****

**Univerzitet u Kragujevcu – Master program Veštačka inteligencija**

**Detekcija Parkinsonove bolesti korišćenjem veštačke inteligencije**

*Seminarski rad iz predmeta Veštačka inteligencija u medicini*

Student: 4vi/2023, Anđela Stanojević

Mentor: prof Arso Vukićević

**Summary:**

This seminar paper explores the application of artificial intelligence (AI) in the diagnosis of Parkinson's disease (PD), a neurodegenerative disorder that affects motor functions. It examines how different classification algorithms can be utilized for the early detection of disease based on the analysis of patients' voice data. The paper includes an analysis of the dataset, the implementation of algorithms such as logistic regression, KNN, and SVM, and the evaluation of their performance in predicting disease status. The results demonstrate the high potential of AI in the early detection of PD, which can significantly contribute to the advancement of medical diagnostics.

**Key Words:**

Artificial intelligence, Parkinson's disease, machine learning, classification algorithms, SVM, KNN, early diagnosis, voice data, dataset.

**Rezime:**

Ovaj seminarski rad bavi se primenom veštačke inteligencije (VI) u dijagnostici Parkinsonove bolesti (PB), neurodegenerativnog poremećaja koji utiče na motoričke funkcije. Istražuje se kako se različiti klasifikacioni algoritmi mogu koristiti za ranu detekciju bolesti na osnovu analize glasovnih podataka pacijenata. Rad uključuje analizu dataseta, implementaciju algoritama kao što su logistička regresija, KNN i SVM, i evaluaciju njihovih performansi u predikciji statusa bolesti. Rezultati pokazuju visok potencijal VI u ranom otkrivanju PB, što može značajno doprineti unapređenju medicinske dijagnostike.

**Ključne reči:**

Veštačka inteligencija, Parkinsonova bolest, mašinsko učenje, klasifikacioni algoritmi, SVM, KNN, rana dijagnostika, glasovni podaci, dataset

**SADŽAJ**

1. **Uvod**
   1. **Veštačka inteligencija**
   2. **Veštačka inteligencija u medicinskoj dijagnostici**
2. **Parkinsonova bolest** 
   1. **Klinička slika osobe sa PB**
   2. **Dijagnostika PB u relnom životu**
   3. **Izgled osobe sa PB i prvi simptomi**
   4. **Lečenje PB**
3. **Postavka problema**
4. **Dataset**
   1. **Opis dataseta**
   2. **Dobijanje medicinskih parametara iz dataseta**
5. **Klasifikacinoni algoritmi**
   1. **Logistička regresija**
   2. **KNN**
   3. **SVM**
   4. **Perceptron**
6. **Implementacija**
   1. **Prikaz podataka datseta**
   2. **Treniranje modela**
   3. **Rezultati**
   4. **Provera**
7. **Zaključak**
8. **Literatura**
9. **Uvod**

U savremenom dobu, tehnologija postaje ključni deo u svakom aspektu društva, a njen uticaj je posebno značajan u oblasti medicine. Razvoj računarskih nauka i veštačke inteligencije omogućio je primenu naprednih metoda u analizi medicinskih podataka, dijagnostici bolesti i personalizaciji tretmana. Jedna od oblasti gde veštačka inteligencija pokazuje značajan potencijal je rana detekcija i dijagnostika neurodegenerativnih bolesti, među kojima je Parkinsonova bolest, što ćemo detaljno pokazati u ovom radu.

Parkinsonova bolest je progresivni neurološki poremećaj koji utiče na motoričke funkcije, a njen rani stadijum često prolazi nezapaženo, što odlaže početak terapije. Tradicionalne metode dijagnostike, koje se oslanjaju na klinički pregled, često ne uspevaju da identifikuju bolest u njenim najranijim fazama. S obzirom na to da rana intervencija može značajno poboljšati kvalitet života pacijenata, postoji sve veći interes za razvojem novih metoda koje bi omogućile bržu i precizniju dijagnozu.

U ovom seminarskom radu, fokusiraćemo se na primenu veštačke inteligencije u detekciji Parkinsonove bolesti. Kroz analizu medicinskih podataka, posebno onih dobijenih iz govora pacijenata, istražićemo kako se VI može koristiti za prepoznavanje ranih znakova bolesti. Cilj ovog rada je da prikaže potencijal veštačke inteligencije u unapređenju dijagnostičkih procedura i omogući bolje razumevanje njenog doprinosa medicinskoj praksi.

* 1. **Veštačka inteligencija**

Veštačka inteligencija (VI) predstavlja oblast računarstva koja se bavi razvojem sistema sposobnih za obavljanje zadataka koji zahtevaju ljudsku inteligenciju. Ovi zadaci uključuju prepoznavanje obrazaca, rešavanje problema, učenje na osnovu iskustva, planiranje, donošenje odluka i razumevanje prirodnog jezika. VI obuhvata širok spektar tehnika i pristupa u šta spada mašinsko učenje, duboko učenje, obrada prirodnog jezika, računarski vid i ekspertski sistemi.

U osnovi VI nalazi se sposobnost mašinskog učenja koja omogućava računarima da uče iz podataka, prepoznaju obrasce i donose predviđanja ili odluke sa sve većom preciznošću. Mašinsko učenje obuhvata različite metode, kao što su nadgledano, nenadgledano i učenje sa podsticajem, od kojih svaka koristi različite pristupe za obradu podataka i optimizaciju performansi. Na primer, algoritmi nadgledanog učenja koriste obeležene podatke za obuku modela, dok nenadgledano učenje analizira neobeležene podatke radi identifikacije skrivenih struktura ili grupa u podacima.

Duboko učenje, koje je podskup mašinskog učenja, koristi više slojeva neuronskih mreža za obradu podataka i učenje reprezentacija na različitim nivoima apstrakcije. Ovaj pristup omogućava računarima da izvode kompleksne zadatke poput prepoznavanja lica, automatskog prevođenja jezika i analize medicinskih slika. Napredak u dubokom učenju, zajedno sa rastućom dostupnošću velikih količina podataka i računske snage, značajno je unapredio sposobnosti VI u različitim domenima.

Veštačka inteligencija se danas koristi u mnogim industrijama, uključujući finansije, transport, proizvodnju, zabavu i obrazovanje. Njena primena omogućava optimizaciju procesa, smanjenje troškova, poboljšanje kvaliteta proizvoda i usluga, i kreiranje novih inovativnih rešenja. Pa se tako VI u kontekstu medicine pokazala kao izuzetno korisna alatka, koja omogućava naprednu analizu podataka, personalizovanu negu pacijenata i poboljšanje dijagnostičkih i terapijskih postupaka.

* 1. **Veštačka inteligencija u medicinskoj dijagnostici**

Primena veštačke inteligencije u medicinskoj dijagnostici predstavlja jedan od najuzbudljivijih i najperspektivnijih aspekata savremene medicine. VI omogućava analizu velikih količina medicinskih podataka, kao što su slike, genomski podaci, elektronski zdravstveni zapisi, i podaci sa nosivih uređaja, kako bi se identifikovali obrasci i anomalije koje mogu ukazivati na prisustvo bolesti ili rizik od njenog razvoja.

Jedan od ključnih doprinosa VI u medicini je u oblasti dijagnostike, gde se koriste algoritmi za automatsko prepoznavanje obrazaca u slikama dobijenim iz različitih modaliteta, kao što su rendgen, magnetna rezonanca (MRI), kompjuterska tomografija (CT) i ultrazvuk. Na primer, algoritmi zasnovani na dubokom učenju mogu prepoznati tumore, lezije, ili druge abnormalnosti na medicinskim slikama sa visokim stepenom tačnosti, često u ranim fazama bolesti, kada je tretman najefikasniji. Ova tehnologija se koristi u ranoj detekciji karcinoma dojke, pluća, prostate i drugih vrsta raka.

VI takođe nalazi primenu u analizi genetskih podataka, gde se koristi za identifikaciju genetskih mutacija i varijanti povezanih sa različitim bolestima. Na primer, VI algoritmi mogu analizirati sekvence DNK kako bi se identifikovale mutacije koje povećavaju rizik od određenih naslednih bolesti, kao što su cistična fibroza ili hemofilija. Ovi podaci mogu biti korišćeni za personalizaciju terapija, što je poznato kao precizna medicina, omogućavajući lekarima da prilagode tretman svakom pacijentu na osnovu njegovih genetskih karakteristika.

Još jedno područje gde VI doprinosi medicinskoj dijagnostici je analiza elektronskih zdravstvenih zapisa (EZZ). EZZ sadrže ogromne količine informacija o pacijentima, uključujući istoriju bolesti, laboratorijske rezultate, slike, terapijske režime i napomene lekara. VI može analizirati ove podatke kako bi identifikovala obrasce koji mogu ukazivati na određene dijagnoze ili rizik od budućih zdravstvenih problema. Na primer, VI algoritmi mogu predvideti rizik od srčanog udara na osnovu analize podataka o krvnom pritisku, holesterolu, i načinu života pacijenta.

VI se koristi i za razvoj tzv. "decision support" sistema, koji pomažu lekarima u donošenju odluka u vezi sa dijagnostikom i lečenjem. Ovi sistemi mogu predlagati dijagnoze na osnovu unetih simptoma, ili preporučivati odgovarajuće terapije na osnovu najnovijih medicinskih istraživanja i smernica. Ovakvi alati ne samo da poboljšavaju tačnost dijagnoza, već i omogućavaju lekarima da se fokusiraju na kompleksnije aspekte nege pacijenata.

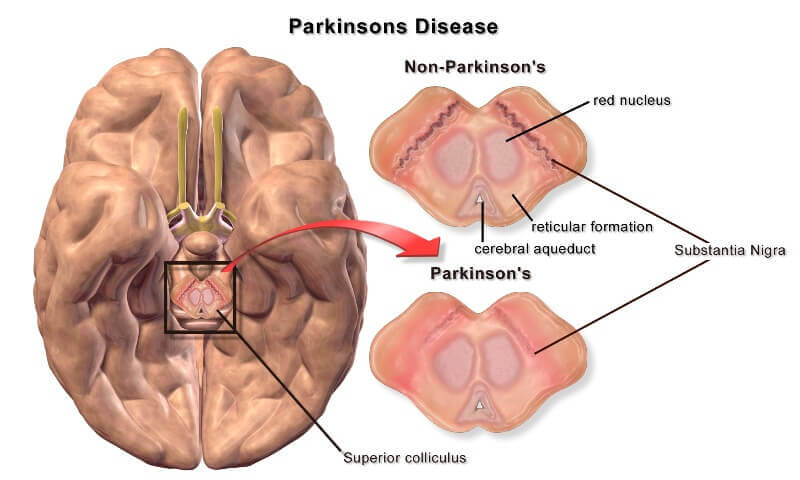
Pored dijagnostike, VI se koristi i za praćenje toka bolesti i odgovor na terapiju, omogućavajući lekarima da prilagode tretmane u realnom vremenu. Na primer, nosivi uređaji opremljeni senzorima mogu kontinuirano pratiti vitalne parametre pacijenta, kao što su srčani ritam, nivo kiseonika u krvi, i krvni pritisak. Podaci sa ovih uređaja mogu biti analizirani pomoću VI algoritama kako bi se detektovale promene koje ukazuju na pogoršanje zdravstvenog stanja pacijenta, što omogućava pravovremenu intervenciju.

U kontekstu Parkinsonove bolesti, primena VI može pomoći u ranoj dijagnostici kroz analizu akustičnih karakteristika glasa ili finih motoričkih sposobnosti, koje mogu biti promenjene u ranim fazama bolesti. Ovime se omogućava ranija intervencija i bolja kontrola simptoma, što može značajno poboljšati kvalitet života pacijenata.

