

# Detekcija epilepsije na osnovu EEG signala

SEMINARSKI RAD IZ PREDMETA BIOMEDICINSKI SIGNALI I SISTEMI

Amila Borančić, 17934/1705

Mahira Buturović, 17760/1708

Sarajevo, 22.07.2020.

# Sadržaj

---

Uvod.....	3
Poglavlje 1.....	4
Epilepsija.....	4
Uzroci.....	5
EEG (elektroencefalogram).....	6
Dijagnoza.....	7
Poglavlje 2.....	9
Machine Learning.....	9
SVM.....	10
Poglavlje 3.....	13
Dataset.....	13
Preprocesiranje za potrebe SVM-a.....	14
Dobijeni model.....	15
Program za klasifikaciju EEG signala pacijenta.....	16
Zaključak.....	19
Reference.....	20
Popis slika.....	22
Listinzi.....	22

# Uvod

---

Epilepsija je neurološki poremećaj od kojeg boluje oko 50 miliona ljudi diljem svijeta. Smatra se četvrtim najčešćim neurološkim poremećajem, i svake godine oko 5 miliona ljudi dobije tu dijagnozu [1]. Glavna karakteristika epilepsije su tzv. *epileptički napadi*. Napadi mogu varirati u svojoj jačini, od toga da su toliko slabi da ih je teško i svrstati u grupu epileptičnih napada, do toga da su toliko jaki da se tijelo pacijenta vidno i veoma jako trza i grči. Epileptični napadi su često okarakterisani trzajima, gubitkom svijesti i osjetila, kao i poteškoće pri kretanju i govoru [2].

U sklopu ovog seminarskog rada se vrši upoznavanje sa EEG-om (skr. *elektroencefalografom*) mozga pacijenata koji pate od epilepsije, kao i od onih koji ne pate, te se vrši njihova usporedba.

U prvom poglavlju se daje detaljniji uvid u epilepsiju kao bolest, njene uzroke i načine dijagnoze, kao i kratki osvrt na EEG i kako izgleda EEG signal epileptičkog napada.

Drugo poglavlje je posvećeno uvodu u tehniku mašinskog učenja (*eng. Machine learning*) koja će se koristiti za detekciju epileptičkih napada. Razjašnjeno je šta je klasifikacija i koja je njena uloga, te je opisan jedan od boljih klasifikatora, poznat pod nazivom *Support Vector Machine*.

Treće poglavlje se bavi implementacijskim detaljima. Prvo je objašnjeno koji dataset je odabran i kako je on prilagođen obradi pomoću SVM klasifikatora, zatim dalje preprocesiranje podataka i konačno, program koji vrši traženu klasifikaciju.

# Poglavlje 1

---

## Epilepsija

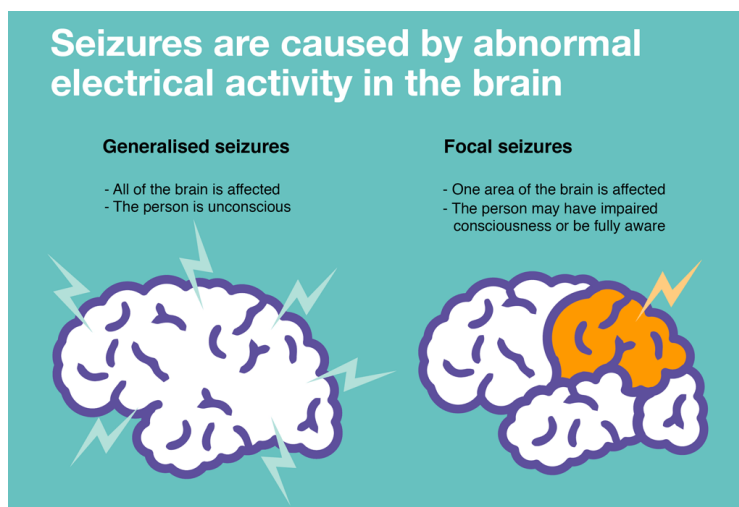
Epilepsija je poremećaj koji spada u veoma česta neurološka oboljenja. Prisutan je i kod domaćih životinja (najviše pasa i mačaka), ne samo kod ljudi [3]. Epilepsija nije zarazna bolest. Ona se može naslijediti i takav oblik epilepsije se naziva *prava* ili *idiopatska* epilepsija. Također, epilepsija se može steći i kao posljedica tumora ili povrede mozga, i takav oblik epilepsije se naziva *sekundarna* ili *simptomska* epilepsija [4].

Epilepsiju karakterišu epizode zvane epileptički napadi, koji se manifestuju kada je pacijent koji pati od epilepsije pod velikim stresom, od pretjerane konzumacije alkohola, pa čak i neki blaži vanjski uticaji kao recimo svjetlo koje titra. Napadi variraju od osobe do osobe u svojoj jačini i simptomima koje ispoljavaju, ali neki od najčešćih su:

- jako grčenje mišića
- gubljenje sposobnosti govora
- oslabljena čula
- mentalne poteškoće
- gubljenje sposobnosti kretanja [5].

