jer točno klasificirao samo 39% primjeraka. No treba uzeti u obzir da na nekim natjecanjima različite kategorije mogu imati istu stazu. Pogledajmo redak u matrici konfuzije na slici 4.7. Model je 14 primjeraka svrstao u točnu kategoriju M21B no 5 ih je klasificirao kao kategoriju M16. Ispod matrice konfuzije nalaze se podatci o zajedničkim stazama te vidimo da su na ovom natjecanju kategorije M16, M21B i Open Long imale istu stazu što znači da je model bio u pravu.

CATEGORY CLASSIFICATION RESULTS	
Correctly Classified Instances	92 38.4937 %
Incorrectly Classified Instances	147 61.5063 %
Kappa statistic	0.3458
Mean absolute error	0.057
Root mean squared error	0.2097
Relative absolute error	66.4476 %
Root relative squared error	101.1057 %
Total Number of Instances	239

Slika 4.6 Prikaz numeričkih podataka rezultata klasifikacije

U pravilu muške kategorije jednakog godišta kao ženske kategorije imaju fizički zahtjevniju, ali tehnički jednako tešku stazu. Ako je npr. model dio ženskih natjecatelja jedne kategorije klasificirao kao natjecatelje iz muške kategorije istog godišta (bez da oni imaju istu stazu) to pokazuje da staza za tu žensku kategoriju nije adekvatno zadana.

Trenutno je najjača muška kategorija M21A te što je druga kategorija mlađa ili starija od nje to ona ima fizički i tehnički manje zahtjevnu stazu. Za starije kategorije razlika u tehničkoj težini je relativno malena jer tu spadaju natjecatelji koji se dobro znaju orijentirati no zbog godina se ne mogu natjecati na fizički zahtjevnim stazama.

Ako model dio natjecatelja iz kategorije M35 svrsta u kategoriju M45 to pokazuje da je kategorija M35 imala fizički prelaganu stazu. Ako pak dio natjecatelja iz kategorije Ž16 svrsta u Ž18 to pokazuje da je kategorija Ž16 imala fizički pretešku stazu.

```

Confusion matrix
a b c d e f g h i j k l m n o p q r s t u v <-- classified as
3 0 0 0 0 0 0 0 0 0 0 5 2 0 0 0 0 0 0 0 0 0 | a = M12
0 2 0 0 0 0 0 0 0 0 0 0 3 1 3 0 0 0 0 0 0 0 0 | b = M14
0 0 3 0 0 4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | c = M16
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 5 0 0 0 0 0 | d = M20
0 0 0 0 15 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | e = M21A
0 0 5 0 0 14 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | f = M21B
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 15 0 0 0 0 0 | g = M35
0 0 0 0 0 0 0 12 0 0 0 0 0 0 0 0 0 0 0 7 0 0 0 | h = M45
0 0 0 0 0 0 0 5 0 0 0 0 0 0 0 0 0 0 0 8 0 0 0 | i = M55
0 0 0 0 0 0 0 0 7 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | j = M65
0 0 0 0 0 5 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | k = OPEN LONG
4 0 0 0 0 0 0 0 0 0 0 4 2 0 0 0 0 0 0 0 0 0 0 | l = OPEN SHORT
3 0 0 0 0 0 0 0 0 0 0 5 1 0 0 0 0 0 0 0 0 0 0 | m = Ž12
0 6 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 | n = Ž14
0 0 3 0 0 3 0 0 0 0 0 0 0 0 1 0 0 2 0 0 0 0 0 | o = Ž16
0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 | p = Ž20
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 13 0 0 0 0 0 | q = Ž21A
0 0 2 0 0 4 0 0 0 0 0 0 0 0 2 0 0 11 0 0 0 0 0 | r = Ž21B
0 0 0 0 0 0 0 4 0 0 0 0 0 0 0 0 0 0 13 0 0 0 0 | s = Ž35
0 0 0 0 0 0 0 0 18 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | t = Ž45
0 0 0 0 0 0 0 0 1 0 0 0 0 0 3 0 0 2 0 0 0 0 0 | u = Ž55
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 3 0 0 0 0 | v = Ž65

```

SHARED COURSES:

-----  
M12; Ž12; Open Short

-----  
M14; Ž14

-----  
M65; Ž45

-----  
M45; Ž20; Ž35

-----  
Ž55; Ž65

-----  
M20; M35

-----  
M16; M21B; Open Long

-----  
Ž16; Ž21B

Slika 4.5 Matrica konfuzije i zajedničke staze

## Zaključak

Ovaj završni rad nastoji ispuniti rupu koja trenutno postoji u programskoj potpori orijentacijskog sporta stvarajući novu aplikaciju za obradu podataka pojedinih natjecanja. Aplikacija omogućuje unos i obradu podataka s natjecanja te radi deskriptivnu statističku analizu nad podacima. Algoritmi strojnog učenja koriste se za klasifikaciju kategorija iz kojih se vidi koliko je koja staza pogodna za koju kategoriju. Sve to pomaže budućim organizatorima za pripremu natjecanja prikladnih težina.