Veštačka inteligencija u medicinskoj dijagnostici predstavlja revolucionarni alat koji može poboljšati tačnost i brzinu dijagnoze, omogućiti personalizovan pristup lečenju i unaprediti ukupni kvalitet zdravstvene nege. Kako tehnologija napreduje, očekuje se da će VI postati sve više uključena u svakodnevnu kliničku praksu, omogućavajući zdravstvenim radnicima da pružaju najbolju moguću negu svojim pacijentima.

1. **Parkinsonova bolest**

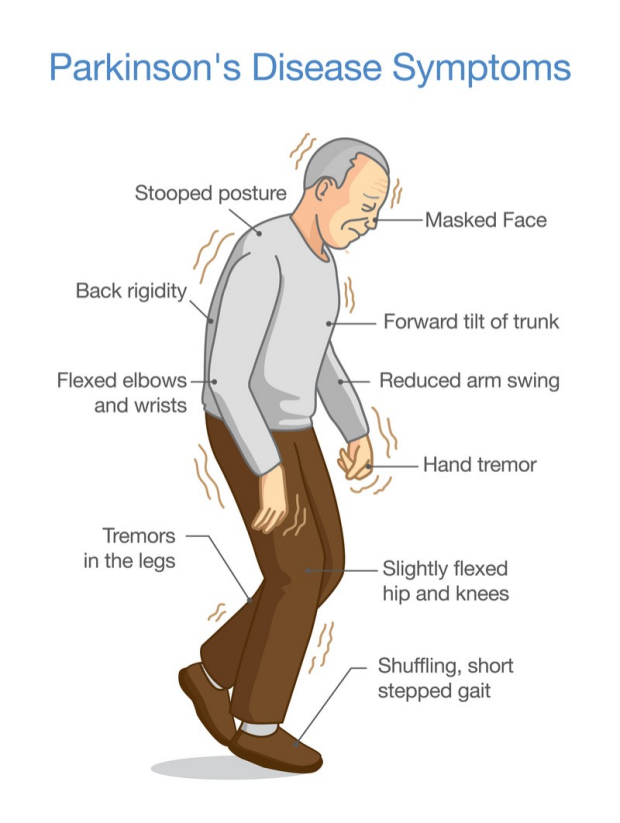
Parkinsonova bolest (PB) je progresivna neurodegenerativna bolest koja prvenstveno pogađa motoričke funkcije, ali može imati uticaj i na druge telesne sisteme. Bolest se razvija usled degeneracije dopaminergičkih neurona u delu srednjeg mozga koji je ključan za regulaciju pokreta. Smanjena proizvodnja dopamina, neurotransmitera odgovornog za kontrolu pokreta, dovodi do klasičnih simptoma Parkinsonove bolesti. Ova bolest se često javlja kod starijih osoba, ali može pogoditi i mlađu populaciju.



*Slika 1 – Izgled mozga zdrave i osobe sa Parkinsonovom bolešću*

* 1. **Klinička slika osobe sa Parkinsonovom bolešću**

Klinička slika Parkinsonove bolesti varira među pacijentima, ali se bolest najčešće manifestuje sledećim karakterističnim simptomima:



*Slika 2 – simptomi osobe sa PB*

**Tremor u mirovanju** jedan je od najprepoznatljivijih simptoma Parkinsonove bolesti i često je prvi simptom koji pacijenti primećuju. Obično se pojavljuje u jednoj ruci, ali se kasnije može proširiti na druge delove tela. Tremor se najjasnije manifestuje kada su mišići opušteni, a smanjuje se ili nestaje tokom voljnih pokreta i u snu. Tremor kod Parkinsonove bolesti je ritmičan, s frekvencijom od 4-6 Hz.

**Rigidnost** podrazumeva povećanu otpornost na pasivne pokrete, koja je prisutna u svim mišićima i zglobovima. Ona može dovesti do ukočenosti i ograničenja pokreta, što pacijentima otežava svakodnevne aktivnosti. Rigidnost je najizraženija u zglobovima i ekstremitetima i može izazvati bolove u mišićima. Klasičan znak rigidnosti je pojava otpora prilikom savijanja ili ispravljanja ekstremiteta.

**Bradikinezija** predstavlja usporenost pokreta i smanjenu sposobnost za započinjanje i izvođenje voljnih pokreta. Ovaj simptom može značajno otežati svakodnevne aktivnosti, kao što su hodanje, oblačenje ili pisanje. Bradikinezija se takođe manifestuje smanjenjem spontane motorike, što dovodi do smanjenog zamaha ruku tokom hoda, sitnog pisanja (mikrografija) i otežanog govora.

**Posturalna nestabilnost** se odnosi na poteškoće u održavanju ravnoteže i koordinacije. Kako bolest napreduje, pacijenti postaju skloni padovima zbog smanjene sposobnosti da održe stabilan stav ili reaguju na promene u položaju tela. Ovaj simptom je često povezan sa kasnijim fazama bolesti i predstavlja jedan od glavnih uzroka invaliditeta kod pacijenata sa Parkinsonovom bolešću.

**Hipomimija**, poznata i kao "mask-like face" (lice poput maske), je smanjenje izraza lica uzrokovano rigidnošću mišića lica. Ovo može dovesti do smanjenja treptanja i osećaja emocionalne praznine u izrazu lica pacijenta, što može uticati na komunikaciju i društvene interakcije.

**Problemi sa govorom i glasom** koje često uzrokujeParkinsonova bolest. Glas može postati slabiji, tiši, monoton ili nejasan, što je rezultat rigidnosti i disfunkcije mišića koji kontrolišu govor. Ove promene u glasu mogu otežati pacijentima da se izraze i komuniciraju sa drugima.

Pored motoričkih simptoma, pacijenti sa Parkinsonovom bolešću mogu iskusiti i niz **nemotoričkih simptoma**, kao što su depresija, anksioznost, poremećaji spavanja, gubitak čula mirisa, konstipacija i kognitivni pad. Ovi simptomi mogu značajno narušiti kvalitet života pacijenata i često su manje prepoznati u ranim fazama bolesti.

* 1. **Dijagnostika Parkinsonove bolesti u realnom životu**

Dijagnostikovanje Parkinsonove bolesti je prvenstveno klinički proces, koji se zasniva na proceni simptoma i znakovima koje pacijent pokazuje. Ne postoji specifičan laboratorijski test koji može potvrditi dijagnozu Parkinsonove bolesti, što dijagnostiku čini izazovnom, posebno u ranim fazama bolesti kada su simptomi suptilni.

Dijagnoza Parkinsonove bolesti počinje detaljnom **anamnezom i kliničkim pregledom.** Lekar će postaviti pitanja o simptomima, njihovom početku i razvoju, kao i o porodičnoj istoriji neuroloških poremećaja. Tokom pregleda, lekar će proceniti prisutnost karakterističnih motoričkih simptoma kao što su tremor, rigidnost, i bradikinezija. Klinički pregled obuhvata i procenu posturalne stabilnosti i koordinacije, kao i testove za identifikaciju nemotoričkih simptoma.

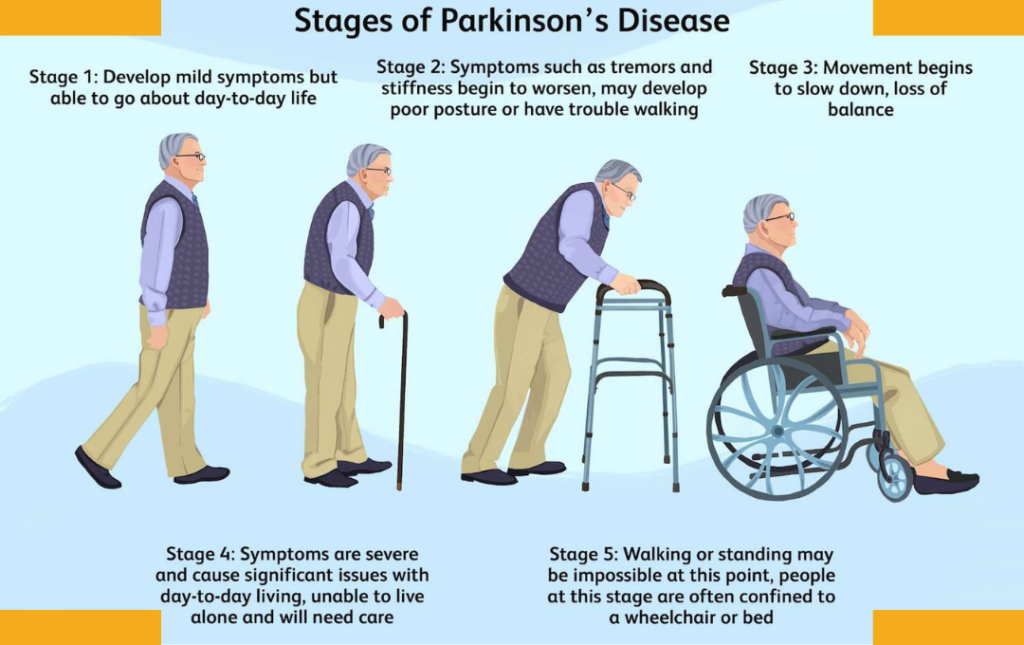
**Neuropatološki testovi,** kao što je UPDRS (Unified Parkinson's Disease Rating Scale), koriste se za kvantifikaciju težine simptoma i praćenje progresije bolesti. UPDRS je sveobuhvatna skala koja ocenjuje motoričke i nemotoričke simptome, uključujući funkcionalne sposobnosti pacijenta. Ovaj test je koristan alat za praćenje odgovora na terapiju i prilagođavanje lečenja.

**Neuroimaging metode**, kao što su MRI (magnetna rezonanca) i CT (kompjuterska tomografija), koriste se za isključivanje drugih mogućih uzroka simptoma, poput tumora ili moždanog udara. Specijalizovane metode, kao što je DaTscan (dopamin transporter imaging), mogu pomoći u potvrđivanju smanjenja dopamina u mozgu, što je karakteristično za Parkinsonovu bolest. DaTscan omogućava vizualizaciju dopaminergičkih neurona u mozgu i može biti koristan u razlikovanju Parkinsonove bolesti od drugih poremećaja sa sličnim simptomima.

Poboljšanje simptoma **nakon uzimanja levodope**, glavnog leka za Parkinsonovu bolest, može potvrditi dijagnozu. Levodopa se u telu pretvara u dopamin, čime se nadoknađuje njegov nedostatak u mozgu. Pacijenti sa Parkinsonovom bolešću obično pokazuju značajno poboljšanje simptoma nakon uzimanja levodope, što se koristi kao dijagnostički kriterijum.

* 1. **Izgled osobe sa Parkinsonovom bolešću i prvi simptomi**

Izgled osobe sa Parkinsonovom bolešću često odražava karakteristične motoričke simptome, kao što su tremor, rigidnost i bradikinezija. Prvi simptomi bolesti mogu biti suptilni i često se pripisuju starenju ili drugim manje ozbiljnim stanjima.