Tokom epileptičkog napada, abnormalno visoka električna aktivnost je prisutna u mozgu. U ovisnosti od toga u kojoj regiji mozga se električna aktivnost dešava, razlikuju se sljedeći oblici napada:

- fokalni - jedna polovica mozga - jedna hemisfera, veliki, srednji ili mali mozak
- generalizirani - obje hemisfere mozga [6].



Slika 1. Fokalni i generalizirani napadi, preuzeto sa <https://www.youngepilepsy.org.uk/guide-for-schools/epileptic-seizures/seizure-types/>

Normalna električna aktivnost u mozgu nije sinhrona. Neuron ne radi kao jedna jedinica, nego svaki za sebe. U epileptičkim napadima, određena količina neurona počne da ispoljava abnormalno jaku električnu aktivnost, i to sinhrono, odnosno istovremeno. Fizički, ova električna aktivnost se manifestuje kao trzanje, grčenje mišića ili kao gubitak svijesti, što su zapravo i ranije navedeni simptomi napada. Pored toga, pacijent može početi umišljati da osjeti čudne mirisove, da čuje zvukove koji nisu tu ili slično. Simptomi napada ovise od toga u kojoj regiji mozga su se neuroni aktivirali [7].

Srećom, epilepsija je bolest koja se može veoma efikasno tretirati lijekovima. Nije izlječiva, međutim moguće ju je držati pod kontrolom redovnim uzimanjem antiepileptika, što je dobra vijest za pacijente [8].

## Uzroci

Uzroci epileptičnih napada mogu biti genetski i organski faktori. Epilepsija izazvana genetskim faktorima naziva se primarna, genetska ili genuina epilepsija. Epilepsija izazvana organskim faktorima naziva se sekundarna, simptomatska ili organska epilepsija. Ranije se je smatralo da je daleko najveći procenat epilepsija uvjetovan genetskim faktorima. Statistički podaci pokazuju da ako osoba ima genetsku epilepsiju, onda u njegovoj porodici 3.2% najbližih srodnika također boluje od epilepsije. Međutim, napretkom u dijagnostici, ovaj procenat se sve više smanjuje i sve više se otkrivaju različiti organski faktori u etiologiji napada. Najčešći organski faktori u etiologiji epileptičnih napada su:

1. paranatalni faktori (faktori vezani za period trudnoće, porođaja; razvojne anomalije ploda, hromozomne abnormalnosti itd.)
2. infektivne bolesti (meningitis; TBC, parazitarne, bakterijski, virusni; encephalitis, moždani apcesi)
3. toksični faktori (alkohol, ugljen monoksid, medikamenti)
4. trauma i fizički agensi (akutne povrede mozga, krvarenja u moždanoj ovojnici)
5. cirkulatorni poremećaji (arterioskleroza, tromboza, krvarenja, hipertenzivna encefalopatija)
6. poremećaji metabolizma i nutricije (izmjene materija u tijelu)
7. heredo-familijarne i degenerativne bolesti (multipla skleroza, tuberozna skleroza, cerebelarne degeneracije)
8. tumor mozga [10].

## EEG (elektroencefalogram)

EEG (*elektroencefalogram*) je grafički prikaz električne aktivnosti mozga. Pacijentu se na glavu postave tzv. *scalp* elektrode, te se signal mjeri kao razlika potencijala između dvije postavljene elektrode. EEG ima ogroman historijski značaj zato što je pravilna analiza EEG signala u stanju da otkrije mnoge anomalije i oboljenja na mozgu, od kojih su samo neki:

- oštećenje mozga nakon traume glave
- nadzor anestezije
- testiranje lijekova na moždanu aktivnost
- lokalizacija i otkrivanje epileptičkih napada
- istraživanje mnogih mentalnih poremećaja [9].



Slika 2. Prikaz normalnog EEG signala i signala epileptičkog napada. Preuzeto sa:

[https://www.researchgate.net/publication/324720694\\_EEG\\_monitoring\\_based\\_on\\_automatic\\_detection\\_of\\_seizures\\_and\\_repetitive\\_discharges/figures?lo=1](https://www.researchgate.net/publication/324720694_EEG_monitoring_based_on_automatic_detection_of_seizures_and_repetitive_discharges/figures?lo=1)

Na slici 2 je prikazan primjer EEG signala. Na gornjem dijelu slike se nalazi EEG osobe sa normalnom moždanom aktivnošću, dok je donji dio EEG signal osobe tokom epileptičkog napada. Kao što je i za očekivati na osnovu dosada izloženog, EEG pacijenta tokom napada prikazuje drastičnu promjenu EEG signala. Na samom kraju snimanja, pacijentov EEG signal je podivljao, što ukazuje na to da je napad tada dosegao svoj vrhunac.

## Dijagnoza

Elektroencefalografija je jedinstven i najvrijedniji laboratorijski test u evaluaciji pacijenata sa epilepsijom. U toku EEG snimanja, od pacijenta se traži da diše duboko, pošto duboko disanje (hiperventilacija) može da otkrije određene abnormalnosti. Također od pacijenta se traži da gleda u trepereće svjetlo, jer su neke osobe osjetljive na treperavo svjetlo i to može biti uzrok napada. Kod nekih pacijenata će se izvršiti EEG registracija u spontanom spavanju, a

mali broj će dobiti portabilni magnetofon na kojem će biti izvršeno snimanje dvadesetčetverostnog EEG-a, za vrijeme normalnih svakodnevnih aktivnosti i spavanja [10].