Glavni i najveći problem s kojim se susreće ova aplikacija jest malen broj natjecatelja koji sa sobom povlači i malen broj dostupnih podataka. Nadalje dio dostupnih podataka nije normiran te to dodatno smanjuje količinu radnih podataka. Malena količina podataka je loša u smislu da ako imamo malen uzorak, veći je rizik da će taj uzorak biti slučajno neobičan. Rad s takvim podacima donosi rezultate koji su često nereprezentativni za stvarnu primjenu. Model strojnog učenja koji nastaje nad tim podacima ima relativno veliku pogrešku te je neprecizan za kategorije natjecanja u kojima je samo nekoliko natjecatelja.

Moguće rješenje za ovaj problem je nabava podataka iz drugih država, no tu opet dolazi do problema, jer Hrvatska ima drugačiju normu zapisa u orijentacijskom trčanju te zbog toga ovo rješenje nije jednostavno primjenjivo u drugim državama.

Nadam se da će ova aplikacija privući nove potencijalne natjecatelje i upoznati ih s čarima ovog fizički i psihički zahtjevnog sporta.

## Literatura

- [1] <https://www.ok-orion.hr/preo/precizna-orijentacija.html>; pristupljeno 11. lipnja 2021.
- [2] Orijentacijski klub Orion – karta, listopad 2020.
- [3] [https://orienteering.hr/docs/pu/2021\\_pravilnik\\_za\\_natjecanja.pdf](https://orienteering.hr/docs/pu/2021_pravilnik_za_natjecanja.pdf); pristupljeno 11. lipnja 2021.
- [4] Orijentacijski klub Orion – rezultati natjecanja, listopad 2020.
- [5] Wikipedia - Training, validation, and test sets, Poveznica: [https://en.wikipedia.org/wiki/Training,\\_validation,\\_and\\_test\\_sets](https://en.wikipedia.org/wiki/Training,_validation,_and_test_sets); pristupljeno 10. lipnja 2021.
- [6] B. D. Bašić, Stabla odluke – slajdovi, UNIZG FER 2010.
- [7] Attribute-Relation File Format (ARFF). Poveznica: <https://www.cs.waikato.ac.nz/ml/weka/arff.html>; pristupljeno 10. lipnja 2021.

# Sažetak

**Naslov:**

Aplikacija za analizu podataka o natjecanjima u orijentacijskom sportu

**Sažetak:**

U ovom završnom radu izrađena je web aplikacija za obradu i analizu podataka orijentacijskih natjecanja. Aplikacija omogućuje unos i obradu podataka te pruža deskriptivnu statističku analizu iz koje se vidi kako koji faktor staze utječe na težinu staze. Također koristi algoritme strojnog učenja za presudu težine pojedine staze te pokazuje koja staza je pogodna za koju kategoriju uzimajući u obzir prethodna natjecanja. Svi ti podatci služe kao pomoć u organiziranju i planiranju budućih natjecanja. Najveći problem s kojim se aplikacija susreće je nedostatak podataka i ne pridržavanje norme pri stvaranju novih podataka.

**Ključne riječi:**

web aplikacija, orijentacijsko natjecanje, strojno učenje, klasifikacija, orijentacijsko trčanje

# Summary

**Title:**

Application for data analysis in orienteering competitions

**Summary:**

In this thesis, a web application for data processing and analysis of orienteering competitions was developed. The application enables data entry and processing and provides a descriptive statistical analysis that shows how much each factor of the track affects the difficulty of the track. It also uses machine learning algorithms to judge the difficulty of an individual track and shows which track is suitable for which category taking into account previous competitions. All this information serves as an aid in organizing and planning future competitions. The biggest problem this application faces is lack of data and non-compliance with standards when creating new data.

**Keywords:**

web application, orienteering competition, machine learning, classification, orienteering



## Skraćenice

FootO	<i>Foot Orienteering</i>	Orijentacijsko trčanje
MTBO	<i>Mountain Bike Orienteering</i>	Orijentacija brdskim biciklima
SkiO	<i>Ski Orienteering</i>	Skijaška orijentacija
TrailO	<i>Trail Orienteering</i>	Precizna orijentacija
ARFF	<i>Attribute-Relation File Format</i>	Format datoteke atribut-veza
CSV	<i>Comma-Separated Values</i>	Vrijednosti odvojene zarezom
WEKA	<i>Waikato Environment for Knowledge Analysis</i>	Waikato okruženje za analizu znanja
MAP	<i>Maximum a posteriori</i>	Maximalna a posteriori vjerojatnost

# Privitak

## **Upute za korištenje programske podrške:**

Aplikacija je jednostavna i intuitivna, pritiskom na gumb „select“ otvara se prozor u kojem se odabire novo natjecanje. Pritiskom na gumb „submit“ aplikacija prikazuje razne podatke o natjecanju.