*Slika 3 – prikaz različitih stadijuma PB*

Osoba sa Parkinsonovom bolešću često pokazuje sledeće fizičke karakteristike:  
Tremor je često vidljiv i karakterističan za bolest.   
Pacijenti mogu imati tremor u jednoj ruci ili nozi, koji se kasnije može proširiti na druge delove tela.   
Ukočenost mišića može otežati pokretanje i savijanje zglobova.   
Zbog rigidnosti mišića lica, izrazi lica su jednolični, što dovodi do izgleda "maske" lica.   
Hod postaje usporen i nestabilan, sa smanjenim zamahom ruku i tendencijom da se pacijent naginje napred. Posturalna nestabilnost može izazvati poteškoće u održavanju ravnoteže, što povećava rizik od padova.  
Govor može postati tih, monoton ili nerazumljiv.   
Rani znaci mogu uključivati nesanicu, nemirno spavanje ili poremećaje u REM fazi sna, kao što su nasilni pokreti tokom snova.  
Jedan od ranih, nemotoričkih simptoma može biti smanjenje ili gubitak čula mirisa (anosmija), često godinama pre pojave motoričkih simptoma.  
Mikrografija, ili postepeno smanjenje veličine rukopisa, može biti jedan od prvih simptoma. Rukopis postaje sitan i teško čitljiv.

Ovi rani simptomi mogu varirati u intenzitetu i brzini napredovanja, a njihovo prepoznavanje i rano dijagnostikovanje može značajno uticati na planiranje i efikasnost lečenja Parkinsonove bolesti.

* 1. **Lečenje Parkinsonove bolesti**

Lečenje Parkinsonove bolesti usmereno je na ublažavanje simptoma i poboljšanje kvaliteta života pacijenata, jer trenutno ne postoji lek koji može zaustaviti ili usporiti progresiju bolesti. Terapija obuhvata farmakološke, hirurške i nefarmakološke pristupe.

Farmakološka terapija je osnovni oblik lečenja Parkinsonove bolesti i uključuje upotrebu lekova koji povećavaju dopaminergičku aktivnost u mozgu ili imitiraju dopaminske efekte.  
Levodopa je najefikasniji lek za lečenje simptoma Parkinsonove bolesti. U telu se pretvara u dopamin, čime se nadoknađuje njegov nedostatak u mozgu. Levodopa se često kombinuje sa inhibitorom dopa dekarboksilaze (kao što je karbidopa) kako bi se smanjili neželjeni efekti i povećala efikasnost.  
Dopaminski agonisti, poput pramipeksola, ropinirola i rotigotina, imitiraju efekte dopamina u mozgu. Koriste se kao monoterapija u ranim fazama bolesti ili u kombinaciji sa levodopom u kasnijim fazama.  
Inhibitori MAO-B kao što su selegilin i rasagilin inhibiraju enzim monoaminooksidazu-B, koji razgrađuje dopamin, čime povećavaju njegovu koncentraciju u mozgu.  
Inhibitori COMT entakapon i tolkapon inhibiraju katehol-O-metiltransferazu, enzim koji razgrađuje dopamin, čime produžavaju efekat levodope.  
Antikolinergici, kao što je benztropin, koriste se za ublažavanje tremora i rigidnosti, ali imaju ograničenu upotrebu zbog neželjenih efekata kao što su konfuzija i halucinacije.  
Amantadin koristi se za ublažavanje diskinezija izazvanih levodopom i ima blage antiparkinsonske efekte.

Hirurške metode se koriste kada farmakološka terapija nije dovoljno efikasna u kontroli simptoma.  
DBS(Deep brain stimulation) je najčešće korišćena hirurška metoda za lečenje Parkinsonove bolesti. Uključuje implantaciju elektroda u specifične delove mozga, kao što su subthalamic nucleus ili globus pallidus internus, koje se povezuju sa uređajem koji generiše električne impulse. DBS može značajno smanjiti motoričke simptome i poboljšati kvalitet života pacijenata.  
Lesioning metode kao što su pallidotomija ili thalamotomija uključuju stvaranje lezija u specifičnim delovima mozga kako bi se smanjili motorički simptomi. Ove metode su danas ređe korišćene zbog uspeha DBS-a.

Nefarmakološki pristupi su važan deo sveobuhvatnog lečenja Parkinsonove bolesti i uključuju rad fizioterapeuta, logopeda, psihoterapeuta i promene u ishrani.  
Fizioterapija pomaže u održavanju pokretljivosti, snage i ravnoteže. Vežbe istezanja, jačanja mišića i koordinacije mogu smanjiti ukočenost i poboljšati funkcionalnost.   
Logopedi rade sa pacijentima na poboljšanju govora, glasovne projekcije i gutanja, što može biti otežano kod Parkinsonove bolesti.  
Psihoterapija može pomoći pacijentima u suočavanju sa emocionalnim i psihološkim izazovima koje nosi Parkinsonova bolest, kao što su depresija i anksioznost.  
Uravnotežena ishrana može pomoći u upravljanju simptomima i poboljšanju opšteg zdravlja. Unos vlakana i tečnosti je važan za prevenciju konstipacije, a unos proteina može biti prilagođen kako bi se optimizovala apsorpcija levodope.

Sveobuhvatan pristup lečenju Parkinsonove bolesti zahteva multidisciplinarni tim lekara, terapeuta i stručnjaka koji zajedno rade na pružanju najbolje moguće nege za pacijente.



*Slika 4 – napredno lečenje PB*

1. **Postavka problema**

Parkinsonova bolest je neurodegenerativni poremećaj koji utiče na motoričke funkcije, a simptomi se postepeno pogoršavaju tokom vremena. Identifikacija i dijagnoza Parkinsonove bolesti u ranim fazama je od vitalnog značaja za efikasno upravljanje bolešću i poboljšanje kvaliteta života pacijenata. Međutim, rana dijagnoza je izazovna usled suptilnih i nespecifičnih simptoma koji mogu biti slični drugim poremećajima.

Tradicionalne metode dijagnostikovanja, kao što su klinički pregledi i vizualizacione tehnike, često nisu dovoljno precizne u ranoj fazi bolesti. Zato je potrebno razviti metode koje koriste napredne analize podataka i mašinsko učenje kako bi se identifikovali suptilni obrasci u medicinskim parametrima koji mogu ukazivati na prisustvo Parkinsonove bolesti.

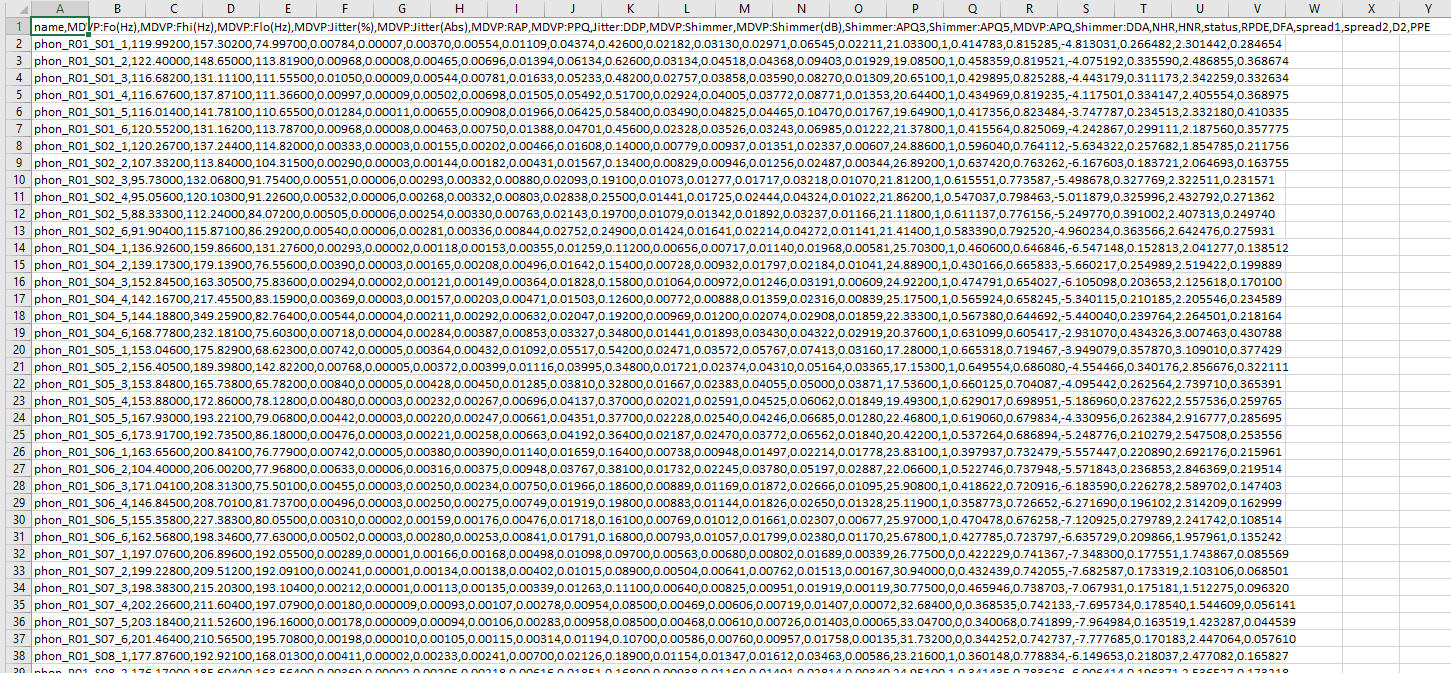
Ovaj projekat ima za cilj da razvije model za prepoznavanje Parkinsonove bolesti analizom skupa podataka koji sadrži različite medicinske parametre dobijene od pacijenata. Korišćenjem metoda poput KNN, SVM, Perceprona i LR, projekat će pokušati da identifikujenn prisustvo bolesti na osnovu ovih parametara, pružajući alat za podršku u dijagnostičkom procesu.

1. **Dataset**

Dataset koji se koristi u ovom projektu sadrži različite medicinske parametre koji su prikupljeni od pacijenata sa i bez Parkinsonove bolesti. Ovi parametri su kvantitativni i služe za procenu različitih aspekata motornih funkcija i govora, koji su ključni za dijagnozu Parkinsonove bolesti.

* 1. **Opis dataset-a**

Dataset koji koristimo za predikciju Parkinsonove bolesti sadrži ukupno 195 uzoraka i 24 kolone. Podaci su izvučeni iz glasovnih zapisa pacijenata, a glavna svrha ovog dataset-a je razlikovanje zdravih pojedinaca od onih sa Parkinsonovom bolešću.



*Slika 5 – dataset ,,Parkinsons.csv“*

Dataset obuhvata sledeće kolone:

**`*name*`**: Identifikator pacijenta. Ime uzorka ili pacijenta.

**`*MDVP:Fo(Hz)*`**: Osnovna frekvencija glasa osobe (Hz) poznata kao Fo (Fundamental Frequency). Meri se u hercima (Hz) i predstavlja broj vibracija glasnih žica u sekundi. Osnovna frekvencija može biti različita kod osoba sa Parkinsonovom bolešću zbog rigidnosti glasnih žica.

**`*MDVP:Fhi(Hz)*`**: Najviša fundamentalna frekvencija glasa. Predstavlja maksimalnu frekvenciju koju glas može proizvesti tokom govora. Promene u ovoj vrednosti mogu ukazivati na gubitak kontrole nad glasnim žicama.

**`*MDVP:Flo(Hz)*`**: Najniža fundamentalna frekvencija glasa. Ovo je najniža frekvencija koju glas može proizvesti. Osobe sa Parkinsonovom bolešću često imaju smanjenu fleksibilnost u promeni osnovne frekvencije.

***`MDVP:Jitter(%)`****:* Kvantitativna mera nesavršenosti glasa (vibracija). Jitter[[1]](#footnote-1) predstavlja male, nasumične promene u osnovnoj frekvenciji glasa. Ove promene mogu biti znak nedostatka stabilnosti glasnica, što je često prisutno kod osoba sa Parkinsonovom bolešću.