EEG se najčešće snima u *interiktalnom* periodu odnosno periodu između dva epileptička napada. Najrjeđi je *iktalni*, a češći tzv. *postiktalni* EEG (snimak u prvih nekoliko sati nakon epileptične krize). Napadi prekidaju ponašanje pacijenta kakvo je do tada bilo, kratkotrajni su, pošto traju najviše jednu do dvije minute. Dakle, epilepsiju definišemo kao hronični poremećaj koji karakteriziraju ponavljani napadi [10].

Bitno je istaknuti da EEG ne postavlja dijagnozu, već samo pokazuje specifične abnormalnosti koje, u sklopu sa kliničkom slikom te rezultatima ostalih ispitivanja, mogu doprinijeti razjašnjenju epilepsije. Međutim, EEG može biti normalan i kod nekih epileptičnih pacijenata. Normalan EEG ne znači da taj pacijent nema epileptične napade. Zbog toga je često potrebno EEG snimati više puta. Na primjer, “pozitivan snimak” (znači postoji epileptiformna aktivnost u EEG-u) nađe se u prvom snimku kod svega 56% pacijenata. Snimajući više puta pacijente koji su imali normalan EEG, može se otkriti “pozitivan snimak” kod još 26% pacijenata. Znači da otprilike 18% svih pacijenata imaju trajno “negativan snimak”, odnosno normalan EEG snimak [10].

Ispitivanja na ogromnom broju pacijenata pokazuju da ishod EEG snimka (normalan ili abnormalan) zavisi od slijedećih faktora:

1. broj snimanja
2. dob pacijenta
3. tip napada
4. učestalost napada
5. efekat liječenja [10].

Što se tiče broja snimaka, ukoliko je prvi snimak u granicama normale, onda je snimanje potrebno provoditi više puta u toku prve godine. Što su pacijenti stariji, to broj “pozitivnih nalaza” opada, tako da se nakon četrdesete godine pojavljuje abnormalan EEG kod 1/3 pacijenata [10].

Što se tiče samih napada, EEG anomalije se najčešće otkrivaju kod *kompleksne parcijalne epilepsije*<sup>1</sup>, te kod *apsansa*<sup>2</sup>. Pacijenti koji imaju češće napade pokazuju više abnormalnosti od onih koji imaju rijetke napade. U rutinskom EEG snimku budnog stanja često se ne otkriju nikakve anomalije [10].

---

<sup>1</sup> Tip epilepsije koji spada u grupu generaliziranih oblika epilepsije. Pretjerana električna aktivnost se javlja u jednoj hemisferi mozga i utiče na svjesnost pacijenta.

<sup>2</sup> Tip epilepsije koji se najčešće manifestuje kod djece. Električna aktivnost se javlja u obje hemisfere mozga istovremeno. Ovi napadi traju oko 10 sekundi i okarakterisani su mentalnim odsustvom pacijenta.



## Poglavlje 2

---

### Machine learning

EEG signale nije lahko interpretirati. Potrebno je dosta truda i iskustva da bi se razumjele i uočile određene osobine iz ovih signala. Kako bi se to postiglo, EEG signal mora proći fazu *procesiranja*. Procesirani EEG signal se može vizualno interpretirati u cilju detekcije anomalija kao što je npr. napad epilepsije, promjene u mentalnom stanju npr. faze sna i slično. Prilikom procesiranja bilo kojeg EEG signala pomoću ML tehnika, postoje tri bitna koraka:

1. *Pre-procesiranje* predstavlja uklanjanje šumova, normalizaciju uzoraka, uklanjanje nedostajućih vrijednosti ili njihova zamjena prosjekom i dr.
2. *Ekstrakcija osobina* je procesiranje signala u cilju kreiranja deskriptora
3. *Dekodiranje* podrazumijeva korištenje klasifikacijskih/regresijskih modela da bi se osobine EEG signala transformisale u signale visokog nivoa npr. smjer kretanja, afektivna stanja, kognitivna stanja i slično [11].

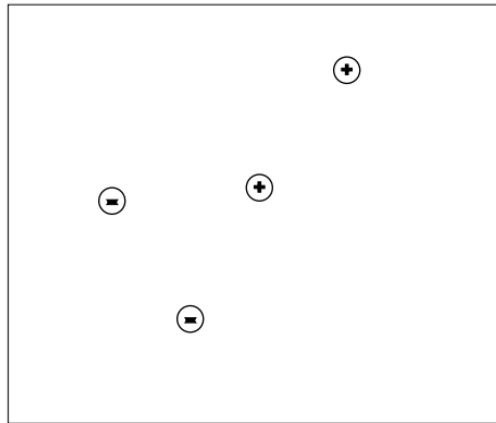
Zadnja faza je veoma bitna i ona predstavlja tzv. *učenje sa supervizijom* (eng. *supervised learning*). Pored učenja sa supervizijom postoji i *učenje bez supervizije* (eng. *unsupervised learning*) u koje spada *klastering* [11].

