

Raspoznavanje aktivnosti životinje na temelju podataka s nosivih senzora

Kikić, Ivan

Undergraduate thesis / Završni rad

2021

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Faculty of Electrical Engineering, Computer Science and Information Technology Osijek / Sveučilište Josipa Jurja Strossmayera u Osijeku, Fakultet elektrotehnike, računarstva i informacijskih tehnologija Osijek**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:200:129799>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-11-20**

Repository / Repozitorij:

[Faculty of Electrical Engineering, Computer Science and Information Technology Osijek](#)



**SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU
FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA I
INFORMACIJSKIH TEHNOLOGIJA**

Sveučilišni preddiplomski studij računarstva

**RASPOZNAVANJE AKTIVNOSTI ŽIVOTINJE NA
TEMELJU PODATAKA S NOSIVIH SENZORA**

Završni rad

Ivan Kikić

Osijek, 2021.

SADRŽAJ

| | |
|--|-----------|
| 1. UVOD | 1 |
| 1.1. Zadatak završnog rada | 1 |
| 2. STROJNO UČENJE U POLJOPRIVREDI | 2 |
| 2.1. Nadzirano i nenadzirano učenje | 2 |
| 2.2. Klasifikacija | 3 |
| 2.2.1. Mjerenje kvalitete klasifikacije | 4 |
| 2.2.2. Deskriptivna statistika | 6 |
| 2.2.3. Kutijasti dijagram | 6 |
| 2.3. Često korišteni klasifikatori | 7 |
| 2.3.1. Algoritam K najbližih susjeda | 7 |
| 2.3.2. Logistička regresija | 8 |
| 2.3.3. Stablo odluke | 9 |
| 2.3.4. Nasumična šuma | 10 |
| 2.4. Unakrsna provjera valjanosti | 11 |
| 2.4.1. Metoda zadržavanja | 11 |
| 2.4.2. K validacija križnog preklapanja | 12 |
| 2.5. Primjena strojnog učenja u poljoprivredi | 12 |
| 2.5.1. Predviđanje prinosa | 13 |
| 2.5.2. Detekcija bolesti | 13 |
| 2.5.3. Prepoznavanje korova | 14 |
| 2.5.4. Sustav navodnjavanja | 14 |
| 2.5.5. Stočarska proizvodnja | 15 |
| 2.5.6. Raspoznavanje aktivnosti životinja | 15 |
| 3. RASPOZNAVANJE AKTIVNOSTI ŽIVOTINJE NA TEMELJU PODATAKA NOSIVIH SENZORA | 16 |
| 3.1. Priprema podatkovnog skupa | 16 |
| 3.2. Metodologija testiranja i opis eksperimenta | 18 |
| 3.3. Usporedba rezultata testiranja | 18 |
| 3.3.1. Testiranje na skupu podataka tri različite životinje | 19 |
| 3.3.2. Testiranje na spojenim skupovima podataka različitih životinja | 22 |
| 4. Zaključak | 25 |

1. UVOD

U današnje vrijeme strojno učenje je još uvijek nedovoljno istražena područje istraživanja. Ljudi je često povezuju samo s apstraktnim programima, uređajima i slično, dok je strojno učenje sve više prisutno u svakodnevnom životu. Njegov rad i dubinska analiza podataka uvelike olakšava svakodnevni rad u mnogim poslovima. Među njima nalazi se i poljoprivreda. Problemi u poljoprivredi gdje bi strojno učenje značajno pomoglo su vrlo česti. Poljoprivrednici se svakodnevno susreću sa bolestima usjeva, bolestima životinja, korovima i raznim drugim problemima u kojima bi strojno učenje dalo brzo i efikasno rješenje problema. Primjena strojnog učenja u poljoprivredi sve je više, a raste i broj korisnika ovakvih pristupa. Sve više poljoprivrednika u svakodnevni rad uključuje alata i sustave koji se, barem u određenoj mjeri, oslanjaju na postupke strojnog učenja. U ovom je radu razmatran problem raspoznavanja aktivnosti životinje. Znanjem o tome što pojedina životinja na nekom gospodarstvu radi u danom trenutku, kao i promatranjem obrazaca njihovih aktivnosti moguće je dobiti bolji uvid u ponašanje životinja te kreirati planove ili donositi različite odluke vezane uz uzgoj.