***`MDVP:Jitter(Abs*)`**: Apsolutna vrednost nesavršenosti glasa. Ovo je mera jitter-a u apsolutnim vrednostima, ne u procentima. Takođe meri nepravilnosti u osnovnoj frekvenciji glasa

***`MDVP:RAP`***: Relativna amplitudna perturbacija. RAP je sličan jitter-u i meri konzistentnost osnovne frekvencije, ali sa fokusom na kratke vremenske intervale.

***`MDVP:PPQ`***: Periodična perturbacija glasa. Ovo je još jedan način merenja jitter-a, ali koristi drugačiji algoritam za izračunavanje.

***`Jitter:DDP`***: Razlika u periodičnim perturbacijama. DDP meri razliku u jitter-u između susednih ciklusa i koristi se kao dodatni indikator nestabilnosti glasa.

***`MDVP:Shimmer`***: Amplitudna nesavršenost glasa. Shimmer meri varijacije u amplitudi glasa, što se često manifestuje kao promena glasnoće tokom govora. Osobe sa Parkinsonovom bolešću mogu imati izraženije varijacije zbog problema sa kontrolom mišića glasnica

***`MDVP:Shimmer(dB)`***: Amplitudna nesavršenost glasa izražena u decibelima.

***`Shimmer:APQ3`***: Amplitudna perturbacija glasa na osnovu tri periode.

`**Shimmer:APQ5**`: Amplitudna perturbacija glasa na osnovu pet perioda.

***`MDVP:APQ`***: Amplitudna perturbacija glasa na osnovu svih perioda.

***`Shimmer:DDA`***: Razlika u amplitudnim perturbacijama. DDA meri razliku u shimmer-u između susednih ciklusa.

***`NHR`***: Odnos šuma prema harmonici. NHR meri odnos između šuma i harmonijskih komponenti u glasu. Osobe sa Parkinsonovom bolešću često imaju povećan nivo šuma zbog nesposobnosti glasnica da pravilno vibriraju.

***`HNR`***: Odnos harmonika prema šumu. HNR meri koliko su izražene harmonijske komponente u odnosu na šum. Viši HNR ukazuje na čistu produkciju glasa, dok niži HNR može ukazivati na prisutnost šuma.

***`status`***: Dijagnostički status (1 - Parkinson, 0 - Zdrav).

***`RPDE`***: Recurrence period density entropy - entropija gustine perioda ponavljanja.

***`DFA`***: Detrended fluctuation analysis - analiza fluktuacije bez trenda. Alfa koeficijent izračunat metodom detrended fluctuation analysis. DFA meri dugoročne korelacije u govornom signalu, što može ukazivati na promene u kontrolnim mehanizmima glasa.

***`spread1`***: Mera spektralne širine. Ovaj parametar meri koliko se glas širi u frekvencijskom spektru.

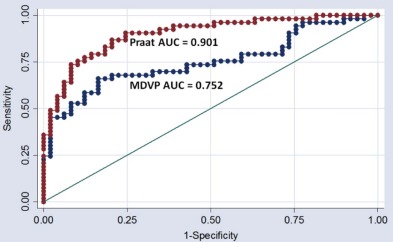
***`spread2`***: Druga mera spektralne širine. Sličan spread1, ali koristi drugačiji algoritam za računanje.

***`D2`***: Korrelaciona dimenzija. Ova mera se koristi za kvantifikovanje složenosti glasa, gde se višim vrednostima D2 pripisuje veća složenost.

***`PPE`***: Pitch period entropy - entropija perioda tona. PPE meri neizvesnost u glasu, gde veće vrednosti mogu ukazivati na manje stabilan glas.

* 1. **Dobijanje medicinskih parametara iz dataseta**

Medicinski parametri u datasetu dobijaju se kroz različite kliničke procedure i korišćenje specifičnih uređaja. Sledeći opis prikazuje kako se svaki od ovih parametara meri i koji uređaji se koriste u procesu.



*Slika 6 – MDVP Multi-Dimensional Voice Program*

Fundamentalna frekvencija glasa (MDVP:Fo, MDVP:Fhi, MDVP:Flo) se dobija analizom glasa pacijenta. Pacijent izgovara određene zvuke ili rečenice u mikrofon, a uređaji kao što su fonometri i specijalizovani softver za analizu glasa (npr. MDVP - Multi-Dimensional Voice Program) mere fundamentalnu frekvenciju glasa i njene varijacije.

Jitter (MDVP:Jitter, MDVP:Jitter(Abs), MDVP:RAP, MDVP:PPQ, Jitter:DDP) predstavlja nesavršenost u frekvenciji glasa, odnosno varijacije u periodima vibracije glasnica. Ove mere se dobijaju takođe analizom glasa pomoću softvera za analizu glasa, koji kvantifikuje ove varijacije. Pacijent izgovara određene fraze ili glasove, a uređaj beleži varijacije u frekvenciji (jitter).

Shimmer (MDVP:Shimmer, MDVP:Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, MDVP:APQ, Shimmer:DDA) predstavlja nesavršenost u amplitudi glasa. Slično kao i jitter, ove mere se dobijaju analizom glasa pomoću specijalizovanog softvera koji meri varijacije u amplitudi zvučnih talasa.

Odnos šuma prema harmonici (NHR) i odnos harmonike prema šumu (HNR) mere količinu šuma u glasu u odnosu na harmonijske komponente. Mere se pomoću softvera za analizu glasa koji razdvaja šum od harmonijskih komponenti i izračunava njihov odnos. Pacijent izgovara dugotrajne glasove, a uređaj meri prisutnost šuma u glasu.

Recurrence period density entropy (RPDE) je kompleksna mera koja se koristi za analizu nelinearnih dinamičkih sistema. U kontekstu analize glasa, ova mera se izračunava pomoću algoritama koji analiziraju gustinu perioda ponavljanja u vremenskim serijama glasa.

Detrended fluctuation analysis (DFA) je metoda koja se koristi za kvantifikaciju dugoročnih korelacija u vremenskim serijama. U kontekstu Parkinsonove bolesti, DFA se koristi za analizu fluktuacija u glasu koje nisu povezane sa predviđanjima.

Spektralne širine (spread1, spread2) mere distribuciju energije u frekvencijskom spektru glasa. Izračunavaju se pomoću spektralne analize koja razlaže zvučni signal na njegove frekvencijske komponente.

Korrelaciona dimenzija (D2) je mera koja kvantifikuje složenost signala. U analizi glasa, koristi se za kvantifikaciju složenosti dinamičkih obrazaca u zvučnom signalu.

Pitch period entropy (PPE) meri entropiju perioda tona, odnosno varijabilnost u trajanju perioda vibracije glasnica. Izračunava se pomoću softverskih algoritama koji analiziraju periodične komponente glasa.



*Slika 7 – kontrolisano okruženje za snimanje glasa pacijenata*

Dakle proces merenja ovih parametara uključuje pacijente koji izgovaraju određene zvuke ili rečenice u kontrolisanom okruženju, dok se njihovi glasovi snimaju visokokvalitetnim mikrofonima. Snimci se zatim analiziraju pomoću specijalizovanog softvera koji kvantifikuje razne aspekte zvučnog signala i generiše potrebne parametre za dalju analizu.

Koristeći ove parametre, moguće je razviti modele mašinskog učenja koji mogu pomoći u ranoj dijagnozi Parkinsonove bolesti, identifikacijom obrazaca u glasu koji su specifični za ovu bolest. Ovi modeli mogu postati značajan alat u kliničkoj praksi, pružajući podršku lekarima u procesu dijagnostike i omogućavajući precizniju i bržu identifikaciju bolesti.

1. **Klasifikacioni algoritmi**

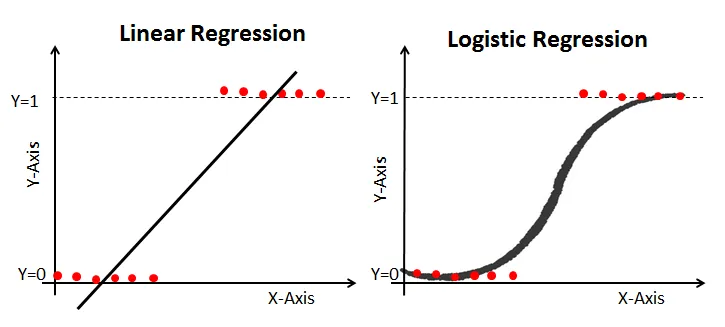
Klasifikacioni algoritmi su ključni za rešavanje mnogih zadataka u oblasti mašinskog učenja, uključujući prepoznavanje obrazaca, otkrivanje anomalija i donošenje odluka. Klasifikacija se odnosi na proces dodeljivanja ulaznih podataka u unapred definisane kategorije.

Klasifikacioni algoritmi funkcionišu tako što analiziraju obeležja (features) skupa podataka i uče obrazac koji razdvaja klase podataka. Ovi algoritmi mogu biti podeljeni na više tipova, kao što su linearni i nelinearni klasifikatori, probabilistički klasifikatori i dr. Svaki algoritam ima svoje prednosti i mane, kao i specifične primene.

U klasifikacione algoritme spada logistička regresija, KNN, SVM,...

* 1. **Logistička regresija**

Logistička regresija je statistički model koji se koristi za binarne klasifikacione probleme. Uprkos svom nazivu, logistička regresija je metoda klasifikacije, a ne regresije. Najčešće se koristi kada je izlazna promenljiva diskretna, poput "da" ili "ne", "1" ili "0". Ona predviđa verovatnoću da podatak pripada jednoj od kategorija.

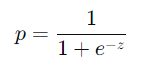


*Slika 8 – moć logističke regresije u mašinskom učenju*

Za razliku od linearne regresije, koja predviđa kontinuirane vrednosti, logistička regresija modelira verovatnoću događaja i ograničava predikciju na opseg između 0 i 1. To se postiže korišćenjem logističke funkcije, poznate i kao sigmoidalna funkcija.

Logistička regresija koristi linearnu kombinaciju ulaznih podataka kako bi izračunala verovatnoću. Ova verovatnoća se transformiše pomoću logističke funkcije kako bi se ograničila između 0 i 1, čime se omogućava binarna klasifikacija.

Sigmoidna funkcija koja se koristi u logističkoj regresiji je definisana kao:



gde je p verovatnoća klasifikacije u klasu 1, a z linearna kombinacija karakteristika.

Ako je p≥0.5, klasifikacija će biti 1, inače će biti 0.

Za treniranje modela koristi se logaritamska funkcija gubitka (log-loss), koja penalizuje netačne predikcije na osnovu njihove udaljenosti od stvarne vrednosti. Funkcija gubitka se računa kao:



gde su y stvarna oznaka, a p predviđena verovatnoća.

Za optimizaciju modela koristi se gradientni spust, koji iterativno ažurira parametre modela u pravcu minimizacije funkcije gubitka. Stopa učenja α određuje veličinu koraka prilikom ažuriranja parametara.

Logistička regresija koristi regularizaciju kako bi izbegla prekomerno prilagođavanje modela podacima (overfitting). Regularizacija može biti L2 (Ridge) ili L1 (Lasso), gde se dodaju kazne za velike težine u funkciji gubitka.

Prednosti LR:

- Jednostavna za implementaciju i interpretaciju.

- Rezultat je verovatnoća, što omogućava lako tumačenje u smislu rizika ili šansi.

- Dobro funkcioniše kada postoji linearna separacija podataka.

Mane LR:

- Ne može efikasno raditi na nelinearnim problemima bez transformacije ulaznih podataka (npr. korišćenjem polinomijalnih karakteristika).