Što se tiče učenja sa supervizijom, za potrebe ovog seminarskog rada je odabrana *klasifikacija*. Kada se vrši klasifikacija, potrebna su dva dataset-a - *trening set* i *test set*. Trening set se koristi kako bi odabrani klasifikacijski model naučio na osnovu kojih osobina klasificira instance trening seta. Sve instance u trening setu imaju poznatu klasu. Testni set se koristi kako bi se provjerilo koliko je dobar odabrani model, odnosno koliko je dobro istreniran za potrebe klasifikacije. U testnom setu nijedna instanca nema poznatu klasu, već joj klasu dodjeljuje model, te se na osnovu toga koliko je dobra klasifikacija određuje da li se model treba dodatno poboljšavati ili je dobar ovakav kakav je trenutno [11].

Odabran je SVM klasifikator i ideja je da se on iskoristi na dataset u kojem se nalaze snimljeni EEG signali za određeni broj ispitanika, od kojih neki boluju od epilepsije, a neki ne boluju. U idućem poglavlju se nalaze svi detalji implementacije - formalniji opis urađenog, koje okruženje je korišteno, diskusija rezultata i ostalo.

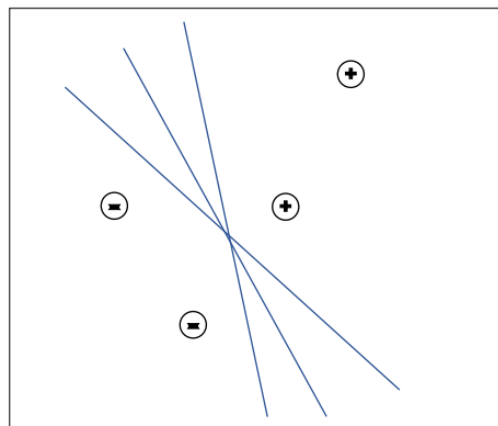
## SVM

Jedna od metoda koja se često koristi u cilju klasifikacije i pri tome daje odlične rezultate je *Support Vector Machine* odnosno SVM skraćeno. SVM je algoritam klasifikacije koji se definira pomoću optimalne hiper-ravni i odabran je za izradu ovog seminarskog rada. U 2D prostoru ova hiper-ravan je linija koja razdvaja instance različitih klasa. Dakle, glavni cilj ove metode je *razdvajanje klasa* pomoću *linije razgraničenja* [13]. Neka je dat prikaz instanci dvije različite klase kao na slici:



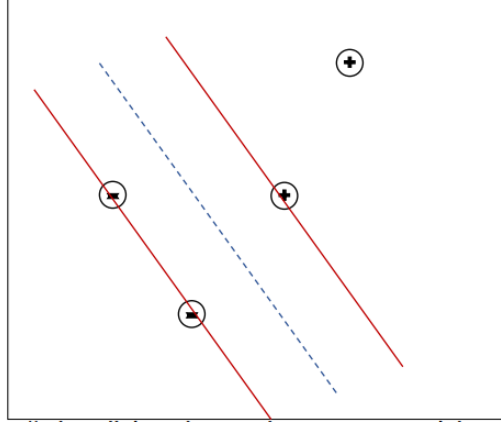
Slika 3. Problem klasifikacije prikazan grafički.

Postavlja se pitanje kako razdvojiti klase + i -. Postoji beskonačno mnogo pravih koje razgraničavaju instance (na slici ispod su prikazane tri):



Slika 4 Razgraničavanje instanci

Potrebno je naći pravu koja najbolje razgraničava prostor. Ideja Vladimira Vapnika, koji je autor SVM metode, jeste da se pronade najšira *ulica* koja razdvaja + i - instance. Linija na sredini te "ulice" je *linija razgraničenja*.



Slika 5 Najbolja linija razgraničenja pomoću SVM

Cilj je da se maksimizira margina koja razdvaja + i - instance, odnosno riješiti problem optimizacije:

$$\arg \min \|\vec{w}\|$$

p.o.

$$y_i(\vec{w} \cdot \vec{x}_i - b) \geq 1 \quad i = 1, 2, \dots, n$$

gdje je  $(x_1, y_1), \dots, (x_n, y_n)$  dataset nad kojim primjenjujemo SVM, pri čemu su vrijednosti  $y_i$  ( $i = 1, 2, \dots, n$ ) jednake 1 ili -1, pri čemu te oznake označavaju kojoj klasi pripada instanca  $x_i$ .  $x_i$  je p-dimenzionalni vektor (broj dimenzija p odgovara broju *kolona* jednog reda dataseta).

Jednačina hiper-ravni se može zapisati u sljedećem obliku:

$$\vec{w} \cdot \vec{x}_i - b = 0$$

gdje je  $\vec{w}$  okomit na ravan, dok veličina  $\frac{b}{\|\vec{w}\|}$  označava udaljenost vektora normale  $\vec{w}$  od koordinatnog početka.