U drugom poglavlju rada opisuje se strojno učenje, navode se podjele strojnog učenja te njegove glavne karakteristike. U trećem poglavlju opisuje se strojno učenje u poljoprivredi, navode se problemi kojima su potrebna rješenja u vidu postupaka strojnog učenja te stvarne primjene strojnog učenja u poljoprivredi. U četvrtom poglavlju opisan je podatkovni skup generiran nosivim sensorima te je isti iskorišten kako bi se odredila trenutna aktivnost životinje uporabom klasifikacijskog modela. Posljednje poglavlje donosi zaključak i smjernice za budući rad.

1.1. Zadatak završnog rada

U teorijskom dijelu rada opisati mogućnosti korištenja postupaka strojnog učenja u poljoprivredi. Naglasak staviti na uporabu postupaka nadzirane klasifikacije na problemu identifikacije aktivnosti životinje. U praktičnom dijelu rada uporabom postupaka opisanih u teorijskom dijelu rada, a korištenjem otvorenog podatkovnog skupa dostupnog u literaturi ostvariti programsko rješenje za raspoznavanje aktivnosti životinje na temelju zapisa senzora postavljenih na različite životinje te ga prikladno testirati.

2. STROJNO UČENJE U POLJOPRIVREDI

Strojno učenje je grana umjetne inteligencije i računalne znanosti koja se usredotočuje na korištenje podataka i algoritama kako bi oponašali način na koji čovjek uči, postepeno poboljšavajući njegovu točnost [1]. Kako bi nešto bilo naučeno i predviđeno, potreban je model. Model je potrebno istrenirati na skupu podataka, dajući mu algoritam uz kojeg može učiti iz tih podataka. Nakon što je model istreniran, može biti korišten za predviđanje rezultata na podacima s kojima još nije radio [2]. Na primjer, prikupljeni su podaci za očitavanje bolesti biljaka. Model je zatim treniran na dijelu podataka tako što će pomoću raznih parametara vidjeti je li biljka bolesna ili ne. Nakon što je model naučen u mogućnosti je za dobivene druge, njemu nepoznate podatke, dati rješenje je li biljka bolesna.

2.1. Nadzirano i nenadzirano učenje

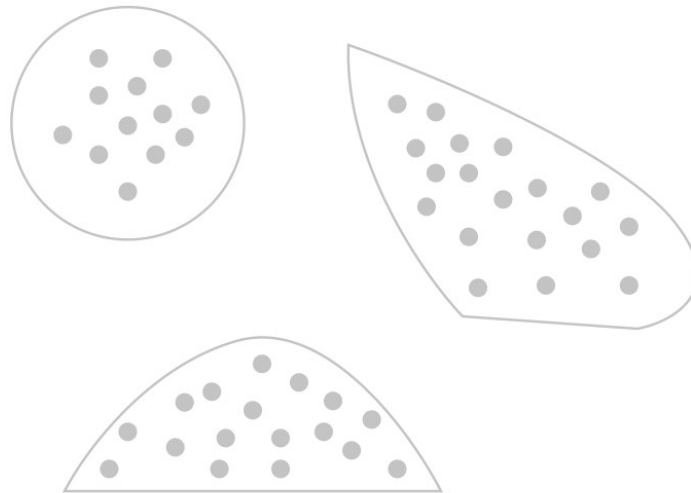
Nadzirano učenje (engl. *Supervised learning*) je vrsta učenja gdje računalni program uči s dva skupa podataka, skupom za treniranje (engl. *Training set*) i skupom za testiranje (engl. *Test set*). Ideja je da uz podatkovni skup algoritam trenira i uči kako bi kasnije mogao na podatkovnom skupu za testiranje razdvojiti i rasporediti s najvećom mogućom točnošću kojoj skupini koji podatak pripada [3]. Slika 2.1. prikazuje tri hrpe voća (podatkovnih skupova). Prva dva skupa služe kao skupovi za treniranje, na kojem će se algoritam istrenirati i kako bi kasnije mogao na trećem skupu odrediti i rasporediti što su jabuke a što breskve.