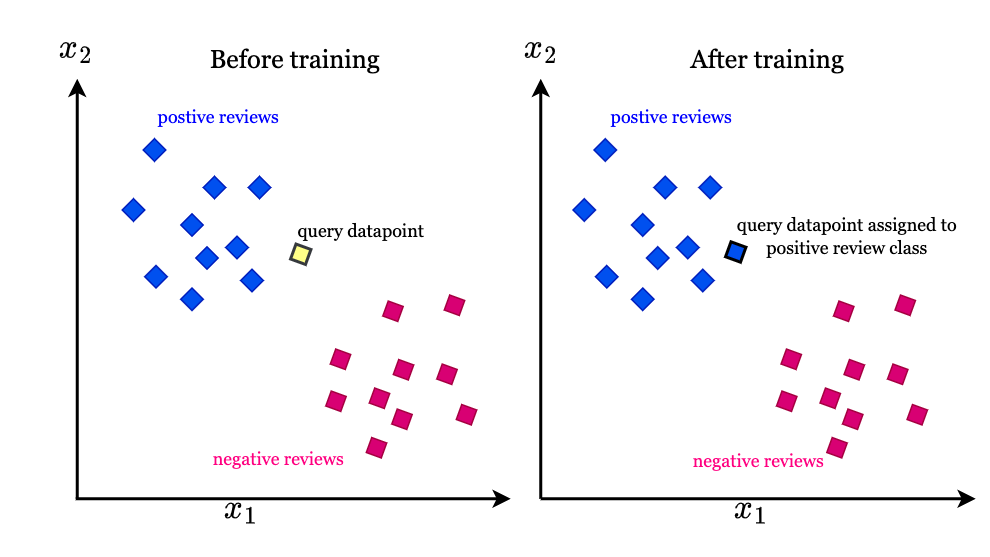
- Osetljiva je na prisustvo multikolinearnosti, odnosno visoke korelacije između ulaznih karakteristika.

- Može se loše prilagoditi složenim dataset-ovima sa puno noise-a, zbog svoje tendencije ka overfiting-u.

* 1. **K-Nearest Neighbors (KNN)**

K-Nearest Neighbors (KNN) je jednostavan, ali efikasan algoritam za klasifikaciju i regresiju. KNN koristi sličnost između podataka kako bi predvideo vrednosti ciljne varijable na osnovu vrednosti najbližih suseda u trening skupu.

KNN je ne-parametarski algoritam što znači da ne pravi nikakav eksplicitni model, već koristi ceo trening skup tokom faze predikcije. Algoritam klasifikuje nepoznat primer na osnovu većinskog glasanja K najbližih primera u trening skupu (za klasifikaciju) ili izračunava prosečnu vrednost ciljne varijable K najbližih suseda (za regresiju).



*Slika 9 – KNN*

Princip rada KNN algoritma može se opisati kroz sledeće korake:

* Izračunavanje udaljenosti:

Algoritam računa udaljenost između nove tačke i svih tačaka u trening skupu. Najčešće korišćena metrika za udaljenost je euklidska udaljenost, ali mogu se koristiti i druge metrike poput Manhattan ili Minkowski udaljenosti.

* Odabir K suseda:

Nakon izračunavanja udaljenosti, KNN algoritam identifikuje K najbližih tačaka (suseda) u trening skupu.

* Klasifikacija ili regresija:

Za klasifikaciju, KNN koristi većinsko glasanje suseda za određivanje klase.

Za regresiju, KNN koristi prosečnu vrednost ciljne varijable suseda za predikciju.

* Predikcija:

Na osnovu informacija dobijenih od K najbližih suseda, KNN algoritam predviđa vrednost ciljne varijable za novu tačku.

Pretpostavimo da imamo dataset sa podacima o tipu cveća, gde svaka tačka predstavlja karakteristike cveta (dužina latica, širina latica, itd.) i ciljna varijabla je vrsta cveta. Za novi cvet, KNN računa euklidsku udaljenost između ovog cveta i svih cvetova u trening skupu. Ako je K postavljeno na 3, KNN bira 3 cveta iz trening skupa koji su najbliži novom cvetu prema izračunatoj udaljenosti. Ako od 3 najbliža cveta, 2 su vrste „setosa”, a 1 je vrste „versicolor”, KNN predviđa da je novi cvet vrste „setosa”. KNN predviđa vrstu cveta na osnovu većinskog glasanja najbližih suseda.

Prednosti KNN:

- KNN je jednostavan za implementaciju i razumevanje.

- Može se koristiti za klasifikaciju i regresiju.

- Ne zahteva pretpostavke o distribuciji podataka.

Mane KNN:

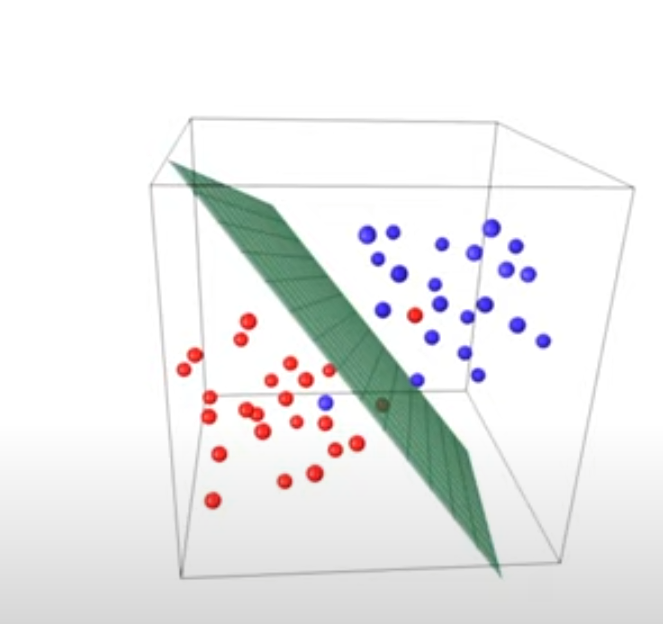
- Algoritam je spor za velike datasete jer zahteva izračunavanje udaljenosti do svih tačaka u trening skupu.

- Performanse zavise od izbora K i metrike udaljenosti.

- Osetljiv je na outliere i irelevantna obeležja.

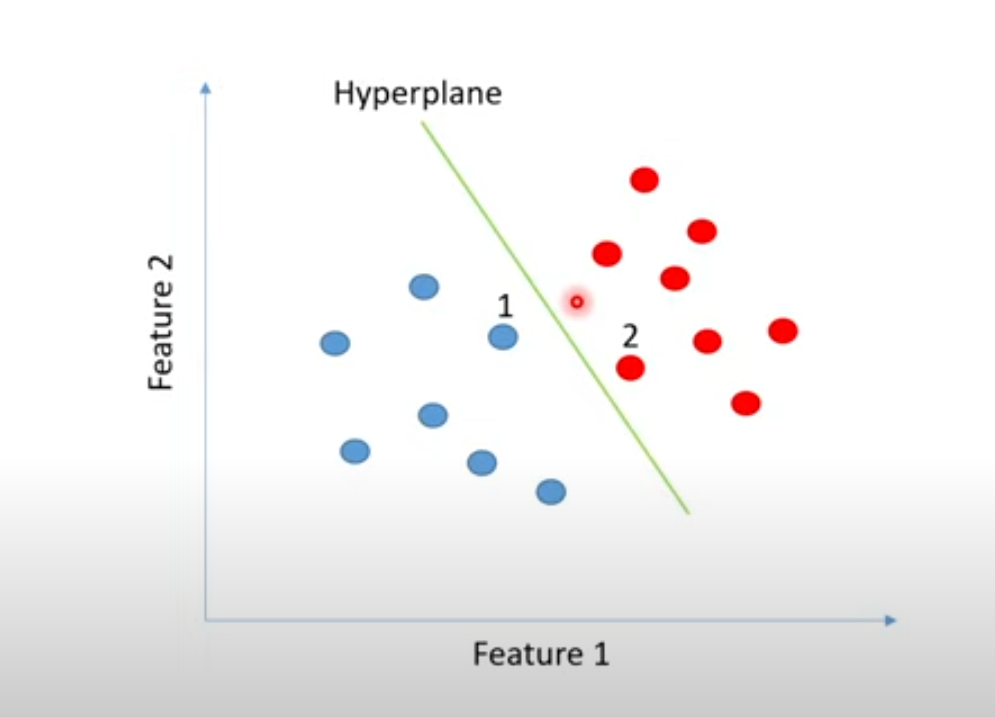
* 1. **Support Vector Machines (SVM)**

Support Vector Machines (SVM) je moćan algoritam za klasifikaciju i regresiju koji koristi hiper-ravne za razdvajanje klasa u višedimenzionalnom prostoru. SVM je naročito efikasan u prostorima visoke dimenzionalnosti i može modelovati kompleksne nelinearne relacije pomoću kernel funkcija.



*Slika 10 – SVM u trodimenzionalnom prostoru*

SVM je algoritam nadgledanog učenja koji se koristi za binarnu klasifikaciju, ali može se proširiti i na višeklasnu klasifikaciju. Glavni cilj SVM-a je pronaći hiper-ravnu koja najbolje razdvaja klase podataka sa maksimalnom marginom.



*Slika 11 – SVM primer u dvodimenzionalnom prostoru*

Princip rada SVM algoritma može se opisati kroz nekoliko ključnih koraka:

* Izbor optimalne hiper-ravni:

SVM traži hiper-ravan koja maksimalno razdvaja dve klase. Optimalna hiper-ravan je ona koja ima najveću marginu, tj. razdaljinu između najbližih tačaka (support vectors) iz svake klase i hiper-ravni

* Kernel trik:

Ako podaci nisu linearno separabilni, SVM koristi kernel trik za mapiranje podataka u viši dimenzionalni prostor gde postaju linearno separabilni. Uobičajeni kernel funkcije su polinomski kernel, RBF (Radial Basis Function) kernel, itd.

* Izračunavanje margina:

Margina je razdaljina između hiper-ravni i najbližih tačaka iz svake klase. SVM optimizuje marginu tako da bude što veća, čime se poboljšava generalizacija modela.

* Predikcija:

Kada se optimalna hiper-ravan izračuna, SVM koristi ovu hiper-ravan za klasifikaciju novih tačaka na osnovu njihove pozicije u odnosu na hiper-ravan.

Pretpostavimo da imamo dataset sa podacima o tipu cveća, gde svaka tačka predstavlja karakteristike cveta, a ciljna varijabla je vrsta cveta. SVM traži hiper-ravan koja razdvaja cvetove različitih vrsta sa maksimalnim marginom. Ako cvetovi nisu linearno separabilni, koristi se RBF kernel za mapiranje podataka u višu dimenzionalnost. SVM optimizuje marginu između najbližih tačaka različitih vrsta. SVM koristi izračunatu hiper-ravan za klasifikaciju novih cvetova.

Prednosti SVM:

- Efikasan u visokodimenzionalnim prostorima.

- Dobro funkcioniše sa jasno razdvojenim klasama.

- Fleksibilan zahvaljujući različitim kernel funkcijama koje omogućavaju modelovanje složenih relacija.

Mane SVM:

- Može biti spor za velike dataset-ove, posebno sa nelinearnim kernel funkcijama.

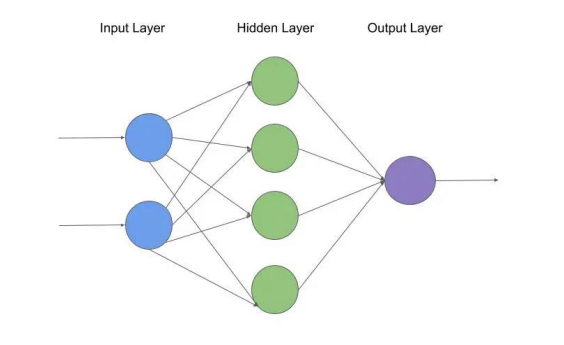
- Performanse zavise od izbora hiperparametara kao što su tip kernela i regularizacija.