Jednačine hiper-ravni za klase 1 i -1 (koje ujedno predstavljaju ograničenja problema) se izražavaju kao:

$$\vec{w} \cdot \vec{x}_i - b = 1, \text{ pri čemu sve na i iznad ove ravni pripada klasi 1}$$

$$\vec{w} \cdot \vec{x}_i - b = -1, \text{ pri čemu sve na i ispod ove ravni pripada klase -1.}$$

Ova ograničenja se mogu zapisati u alternativnom obliku iz kojeg se izvodi oblik koji je naveden u formulaciji problema optimizacije:

$$\vec{w} \cdot \vec{x}_i - b \geq 1 \text{ za } y_i = 1$$

$$\vec{w} \cdot \vec{x}_i - b \leq -1 \text{ za } y_i = -1$$

Geometrijski gledano, udaljenost između ove dvije ravni iznosi  $\frac{2}{\|\vec{w}\|}$ . Kako je cilj maksimizacija širine "ulice", potrebno je *minimizirati* intenzitet vektora  $\vec{w}$ .

Parametri  $\vec{w}$  i  $b$  koji se dobiju kao rješenje ovog problema predstavljaju parametre SVM klasifikatora koji se dobije nakon treniranja [13].

## Poglavlje 3

---

Tematika ovog seminarskog rada je detekcija epilepsije pomoću ML tehnika. U prethodnom poglavlju je dat kratak opis SVM klasifikatora koji je odabran za potrebe rada, a u poglavlju 1 je dat opis epilepsije kao bolesti, te kako izgleda EEG prilikom epileptičkog napada. U ovom poglavlju se kombinuje dosada opisano kako bi se izgradio dobar klasifikator, a zatim kako bi vršio klasifikaciju EEG signala pacijenata. Rad je izrađen u programskom okruženju RStudio.

### Dataset

Za potrebe ovog rada bila je neophodna neka vrsta baze podataka/dataseta u kojem se nalaze snimljeni EEG signali pacijenata. Dataset koji je pronađen i odabran se nalazi na sljedećem linku: <https://www.kaggle.com/harunshimanto/epileptic-seizure-recognition>.

Ovaj dataset sadrži ukupno 23 sekunde snimljenog EEG signala za 500 ispitanika. Od tih 500 ispitanika, ukupno 100 ispitanika ne pati od epilepsije, dok preostalih 400 pati. Dataset je organizovan na sljedeći način:

- prva kolona predstavlja ID ispitanika
- narednih 178 kolona predstavljaju mjerene vrijednosti EEG signala
- 180. kolona po redu predstavlja dodijeljenu klasu pacijentu koja može uzimati vrijednosti 1, 2, 3, 4 ili 5
- svaki red sadrži 1 sekundu EEG signala, dakle 23 reda pripada jednom ispitaniku.

SVM klasifikator kako je opisan u poglavlju 2 vrši *binarnu* klasifikaciju, odnosno, on može raditi sa *samo dvije klase*. Ovo na prvu izgleda kao problem, pošto je navedeno da odabrani dataset sadrži ukupno pet klasa.

Na linku iznad se također nalazi objašnjenje za svaku od klasa:

1. Registrovan epileptički napad
2. Snimanje EEG signala u regiji gdje je osoba patila od tumora mozga ranije u životu
3. Snimanje EEG signala mimo regije gdje je osoba patila od tumora mozga ranije u životu
4. Snimanje EEG signala kada su pacijentu oči zatvorene
5. Snimanje EEG signala kada su pacijentu oči otvorene

Od svih navedenih klasa, samo je klasa 1 indikator nastanka epileptičkog napada, dok preostale klase predstavljaju zdrav (normalan) EEG signal, i upravo ova informacija omogućava da ipak koristimo binarni SVM klasifikator nad ovim datasetom. Naime, sve što je potrebno

uraditi je preimenovati klase 2, 3, 4 i 5 u *jednu te istu klasu*. Odabrano je da se klase zovu 0 i 1 - 0 označava da pacijent ne pati od epilepsije, a 1 da pati. Klase sa prethodnom oznakom 1 su ostale kakve su i bile, a klase sa prethodnim oznakama 2, 3, 4 i 5 su dobile oznaku 0.

Nakon što su se preimenovalе klase, bilo je potrebno izvršiti još nekoliko prepravki. Naime, dataset sa navedenog linka ima preko 11 000 redova, pa se je dataset morao dosta skratiti. Odabrano je 100 ispitanika koji nemaju epilepsiju i 100 koji je imaju (kako bi dataset bio pravičan), te su sastavljena dva manja pomoćna dataseta. `dataset-prune-script.R` je fajl u kojem se nalazi kod za navedene izmjene nad datasetom i nalazi se u prilogu rada.