Slika 2.1. Nadzirano učenje na primjeru voća [3]

Učenje bez nadzora koristi algoritme strojnog učenja za analizu i grupiranje neoznačenih skupova podataka. Grupiranje (engl. *Clustering*) je zadatak podjele podatkovnih točaka u brojne grupe tako da su podatkovne točke u istim skupinama sličnije drugim podatkovnim točkama u istoj skupini za razliku od onih u drugim skupinama. Jednostavnije rečeno, cilj je izdvojiti grupe sa sličnim

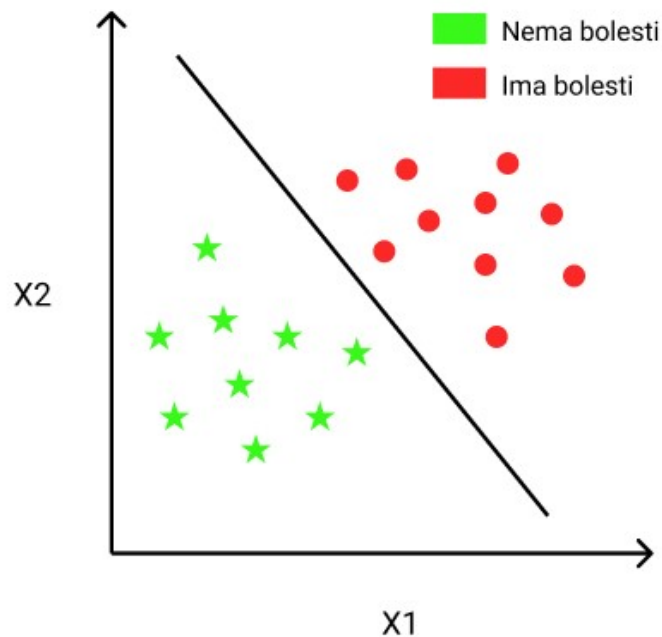
karakteristikama i rasporediti ih u klustere [4]. Primjer grupiranja je vidljiv na slici 2.2. U ovom trenutku nije bitno što je u grupama, najbitnija je sama podjela podatkovnih točaka u grupe po nekim određenim karakteristikama.



Slika 2.2. Učenje bez nadzora

2.2. Klasifikacija

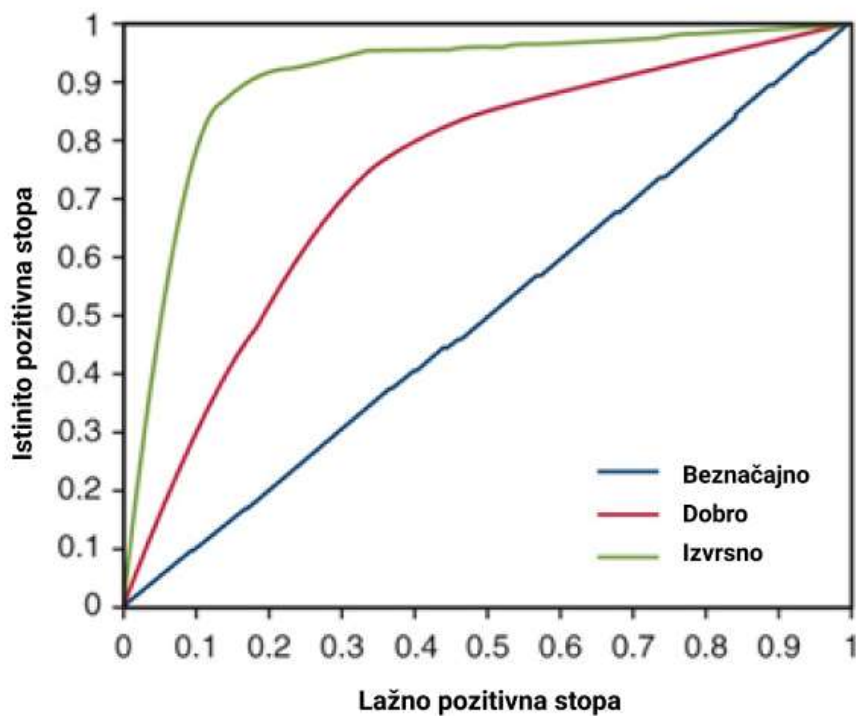
Klasifikacija je proces kategorizacije danog skupa podataka u unaprijed poznat broj klasa. Klase se često nazivaju ciljem, oznakom ili kategorijama. Klasifikator je glavni pojam klasifikacije, a on predstavlja algoritam koji se koristi za razvrstavanje ulaznih podataka u određenu kategoriju, kada dođu novi podaci, cilj je odrediti kojoj klasi pripadaju [5]. Rad klasifikatora se vrlo lako može objasniti na već spomenutom primjeru bolesti biljaka. U ovom primjeru radi se o binarnoj klasifikaciji jer mogu postojati samo dvije klase, biljka je bolesna ili nije. Potrebno je istrenirati klasifikator. Nakon što se klasifikator istrenirao na određenom skupu podataka može se koristiti za otkrivanje postoji li bolest kod biljke ili ne. Kako dolaze novi podaci, on će ih klasificirati i dodijeliti ih u skupinu kojoj pripada, tj. pripada li bolesti ili ne (Slika 2.3.).