- Osetljiv na šum i outliere u podacima.

* 1. **Perceptron**

Perceptron je osnovni algoritam mašinskog učenja koji se koristi za binarnu klasifikaciju. Predstavlja jednostavnu neuronsku mrežu koja može rešavati linearno separabilne probleme. Razvijen od strane Franka Rosenblatta 1957. godine, perceptron postavlja temelje za mnoge naprednije tehnike i algoritme u oblasti dubokog učenja.

Perceptron je najjednostavniji oblik veštačke neuronske mreže koji se sastoji od ulaznog sloja, jednog ili više skrivenih slojeva, i izlaznog sloja. Svaki sloj sadrži određeni broj neurona, gde su svi neuroni povezani sa neuronima sledećeg sloja. U osnovnoj verziji, perceptron ima jedan sloj neurona koji se direktno povezuje sa izlazom.



*Slika 12 – grafički prikaz slojeva perceptrona*

Ulazni sloj: Sadrži broj neurona jednak broju ulaznih promenljivih u skupu podataka.

Izlazni sloj: Sadrži jedan neuron za binarnu klasifikaciju. U slučaju višeklasne klasifikacije, broj neurona u izlaznom sloju odgovara broju klasa.

Skriveni slojevi: Mogu biti prisutni i sastoje se od proizvoljnog broja neurona, ali u osnovnom perceptronu obično nisu prisutni.

Perceptron radi na sledeći način:

* Izračunavanje izlaza:

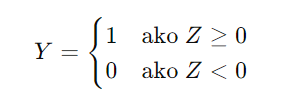
Za ulazni vektor X i težinu W, perceptron izračunava linearnu kombinaciju:

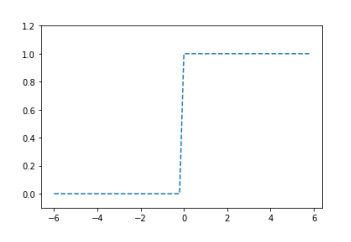


gde je b pristrasnost (bias).

* Aktivaciona funkcija:

Primena aktivacione funkcije na Z kako bi se dobila izlazna vrednost Y:





* Ažuriranje težina:

Ako je predikcija netačna, težine se ažuriraju kako bi se smanjila greška, gde je eta stopa učenja, T stvarna klasa (1 ili 0), a Y predviđena klasa.



* Iteracija:

Proces učenja se ponavlja kroz više epoha, tj. prolazaka kroz čitav trening skup, dok se težine ne stabilizuju ili dok se ne postigne zadovoljavajuća tačnost.

Pretpostavimo da imamo dataset sa dva obeležja koja predstavljaju karakteristike glasa pacijenata sa Parkinsonovom bolešću i zdravih pacijenata. Cilj je klasifikovati pacijente u dve klase: pacijenti sa Parkinsonovom bolešću i zdravi pacijenti.

Težine se inicijalizuju na nule ili male nasumične vrednosti. Bias se takođe postavlja na nulu ili malu nasumičnu vrednost. Za svaki primer iz trening skupa, perceptron računa linearnu kombinaciju ulaznih vrednosti i težina. Ako je rezultat linearne kombinacije veći ili jednak nuli, predikcija je 1; inače, predikcija je 0. Ako je predviđanje tačno, težine se ne menjaju. Ako je predviđanje netačno, težine se ažuriraju kako bi se smanjila greška. Proces se ponavlja kroz više epoha dok se greške ne smanje ili dok se model ne stabilizuje.

Prednosti perceptrona:

- Perceptron je jednostavan za implementaciju i razumevanje.

- Efikasan je za binarnu klasifikaciju kada su podaci linearno separabilni.

- Osnova je za složenije modele neuronskih mreža i dubokog učenja.

Mane perceptrona:

- Ne može rešavati nelinearno separabilne probleme (npr. XOR problem).

- Osetljiv je na izbor stope učenja i inicijalnih težina.

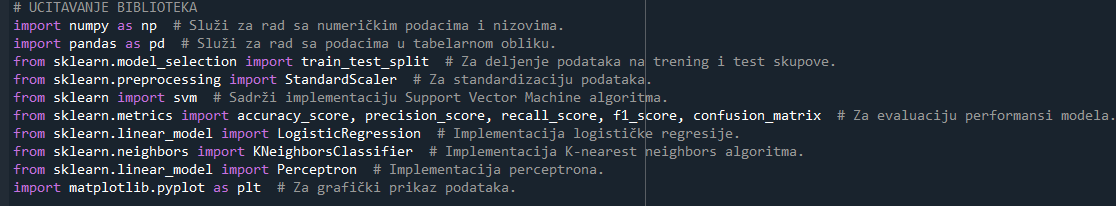
- Performanse mogu biti ograničene na složenijim i višedimenzionalnim podacima.

1. **Implementacija**

U ovom delu seminarskog rada, detaljno ćemo opisati implementaciju klasifikacionih modela za prepoznavanje Parkinsonove bolesti. Implementacija uključuje učitavanje i pregled podataka, preprocesiranje podataka, treniranje modela, evaluaciju modela, prikaz rezultata i proveru predikcija na novim podacima.

* 1. **Prikaz podataka datset-a**

Prvo ćemo učitati potrebne biblioteke i podatke iz CSV fajla, te izvršiti osnovni pregled i analizu podataka.

****

*Slika 13 – učitavanje biblioteka*

Numpy - biblioteka za rad sa numeričkim podacima i nizovima.

Pandas - blioteka za rad sa podacima u tabelarnom obliku (DataFrame).

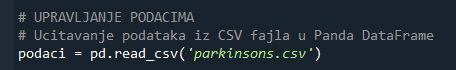
sklearn.model\_selection - modul za deljenje podataka na trening i test skupove.

sklearn.preprocessing - modul za preprocesiranje podataka, uključujući standardizaciju.

Sklearn - sadrži različite algoritme za mašinsko učenje.

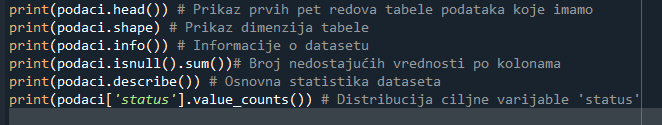
matplotlib.pyplot - biblioteka za pravljenje grafika i vizualizaciju podataka.

Kod za prikaz podataka datseta ima ključnu ulogu u početnoj fazi analize podataka. Ovaj deo koda omogućava razumevanje strukture podataka, njihovih karakteristika, kao i identifikaciju potencijalnih problema kao što su nedostajući podaci.



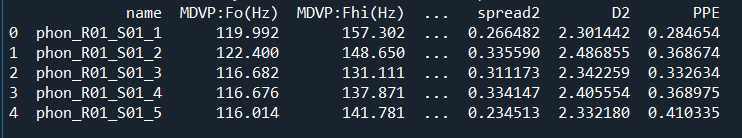
*Slika 14 – učitavanje dataset-a*

Ovaj deo koda učitava podatke o Parkinsonovoj bolesti iz fajla parkinsons.csv u pandas DataFrame koji omogućava jednostavno rukovanje podacima u tabelarnom obliku, olakšava analizu i vizualizaciju.



*Slika 15 – manipulacija i vizualizacija podataka iz dataset-a*

**podaci.head()** je funkcija koja prikazuje prvih pet redova iz dataseta. Cilj je brzo sagledavanje osnovne strukture podataka i provera formata. Prikazuje kolone kao što su name, MDVP:Fo(Hz) (osnovna frekvencija glasa), MDVP:Fhi(Hz) (najviša frekvencija glasa), spread1, D2, PPE, i druge.

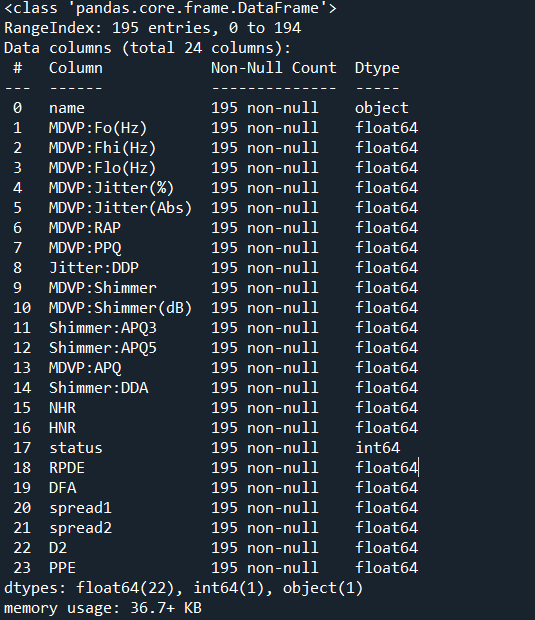


*Slika 16 – izlaz funkcije podaci.head()*



*Slika 17 – izlaz funkcije podaci.shape()*

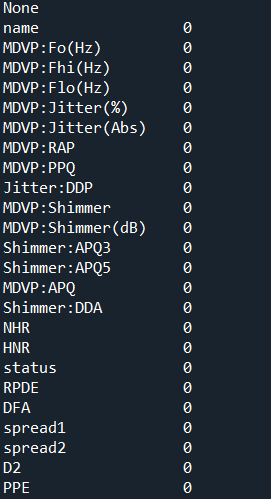
Ovaj izlaz nam govori da imamo **195 uzoraka** (ili pacijenata) i **24 parametra**. Ovaj broj parametara uključuje i ciljnu varijablu 'status', koja označava da li osoba ima Parkinsonovu bolest (1) ili ne (0).



*Slika 18 – izlaz funkcije podaci.info()*

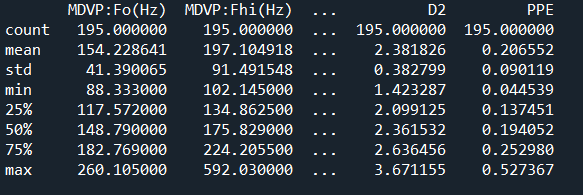
Ova funkcija potvrđuje da dataset nema nedostajućih vrednosti, pošto sve kolone imaju **195 ne-null vrednosti**. Kolone sa merenjima su tipa **float64**, dok je kolona status tipa **int64**, jer označava binarnu klasifikaciju (1 za prisustvo bolesti, 0 za odsustvo).

**podaci.isnull().sum()** prikazuje broj nedostajućih vrednosti po kolonama. Pošto svi rezultati imaju vrednost 0, nema nedostajućih podataka.



*Slika 19 – izlaz funkcije podaci.isnull().sum()*

**podaci.describe()** prikazuje osnovnu statistiku za numeričke kolone, uključujući broj uzoraka, prosečnu vrednost, standardnu devijaciju, minimalnu i maksimalnu vrednost, kao i kvartile (25%, 50%, 75%).



*Slika 20 – izlaz funkcije podaci.describe()*

**MDVP(Hz)**: Osnovna frekvencija glasa — prosečna vrednost je 154.23 Hz, a raspon je od 88.33 do 260.11 Hz.

**MDVP(Hz)**: Najviša frekvencija — prosečna vrednost je 197.10 Hz, ali standardna devijacija ukazuje na varijabilnost, jer maksimalna vrednost dostiže 592 Hz.

Ovo nam ukazuje na to da postoje značajne varijacije u akustičnim karakteristikama glasa među pacijentima, što može biti važno za diferencijaciju osoba sa Parkinsonovom bolešću i zdravih osoba.