## Preprocesiranje za potrebe SVM-a

Nakon što je formiran dataset koji ima prihvatljivu količinu podataka za obradu, bilo je potrebno od njega formirati trening i test set, kao što je objašnjeno u poglavlju 2. U praksi se obično koristi omjer 70:30, pri čemu 70% originalnog dataset-a odlazi na trening, a preostalih 30% na testni set, međutim za potrebe SVM-a u ovom radu bolje se je pokazao omjer 50:50. U sljedećem kratkom listingu se nalazi prikaz koda kojim su izdvojeni setovi za trening i test:

```
no_seizure_data <- read.csv("EpilepticSeizureRecognitionREDUCED.csv", header=T)
seizure_data <- read.csv("EpilepticSeizureRecognitionREDUCED_2.csv", header=T)
#izvadimo svakog drugog ispitanika iz prvog i drugog seta da napravimo trening set
i <- 0
f <- FALSE
train_data <- data.frame()
test_data <- data.frame()
while(i < (nrow(seizure_data)) && i < (nrow(no_seizure_data))){
  dummy <- data.frame(seizure_data[(i+1):(i+23),])
  dummy_2 <- data.frame(no_seizure_data[(i+1):(i+23),])
  if(f){
    train_data <- rbind(train_data, dummy)
    test_data <- rbind(test_data, dummy_2)
  }
  else{
```

```

train_data <- rbind(train_data, dummy_2)

test_data <- rbind(test_data, dummy)

}

i<-i+23

f <- !f

}

train_data["Klasa"] <- lapply(train_data["Klasa"], as.factor)
test_data["Klasa"] <- lapply(test_data["Klasa"], as.factor)

```

Listing 1. Kod za formiranje trening i test dataseta.

Nakon što su formirani trening i test set, idući korak je normalizacija podataka. Dataset se normalizuje iz razloga što se često veličine u datasetovima mnogo rasipaju po veličini, što za mnoge (mada ne i sve) algoritme klasifikacije pravi probleme i rezultira lošijim modelom [14]. Postoji više oblika normalizacije, a ona koja je upotrijebljena u ovom radu je min-max normalizacija [15] koja skalira podatke na opseg od 0 do 1. Za svrhu normalizacije pripremljena je kratka funkcija u R-u:

```

normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x)))
}

```

Listing 2. Kod za normalizaciju podataka.

## Dobijeni model

Klasifikacijski modeli se na razne načine provjeravaju sa svrhom da se utvrdi koliko su dobri, tačni, pouzdani i slično. Konfuzijska matrica je tabela koju nudi programski jezik R a koja u sebi sadrži razne parametre na osnovu kojih se mjeri koliko je neki model dobar. Na slici ispod je prikazana konfuzijska matrica za model koji je dobijen nakon što su se izvršile sve ranije opisane izmjene nad datasetom:

```

Confusion Matrix and Statistics

          Reference
Prediction 0    1
          0 1149    0
          1    1 1150

          Accuracy : 0.9996
          95% CI : (0.9976, 1)
    No Information Rate : 0.5
    P-Value [Acc > NIR] : <2e-16

          Kappa : 0.9991

    McNemar's Test P-Value : 1

          Sensitivity : 0.9991
          Specificity : 1.0000
    Pos Pred Value : 1.0000
    Neg Pred Value : 0.9991
          Prevalence : 0.5000
    Detection Rate : 0.4996
    Detection Prevalence : 0.4996
    Balanced Accuracy : 0.9996

    'Positive' Class : 0

```

Slika 6. Konfuzijska matrica

Prikazana konfuzijska matrica je izračunata na osnovu normaliziranih vrijednosti podataka. Ranije je rečeno da je normalizacija itekako korisna kada je potrebno povećati tačnost klasifikatora, međutim za ovaj dataset normalizacija *nije* dala nikakve značajne promjene. Jedan od razloga zašto normalizacija može da ne utiče na tačnost klasifikatora je ukoliko se instance ne rasipaju mnogo po veličini. Drugi validan razlog je da normalizacija uveliko pomaže kod datasetova čije su varijable heterogene, odnosno čije se varijable razlikuju po tipu [16]. Dataset korišten u ovom radu je homogeni dataset - sve njegove varijable su numeričkog tipa i predstavljaju amplitudu mjerenog signala u jedinici vremena.

Konfuzijska matrica nudi razne informacije. Matrica upoređuje stvarne vrijednosti klasa sa onim koje su predviđene modelom mašinskog učenja. Za binarnu klasifikaciju, kao u našem slučaju, matrica će imati format 2x2 i izgledat će kao na slici:



		ACTUAL VALUES	
		POSITIVE	NEGATIVE
PREDICTED VALUES	POSITIVE	TP	FP
	NEGATIVE	FN	TN

Slika 7. Izgled konfuzijske matrice dimenzija 2x2

Kolone matrice odgovaraju stvarnim vrijednostima, dok redovi odgovaraju predviđenim. Te vrijednosti mogu biti pozitivne ili negativne. Kako bi se mogla ispravno tumačiti ova matrica, potrebno je razumjeti pojmove TP, FP, TN i FN.

*True positive (TP)* znači da je stvarna vrijednost pozitivna i model je predvidio pozitivnu vrijednost.

*False positive (FP)*, odnosno greška tipa 1, znači da je stvarna vrijednost modela negativna, ali je model predvidio pozitivnu vrijednost.