Slika 2.3. Binarna klasifikacija na primjeru bolesti biljaka

2.2.1. Mjerenje kvalitete klasifikacije

Vrlo bitan pojam klasifikacije je operativna karakteristika prijavnika (engl. *Receiver operating characteristic curve, ROC*) ili tzv. ROC krivulja. To je grafički prikaz koji ilustrira dijagnostičku sposobnost binarnog klasifikacijskog sustava s promjenom praga diskriminacije. ROC krivulja stvara se iscrtavanjem prave pozitivne stope (engl. *True positive rate, TPR*) i lažno pozitivne stope (engl. *False positive rate, FPR*) pri različitim postavkama praga [6]. Kvalitetu ROC krivulje određuje površina ispod iste. Krivulja je bolja kada je površina veća. Na slici 2.4. su vidljive tri krivulje. Plava krivulja je najlošija, njena je površina oko 0.5 odnosno 50%, što predstavlja postotak slučajno odabranih parova koji su ispravno klasificirani. Crvena krivulja je puno bolja od plave, njoj je krivulja 0.75, odnosno 75% točnosti, dok zelena pokazuje izvrsnu ROC krivulju sa površinom blizu 1.



Slika 2.4. Uspoređivanje ROC krivulja

Još jedan način mjerenja kvalitete klasifikacije je matrica zabune. Matrica zabune (engl. *Confusion matrix*) je skup rezultata klasifikacije. To je broj netočnih i točnih predviđanja [7]. Na slici 2.5. je vidljivo da matrica zabune ima dva retka i dva stupca:

- IP - Istinito pozitivan (engl. *True positive*)
- LP - Lažno pozitivan (engl. *False positive*)
- LN - Lažno negativan (engl. *False negative*)
- IN - Istinito negativan (engl. *True negative*)

| | | | |
|---------|-----------|------------|-----------|
| Stvarno | Pozitivno | IP | LN |
| | Negativno | LP | IN |
| | | Pozitivno | Negativno |
| | | Predviđeno | |

Slika 2.5. Matrica zabune

2.2.2. Deskriptivna statistika

Deskriptivna statistika bavi se mjerama centralne tendencije (aritmetička sredina, medijan i mod), mjerama varijabiliteta, kao što su standardna devijacija, varijanca i drugi [8]. Također, to su kratki opisni koeficijenti koji sažimaju dati skup podataka, koji mogu biti ili prikaz cijele ili samo uzorak populacije. U radu su korišteni sljedeći:

- Srednja vrijednost
- Standardna devijacija
- Minimum
- Maksimum
- Medijan

2.2.3. Kutijasti dijagram

U deskriptivnoj statistici kutijasti dijagram (engl. *Box plot*) je metoda za grafičko prikazivanje podataka kroz njihove kvartile. Prikazuje pet točaka podataka. Prikazuje minimum, maksimum, medijan te prvi i treći kvartil. Na slici 2.6. su vidljive glavne karakteristike. Vanjske točke koje su udaljene okomitom linijom od same plohe predstavljaju minimum i maksimum. Rubovi plohe su prvi i treći kvartil dok je crta koja se proteže po sredini plohe medijan.