**podaci['status'].value\_counts()** prikazuje distribuciju vrednosti u kolonama. Konkretno, ovde gledamo koliko pacijenata ima Parkinsonovu bolest (status = 1) i koliko je zdravih (status = 0).



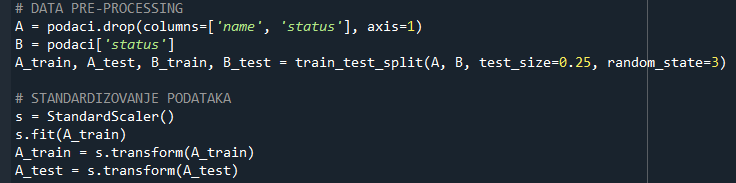
*Slika 21 – status osoba sa i bez PB*

Ovaj rezultat pokazuje da od ukupno 195 uzoraka, **147 pacijenata ima Parkinsonovu bolest (status = 1)**, dok je **48 pacijenata zdravo (status = 0)**. Ova distribucija pokazuje neuravnoteženost klasa, što može uticati na performanse modela klasifikacije, jer će većina modela imati tendenciju da favorizuje većinsku klasu.

* 1. **Treniranje modela**

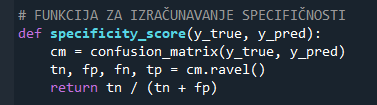
Treniranje modela je ključna faza u mašinskom učenju. Cilj je da modeli nauče obrasce u podacima kako bi mogli da donesu tačne predikcije na novim, neviđenim podacima. U ovom delu ćemo detaljnije opisati proces treniranja modela i analizirati ključne korake kroz kod.

Pre nego što treniramo model, podaci moraju biti pripremljeni. To uključuje izdvajanje ciljne varijable, deljenje na trening i test skupove, kao i standardizaciju podataka. Standardizacija je važna jer mnogi algoritmi uče bolje kada su svi podaci na istoj skali.



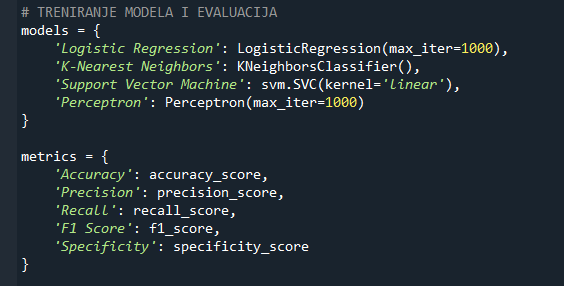
*Slika 22 – priprema i standardizacija podataka*

Kolone 'name' i 'status' se izbacuju iz obeležja (A), dok 'status' postaje ciljna varijabla (B). Podaci se dele na trening (75%) i test (25%) skupove koristeći funkciju train\_test\_split. Kreira se instanca StandardScaler-a koja standardizuje podatke i skaliraju se podaci iz trening skupa. Transformišu se trening i test podaci kako bi imali nultu srednju vrednost i jediničnu standardnu devijaciju.



*Slika 23 – funkcija za izračunavanje specifičnosti*

Funkcija koja računa specifičnost (meri udeo tačno predikovanih negativnih primera u ukupnom broju stvarnih negativnih primera).

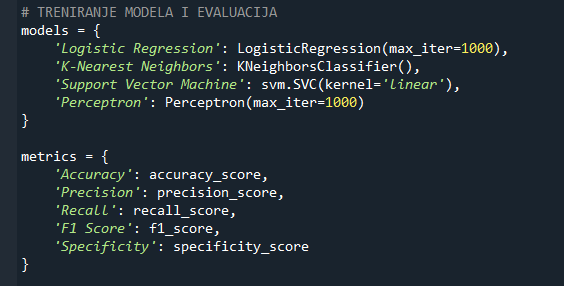


*Slika 24 – kreiranje rečnika korišćenih modela*

Kreira se rečnik models koji sadrži instance različitih modela (Logistic Regression, KNN, SVM, Perceptron).

Treniranje modela podrazumeva da model na osnovu trening skupa nauči obrasce i odnose među obeležjima i ciljnom varijablom. To se postiže optimizacijom težina (u linearnim modelima) ili traženjem najsličnijih primera (KNN).

Nakon treniranja, modeli se testiraju na test skupu. Na ovaj način proveravamo koliko je model uspešno naučio iz podataka i koliko dobro generalizuje na nove podatke.



*Slika 25 – kreiranje rečnika korišćenih metrika*

Kreira se rečnik metrics koji sadrži funkcije za evaluaciju performansi modela. Metrike su ključne za evaluaciju performansi modela, a mi ćemo koristiti sledeće:

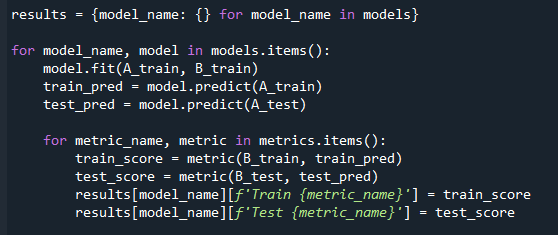
**Accuracy (Tačnost)**: Udeo tačno predviđenih primera u ukupnom broju primera.

**Precision (Preciznost)**: Udeo tačno predviđenih pozitivnih primera u ukupnom broju predikovanih pozitivnih primera.

**Recall (Odziv)**: Udeo tačno predviđenih pozitivnih primera u ukupnom broju stvarnih pozitivnih primera.

**F1 Score**: Harmonična sredina preciznosti i odziva, korisna kod neuravnoteženih klasa.

**Specificity (Specifičnost)**: Udeo tačno predviđenih negativnih primera u ukupnom broju stvarnih negativnih primera.



*Slika 26 – treniranje modela i izračunavanje metrika za svaki model*

**model.fit(A\_train, B\_train)** svaki model se trenira na trening podacima. Ovaj metod uči modelu kako da poveže ulazna obeležja (A\_train) sa ciljnom varijablom (B\_train).

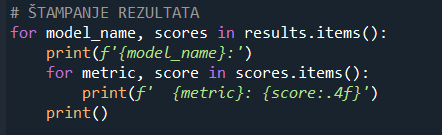
**model.predict()** nakon što je model obučen, koristi se za predikciju vrednosti na trening i test skupovima.

Za svaki model se izračunavaju sve metrike kako za trening, tako i za test podatke, i čuvaju se u rečniku results.

Svaki model se trenira na trening skupu, i predikcije se prave za trening i test skupove. Za svaku metriku (tačnost, preciznost, odziv, F1 skor, specifičnost) računaju se vrednosti za trening i test skupove i čuvaju se u rečniku results.

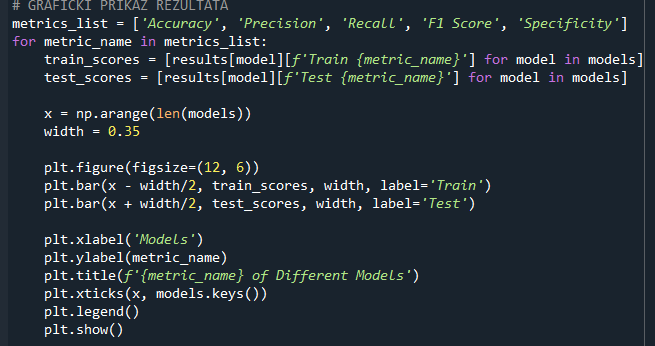
* 1. **Rezultati**

Nakon treniranja modela, prikazuju se rezultati za sve modele.



*Slika 27 – štampanje rezultata*

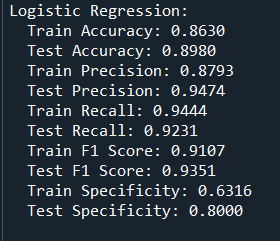
Prikazuju se rezultati za sve modele za svaku metriku (tačnost, preciznost, odziv, F1 skor, specifičnost) za trening i test skupove.



*Slika 28 – grafički prikaz rezultata*

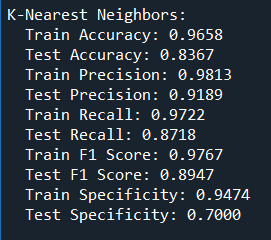
Kreiraju se bar grafici za svaku metriku, prikazujući rezultate za trening i test skupove za sve modele.

Nakon treniranja modela, analiziramo njihove performanse koristeći nekoliko metrika, uključujući tačnost, preciznost, odziv, F1 skor i specifičnost. Rezultati pokazuju kako se modeli ponašaju na trening i test skupovima. Važno je naglasiti da dobra performansa na trening skupu ne garantuje dobru generalizaciju, odnosno performansu na test podacima.



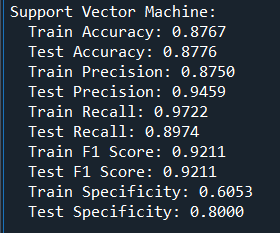
*Slika 29 – rezultati modela logistčke regresije*

Logistic Regression pokazuje odlične rezultate na test skupu sa tačnošću od 89.80% i visokim F1 skorom (93.51%). Odziv je takođe visok, što znači da model uspešno detektuje većinu pozitivnih primera (osoba sa Parkinsonovom bolešću). Međutim, specifičnost je nešto niža (80%), što ukazuje na to da model ne prepoznaje dovoljno dobro negativne primere (zdrave pacijente).



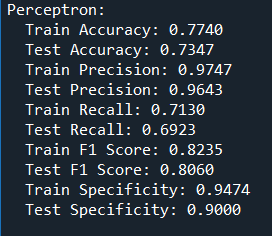
*Slika 30 – rezultati KNN modela*

KNN pokazuje visoku tačnost na trening podacima (96.58%), ali znatno manju na test podacima (83.67%). Ovo može ukazivati na overfitting modela na trening skup (overfitting). F1 skor je solidan, ali specifičnost je relativno niska (70%), što znači da model često greši kod negativnih primera.



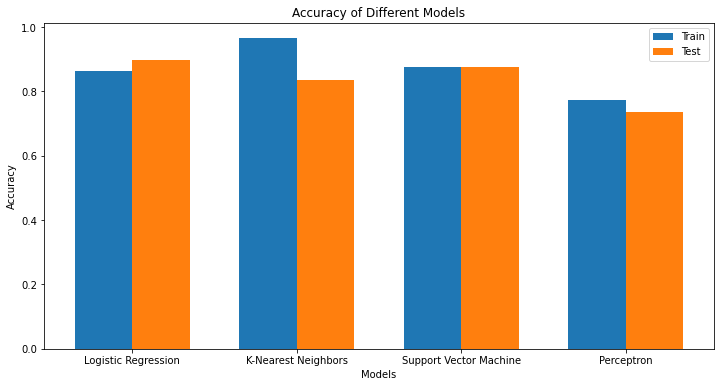
*Slika 31 – rezultati SVM modela*

SVM model pokazuje slične performanse na trening i test skupu, što znači da model dobro generalizuje i nije overfitovan. Tačnost na test podacima je 87.76%, sa odličnim F1 skorom (92.11%). Ipak, specifičnost je samo 80%, što znači da model može napraviti greške kod zdravih pacijenata.