*True negative (TN)* znači da je stvarna vrijednost negativna i da je model predvidio negativnu vrijednost.

*False negative (FN)*, odnosno greška tipa 2, znači da je stvarna vrijednost pozitivna, ali je model predvidio negativnu vrijednost.

Poznavanje ovih vrijednosti je bitno, jer se iz njih mogu odrediti karakteristike performansi modela kao što su: *accuracy* (tačnost), *precision* (preciznost), *recall* (opoziv), *sensitivity* (osjetljivost) i mnoge druge.

Sada se može preći na tumačenje matrice prikazane na slici 6. Na vrhu matrice je prikazan broj tačno i pogrešno svrstanih instance odakle vidimo da su sve instance sa klasom 0 (odnosno, "pacijent nema epilepsiju") korektno svrstane, da su skoro pa sve instance sa klasom 1 također korektno svrstane, te da samo jedna nije dobro svrstana. Za 1 od 200 ispitanika, ovaj model je pogrešno zaključio da ispitanik ima epilepsiju a zapravo je nema. U skladu sa prethodnim objašnjenjem konfuzijske matrice, moglo bi se zaključiti da je 1149 vrijednosti TP, greške tipa 1 odnosno FP nemamo, jedna vrijednost je FN tj. greška tipa 2, dok je 1150 instanci TN.

Očito je da model ima veoma visoku tačnost (*accuracy* iznosi čak 0.9996). Osjetljivost i specifičnost mjere koliko dobro model radi za "negativnu" i "pozitivnu" klasu i ovi parametri su veoma bliski 1, što označava skoro pa 100% tačnost [17].

Na osnovu dosada izloženog jasno je da se radi o veoma dobro istreniranom modelu.

## Program za klasifikaciju EEG signala pacijenta

Nakon što je dataset adekvatno modificiran i nakon što su podaci ispravo preprocesirani, kreiran je kratki program koji vrši klasifikaciju EEG signala pacijenta.

```

user_input<-function(){
  rb <- readline(prompt="Unesite redni broj pacijenta (od 1 do 100): ");
  rb<-as.numeric(rb)
  if(rb>100 || rb<0){
    print("Broj koji ste unijeli nije u opsegu od 1 do 100.");
  }
  else{
    epilepsy_detection(rb);
  }
}

epilepsy_detection<-function(rb){
  tune.out <- tune(svm,Klasa~.,data=nm_train, kernel ="linear",
ranges=list(cost=c(0.01,0.1,1,10)))
  svm_model <- tune.out$best.model
  svm_predict <- predict(svm_model, nm_test)
  svm_model <- tune.out$best.model
  svm_predict<-as.data.frame(svm_predict)
  i<-23*(rb-1)+1
  ima<-0
  nema<-0
  k<-1
  while(k<=23){
    if(svm_predict[i,1]==0){
      nema<-nema+1
    }else{
      ima<-ima+1;
    }
    i<-i+1;
    k<-k+1;
  }
}

```

```

    }
    if(ima>nema){
        print("Pacijent ima epilepsiju.")
    }else if(ima<nema){
        print("Pacijent nema epilepsiju.")
    }else{
        print("Nije moguće pouzdano utvrditi da li pacijent ima epilepsiju ili ne.")
    }
}

user_input()

```

Listing 3. Kod za program koji vrši klasifikaciju EEG pacijenta.

Slijedi kratko objašnjenje koda priloženog u listingu 3. Prvo je kreirana pomoćna funkcija nazvana `user_input` čiji je zadatak da od korisnika traži unos rednog broja pacijenta. Redni broj pacijenta je u opsegu od 1 do 100 uključivo, jer se toliko pacijenata nalazi u test datasetu. Dodana je i validacija kako bi se spriječilo krahiranje programa.

Nakon toga, slijedi definicija funkcije `epilepsy_detection` koja se poziva iz tijela funkcije `user_input`. Pomoću funkcije `tune` (koja dolazi ugrađena u jezik R) vršimo treniranje svm klasifikatora. Prvi parametar su kolone na osnovu kojih vršimo klasifikaciju (sve osim zadnje kolone koja predstavlja naziv klase), drugi je dataset koji koristimo za treniranje, treći je vrsta SVM modela (linearni u našem slučaju) i posljednji parametar je cost vektor koji govori SVM-u koliko se smije previjati oko podataka (iako se radi o linearnom kernelu, ponekad nije moguće povući u potpunosti pravu liniju razgraničenja pa je korisno navesti ovaj parametar).

Potrebno je sačekati neko vrijeme da se model istrenira pošto podataka ima mnogo. Nakon faze treniranja slijedi predikcija. Pomoću ugrađene funkcije `predict` vrši se predikcija klase za EEG signal odabranog pacijenta, te se nakon toga rezultat predikcije ispisuje na ekran. Moguća su tri scenarija: pacijent ima epilepsiju, pacijent nema epilepsiju i nije moguće utvrditi da li pacijent ima epilepsiju.