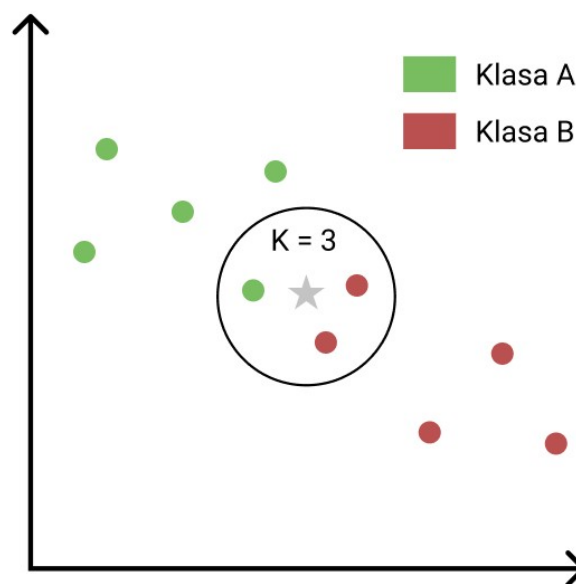
Slika 2.6. Prikaz glavnih karakteristika kutijastog dijagrama

2.3. Često korišteni klasifikatori

Klasifikator u strojnom učenju je algoritam koji automatski slaže ili kategorizira podatke u jednu ili više grupa, klasa. Jedan od najčešćih je klasifikator e-pošte koji skenira e-poštu i filtrira poruke prema oznaci klase kao što su normalna pošta ili neželjena pošta [9]. Neki od najčešćih klasifikatora su klasifikator K najbližih susjeda (engl. *K nearest neighbors*, KNN) ($N = 1$ ili $N = 5$), Stablo odluke (engl. *Decision Tree*), Logistička regresija (engl. *Logistic Regression*) te Nasumična šuma (engl. *Random Forest*).

2.3.1. Algoritam K najbližih susjeda

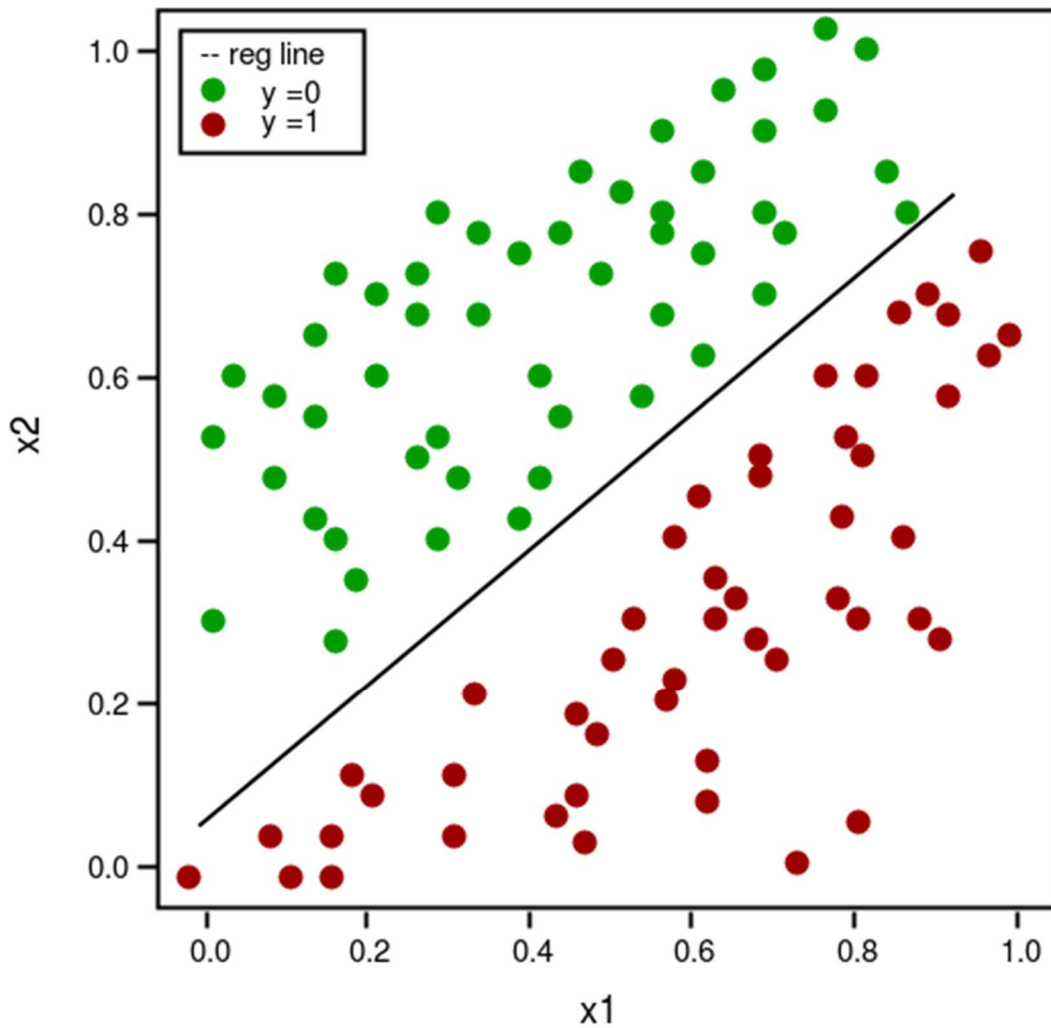
Algoritam K najbližih susjeda jednostavan je algoritam strojnog učenja u pravilu dobrih performansi. Često je upotrebljavan za klasifikaciju, ali ga je moguće rabiti i za regresiju. KNN klasificira nove unesene podatke na temelju njihove sličnosti s prethodno dostupnim podacima poznatih oznaka klasa, prema nekoj mjeri udaljenosti (najčešće Euklidska udaljenost) [10]. Na slici 2.7. vidljiv je primjer načina rada algoritma KNN. U ovom slučaju broj K je 3. K je srž ovog algoritma i on predstavlja broj najbližih susjeda koje se promatra te u odnosu na koje će na koncu donijeti odluku o klasi novog podatka. Vidljivo je da je siva zvijezda novi podatak kojemu će klasa biti odlučena na osnovu njegovih K susjeda, u ovom slučaju tri najbliža susjeda. Zaključak je da novi podatak pripada klasi B jer se u tri promatrana susjeda dvije točke crvene, a jedna zelena.