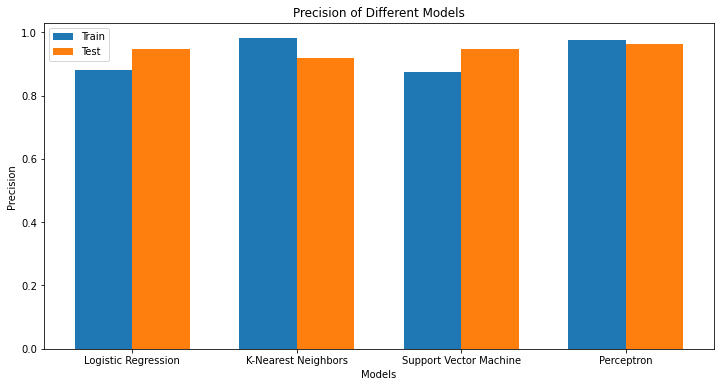


*Slika 32 – rezultati perceprton modela*

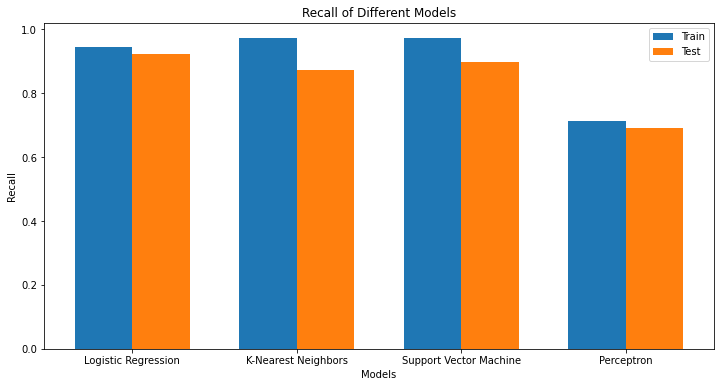
Perceptron pokazuje najnižu tačnost (73.47%) među svim modelima, ali iznenađujuće visoku specifičnost (90%) na test podacima. Međutim, model pokazuje slab odziv (69.23%), što znači da često ne uspeva da prepozna pozitivne primere (osobe sa Parkinsonovom bolešću).



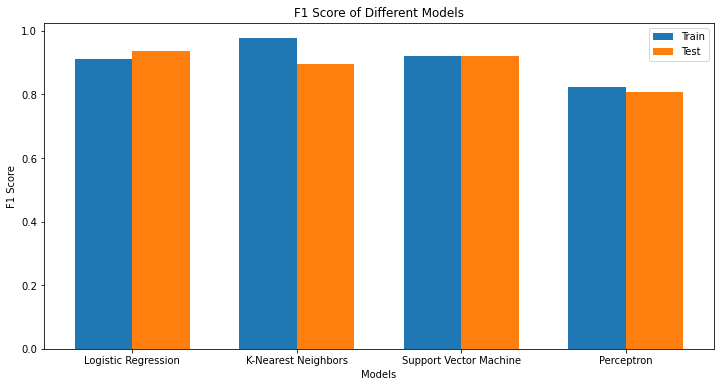
*Slika 33 – tačnost modela*



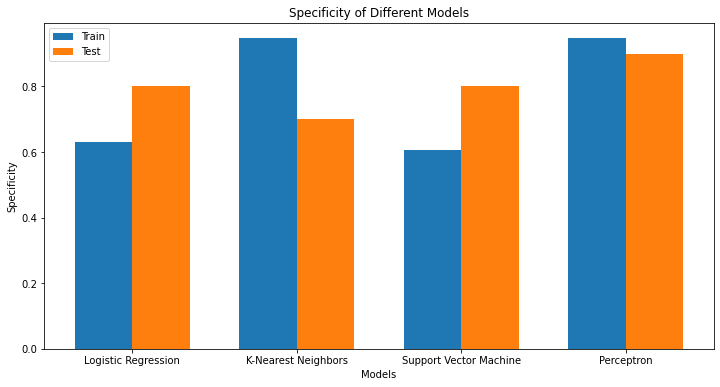
*Slika 34 – preciznost modela*



*Slika 35 – odziv modela*



*Slika 36 – F1 score*

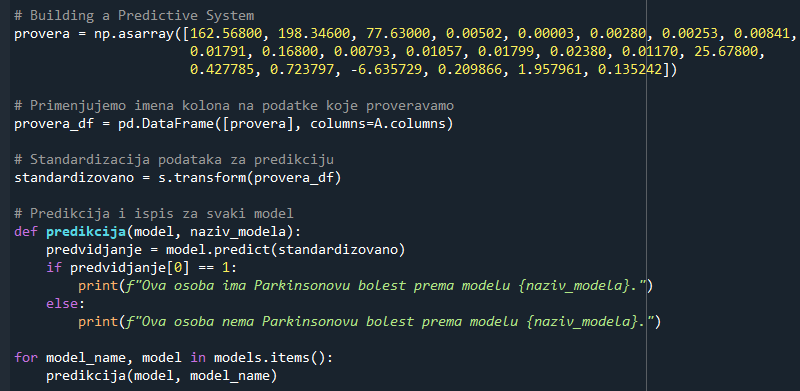


*Slika 37 – specifičnost modela*

Logistic Regression i Support Vector Machine (SVM) pokazuju visoku tačnost na test skupu. Međutim, SVM ima nešto bolji balans između preciznosti, odziva i F1 skora na test skupu, što ga čini boljim izborom. K-Nearest Neighbors (KNN) pokazuje vrlo visoku tačnost na trening skupu, ali nižu na test skupu, što može ukazivati na overfitting. Perceptron ima najnižu tačnost, ali pokazuje visoku specifičnost na test skupu.

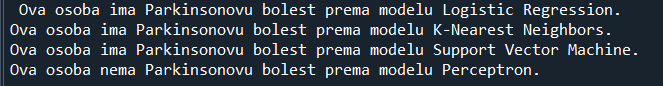
* 1. **Provera**

Nakon što smo obučili modele i analizirali njihove performanse na test podacima, sledeći korak je provera modela na novim, neviđenim podacima. Ova provera se vrši kako bismo videli kako se različiti modeli ponašaju kada dobiju nove podatke za klasifikaciju i da li mogu precizno predvideti da li osoba ima Parkinsonovu bolest ili ne. U ovom delu ćemo analizirati rezultate predikcija svakog modela na osnovu jednog primera.



*Slika 33 – provera funkcionalnosti modela na datom primeru*

Definiše se novi primer sa podacima koji se nalaze u provera. Novi primer se konvertuje u pandas DataFrame sa istim kolonama kao i obeležja u trening skupu. Podaci se standardizuju pomoću istog StandardScaler objekta koji je korišćen za trening podatke. Za svaki model se vrši predikcija na standardizovanim podacima i ispisuje se rezultat (da li osoba ima Parkinsonovu bolest ili ne).



*Slika 34 – ispis rezultata modela za dati primer*

Tri od četiri modela (Logistic Regression, KNN, i SVM) su predvidela da pacijent ima Parkinsonovu bolest. Ovi modeli su pokazali visoke vrednosti za metrike kao što su preciznost, F1 skor i odziv tokom evaluacije na test podacima, što znači da su uspešni u prepoznavanju osoba sa Parkinsonovom bolešću.

Perceptron, s druge strane, predviđa da pacijent nema Parkinsonovu bolest, što je u suprotnosti sa ostalim modelima. Kao što smo ranije videli, Perceptron ima lošije performanse u poređenju sa drugim modelima, posebno kada je u pitanju odziv i tačnost, pa je manje pouzdan u predikcijama. Ova greška može se pripisati njegovoj jednostavnosti i ograničenim sposobnostima u rešavanju složenijih klasifikacionih problema kao što je ovaj.

1. **Zaključak**

U ovom radu istražena je primena veštačke inteligencije za dijagnostiku Parkinsonove bolesti korišćenjem klasifikacionih algoritama. Razmatrani su različiti modeli mašinskog učenja: logistička regresija, K-Nearest Neighbors (KNN), Support Vector Machine (SVM) i Perceptron. Kroz implementaciju ovih algoritama na datasetu glasovnih karakteristika pacijenata sa i bez Parkinsonove bolesti, evaluirani su njihovi učinci koristeći metrike tačnosti, preciznosti, odziva, F1 skora i specifičnosti.

Na osnovu rezultata eksperimenta, zaključeno je da SVM algoritam pokazuje najbolji balans između tačnosti i ostalih metrika, dok je logistička regresija takođe pokazala vrlo dobre rezultate, ali nešto nižu specifičnost u poređenju sa SVM-om. KNN model je pokazao dobru tačnost na trening skupu, ali je bio sklon overfitting-u na test podacima. Perceptron, iako jednostavniji model, imao je najlošije rezultate u predikciji pozitivnih primera Parkinsonove bolesti, ali je pokazao relativno visoku specifičnost, što znači da je bolje prepoznavao zdrave pacijente.

Ovi rezultati ukazuju na značajan potencijal primene mašinskog učenja u ranom prepoznavanju Parkinsonove bolesti. Kombinacija različitih metrika, kao što su preciznost i odziv, omogućava dublju analizu performansi modela i pomaže u identifikaciji najboljeg pristupa za dijagnostiku. Iako su modeli poput SVM-a pokazali robusne performanse, dalje istraživanje bi moglo uključiti optimizaciju hiperparametara i proširenje dataseta kako bi se povećala tačnost predviđanja.

Implementacija ovih algoritama može značajno unaprediti medicinsku dijagnostiku, omogućavajući brže i preciznije prepoznavanje Parkinsonove bolesti na osnovu glasovnih uzoraka, čime se doprinosi ranijem započinjanju terapije i poboljšanju kvaliteta života pacijenata.

Buduća istraživanja bi mogla uključiti dodatne modele i tehnike kao što su duboko učenje ili kombinovani modeli, kao i integraciju drugih vrsta podataka, poput neuroloških snimaka, kako bi se poboljšala tačnost i pouzdanost sistema za dijagnostiku. Takođe, preporučuje se proširenje dataseta i optimizacija parametara kako bi se smanjila mogućnost overfitting-a modela i poboljšala generalizacija na neviđene podatke.

1. **Literatura**
2. Support Vector Machines for Machine Learning. (2023). <https://en.wikipedia.org/wiki/Bell_Labs>

<https://medium.com/@sumbatilinda/support-vector-machine-svm-algorithm-064566b5d411>

<https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm>

1. Logistic Regression Explained. (2023).

<https://en.wikipedia.org/wiki/Logistic_regression>

<https://www.geeksforgeeks.org/understanding-logistic-regression/>

<https://www.ibm.com/topics/logistic-regression>

1. K-Nearest Neighbors (KNN) Algorithm for Machine Learning. (2023).

<https://www.geeksforgeeks.org/k-nearest-neighbours/>

<https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm>

<https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning>

1. Perceptron. (2023).

<https://en.wikipedia.org/wiki/Perceptron>

<https://www.geeksforgeeks.org/what-is-perceptron-the-simplest-artificial-neural-network/>

<https://www.javatpoint.com/perceptron-in-machine-learning>

1. Evaluation Metrics for Classification Models

<https://medium.com/@mlmind/evaluation-metrics-for-classification-fc770511052d>

<https://cohere.com/blog/classification-eval-metrics>

1. Parkinson's Disease.

<https://www.rehaklinik-zihlschlacht.ch/en/?mtm_source=google&mtm_medium=cpc&mtm_campaign=zihlschlacht_int_en&mtm_kwd=parkinsons%20disease%20therapy&gad_source=1&gclid=CjwKCAjwxY-3BhAuEiwAu7Y6s6eZTwQDq0IXx0Xgg4TVZXbnpe-uTH5iU8wwLIIupeiH5O8zJgHvfhoC8PYQAvD_BwE>

<https://en.wikipedia.org/wiki/Parkinson%27s_disease>

1. Jitter – neželjena devijacija od prave periodičnosti pretpostavljenog signala u elektronici i telekomunikacijama [↑](#footnote-ref-1)