## Zaključak

---

Epilepsija je bolest od koje boluje mnogo ljudi diljem svijeta. Napretci u medicini i procesiranju biomedicinskih signala su uveliko doprinijeli razumijevanju epilepsije kao bolesti i na tom polju se još uvijek uveliko radi. Primjena mašinskog učenja na detekciju epilepsije iz EEG signala je od velike koristi, s obzirom da je dobro istreniran klasifikacijski model u stanju uočiti trendove i karakteristike koje bi promakle kada bi se signal analizirao golim okom. Predviđanje da li pacijent pati od epilepsije ili ima šanse da pati od nje može se vršiti i na osnovu mješovitih datasetova (koji imaju druge parametre vezane za stil života kao recimo konzumacija alkohola, nasljedne bolesti i dr.) a ne samo na osnovu amplitude EEG signala kao što je rađeno u ovom radu. Normalizacija podataka dataseta može, ali ne mora nužno unaprijediti kvalitet odabranog klasifikatora, što je također demonstrirano. Razvoj dobrog modela nije nimalo lak zadatak i uvijek postoji prostora za poboljšanje i napredak.

## Reference

---

[1] Epilepsija kao neurološki poremećaj, World Health Organization:

<https://www.who.int/news-room/fact-sheets/detail/epilepsy#:~:text=The%20estimated%20proportion%20of%20the,diagnosed%20with%20epilepsy%20each%20year>. [Posljednji pristup: 09.06.2020.]

[2] Epileptički napadi, Wikipedia:

<https://hr.wikipedia.org/wiki/Epilepsija> [Posljednji pristup: 09.06.2020.]

[3, 4] T. Božić, D. Gvozdić, M. Kovačević-Filipović, I. B. Jovanović, J. Stevanović, S. Nikolić, D. Kirovski, M. Jerkić, A. Radovanović. A. Epštajn, N. Andrić, M. Robić, Z. Faixova, Š. Faix, Z. Ivanović, D. Trailoviš, Ž. Maličević, "Patološka fiziologija domaćih životinja", Univerzitet u Beogradu, Beograd 2012.

[5] Simptomi napada:

<https://en.wikipedia.org/wiki/Epilepsy> [Posljednji pristup: 09.06.2020.]

[6] Podjele napada:

[https://en.wikipedia.org/wiki/Epileptic\\_seizure](https://en.wikipedia.org/wiki/Epileptic_seizure) [Posljednji pristup: 09.06.2020.]

[7] Kako epileptički napad izgleda

<https://www.youtube.com/watch?v=uwB8AZ2vK1g> [Posljednji pristup: 09.06.2020.]

[8] Da li je epilepsija izlječiva:

<https://www.healthline.com/health/epilepsy#:~:text=There's%20no%20cure%20for%20epilepsy,with%20medications%20and%20other%20strategies>. [Posljednji pristup: 09.06.2020.]

[9] D. Bošković, O. Lepara, "Biomedicinski signali i sistemi", Univerzitet Sarajevo, Sarejvo, 2018.

[10] Dželaludin Kantardžić i saradnici, "Epilepsija-znanjem protiv straha", Sarajevo, 1989.

[11] Why do we need machine learning for EEG data: <https://www.bitbrain.com/blog/ai-eeg-data-processing> [Posljednji pristup: 15.07.2020.]

[12] Why and how to Cross Validate a Model: <https://towardsdatascience.com/why-and-how-to-cross-validate-a-model-d6424b45261f> [Posljednji pristup: 20.07.2020.]

[13] SVM: [https://en.wikipedia.org/wiki/Support\\_vector\\_machine](https://en.wikipedia.org/wiki/Support_vector_machine) [Posljednji pristup: 20.07.2020.]

[14] Normalizacija podataka i zašto je važna: <https://medium.com/@urvashilluniya/why-data-normalization-is-necessary-for-machine-learning-models-681b65a05029>

[Posljednji pristup: 22.07.2020.]

[15] Oblici normalizacije: <https://www.analyticsvidhya.com/blog/2020/04/feature-scaling-machine-learning-normalization-standardization/> [Posljednji pristup: 22.07.2020.]

[16] Normalizacija: <https://www.quora.com/When-should-you-normalize-a-dataset?fbclid=IwAR0n117ti4-WQYh5vYXdHbhT4ZY8keAjNGh-qZWAOAAQe183ojOtzeXPdk> [Posljednji pristup: 28.07.2020.]

[17] Everything you should know about confusion matrix for machine learning:

<https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/>. [Posljednji pristup: 22.07.2020]

## Popis slika

---

Slika 1 Fokalni i generalizirani napadi.....	5
Slika 2 Prikaz normalnog EEG signala i signala epileptičkog napada.....	7
Slika 3 Problem klasifikacije prikazan grafički.....	10
Slika 4 Razgraničavanje instanci.....	10
Slika 5 Najbolja linija razgraničenja pomoću SVM.....	11
Slika 6 Konfuzijska matrica.....	16
Slika 7 Konfuzijska matrica.....	16

## Listinzi

---

Listing 1 Kod za formiranje trening i test dataseta.....	15
Listing 2 Kod za normalizaciju podataka.....	15
Listing 3 Kod za program koji vrši klasifikaciju EEG pacijenta.....	18