Slika 2.7. Primjer algoritma K najbližih susjeda

2.3.2. Logistička regresija

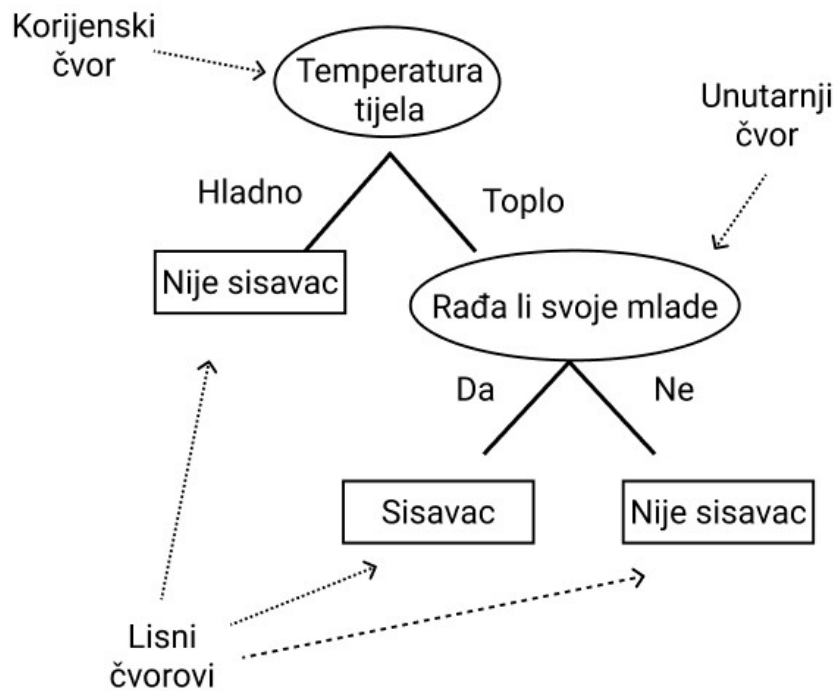
Logistička regresija (engl. *Logistic regression*, LR) jedno je od najvažnijih analitičkih alata u društvenoj i prirodnoj znanosti. Osnovni je nadzirani algoritam strojnog učenja za klasifikaciju, a ima i vrlo blizak odnos s neuronskim mrežama. [11]. LR se može podijeliti u dvije skupine, binarnu i multi-linearnu skupinu. Binarna LR predstavlja skupinu gdje su mogući ishodi binarni, kao na primjer detekcija tumora, tumor može biti zloćudan i benigan, dok multi-linearna LR predstavlja više klasa kao što su recimo tipovi životinja na farmi, koze, krave, ovce itd. Slika 2.8. predstavlja binarnu logističku regresiju. Na slici je vidljiva podjela dvije skupine, u ovom slučaju crvenih i zelenih točaka. Crta koja razdvaja ove dvije klase naziva se crta odluke (engl. *Decision Boundary*).



Slika 2.8. Binarna logistička regresija

2.3.3. Stablo odluke

Stablo odluke (engl. *Decision Tree*) klasifikator, jedan je od često korištenih jednostavnih tehnika klasifikacije [12]. Ovaj algoritam funkcionira tako što se uz pomoć pitanja i praga dolazi do binarnog rješenja, do jedne od dvije kategorije, na primjer pripada li životinja skupini sisavaca ili ne. Kada bi došla nova vrsta životinja, potrebno je odrediti kojoj od ove dvije kategorije ona pripada. Uz pomoć nekoliko jednostavnih pitanja dolazi se do rješenja. Slika 2.9. predstavlja primjer takvog stabla odluke. Prvo pitanje može biti, je li ta vrsta hladnokrvna ili toplokrvna, ako je hladnokrvna, odmah se može zaključiti da nije sisavac. U suprotnom, nova vrsta može biti ptica ili sisavac. Time dovodi u pitanje može li ta vrsta rađati svoje mlade, ako je odgovor da, onda je definitivno sisavac (ptica nese jaja, ne rađa).



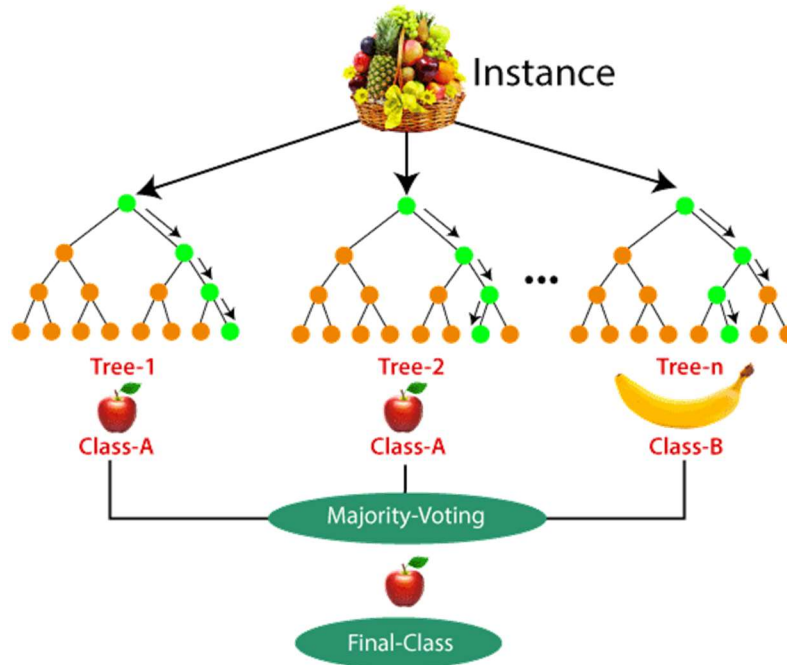
Slika 2.9. Stablo odluke sa glavnim karakteristikama

Stablo odluke sastoji se od tri vrste čvorova, a to su korijenski čvor, unutarnji čvor i list, odnosno lisni čvor. Korijenski čvor ima nula ulaznih rubova, dok može imati nula ili više izlaznih. U ovom slučaju vidljiva su dva izlazna ruba iz korijenskog čvora, a to su toplo i hladno. Unutarnji čvor predstavlja čvor koji ima točno jedan ulazni rub i dva ili više izlaznih rubova. Također, u ovom primjeru ima dva izlazna, da i ne. Treća vrsta čvorova su listovi, to su krajnji čvorovi koji imaju točno jedan ulazni i nijedan izlazni čvor jer predstavljaju kraj stabla. U ovom primjeru oni daju rješenja stabla odluke, a to su je li neka životinja sisavac ili ne.

2.3.4. Nasumična šuma

Nasumična šuma (engl. *Random Forest*) tehnika je strojnog učenja koja se koristi za rješavanje klasifikacijskih i regresijskih problema. Algoritam se sastoji mnogih stabala odluke. Uspostavlja ishod na temelju predviđanja stabala odluke. Zatim uzima prosjek predviđanja stabala i time dolazi do rješenja. Povećanjem broja stabala povećava se preciznost rezultata [13]. Slika 2.10. pokazuje primjer kada su samo tri stabla izabrana, što naravno nikad nije tako malen broj, ali za ovaj primjer će biti sasvim u redu da se lakše pojasni. Prvo stablo pokazuje jabuku kao odgovor, drugo isto jabuku, a treće bananu. Kao što je objašnjeno ranije, uzima se prosjek rezultata i dolazi se do

krajnjeg rezultata. Stabla su pokazala dvije jabuke i jednu bananu kao rješenje, tako da je krajnji rezultat jabuka.



Slika 2.10. Prikaz Random Forest algoritma [14]

2.4. Unakrsna provjera valjanosti

Unakrsna provjera valjanosti (engl. *Cross-validation*) tehnika je ponovnog uzorkovanja s temeljnom idejom podjele skupa podataka na 2 dijela, na podatke za treniranje i podatke za testiranje. Podaci za testiranje koriste se za treniranje modela, a podaci za testiranje su dosad neviđeni podaci za model i služe za predviđanje [15]. Postoji nekoliko metoda vrednovanja modela, a to su metoda zadržavanja (engl. *Hold out method*), k validacija križnog preklapanja (engl. *K-fold cross-validation*), stratificirana k validacija križnog preklapanja (engl. *Stratified k-fold cross-validation*) i druge.

2.4.1. Metoda zadržavanja

Metoda zadržavanja je najjednostavnija metoda vrednovanja i naširoko se koristi u projektima strojnog učenja [14]. Radi tako da su podaci podijeljeni u dvije skupine, odnosno u dva seta podataka, podaci za treniranje i podaci za testiranje. Podaci se mogu podijeliti na više načina na 70-30, 60-40, 80-20, 50-50 i razne druge kombinacije. Podjela podataka na omjere se događa nasumično, nije moguće znati koji će podaci završiti u setu za treniranje, a koji za testiranje.

2.4.2. K validacija križnog preklapanja

U ovoj metodi ponovnog uzorkovanja cjelokupni su podaci podijeljeni u k skupova gotovo jednakih veličina. Prvi je set podataka odabran kao testni skup podataka i model treniran na ostalim k-1 skupovima. Stopa pogreške pri testiranju se izračunava nakon što se model prilagodi ispitnim podacima. U drugom ponavljanju, drugi skup je odabran kao testni skup i preostali k-1 skupovi koriste se za treniranje podataka. Ovaj proces se izvršava za sve k skupove [15].

2.5. Primjena strojnog učenja u poljoprivredi

Vjerojatno je vrlo jasno da strojno učenje kao i umjetna inteligencija počinje biti korisna u raznim područjima i izvan same tehnologije. Strojno učenje se primjenjuje u svim uspješnim poduzećima, kojima olakšava razne odluke i svakodnevne probleme. Primjenom strojnog učenja vlasnici raznih poljoprivrednih poduzeća mogu pristupiti raznim alatima za podatke, analitiku i statistiku, što omogućuje bolje i kvalitetnije odluke, poboljšanu učinkovitost i smanjenje otpada.

Slika 2.11. prikazuje da je strojno učenje sve popularnije, da postaje sve primjenjenije u poljoprivredi. Broj objavljenih radova govori o tome da postoje mnoge dobre stvari strojnog učenja u ovoj grani pa se potreba i područja primjene s godinama povećavaju.



Slika 2.11. Broj objavljenih radova na temu strojnog učenja u poljoprivredi [13]

Strojno učenje u poljoprivredi uvelike pomaže malim i velikim poljoprivrednicima obavljanju njima teških ili jednostavno monotonih poslova koje bi jedan algoritam vrlo brzo i precizno odradio. Podskup umjetne inteligencije, strojno učenje, ima znatan potencijal za rješavanje brojnih izazova u uspostavi poljoprivrednih sustava [15]. Razvoj strojnog učenja u posljednjih dvadesetak godina doveo je do značajnog povećanja broja projekata u području poljoprivrede [13].

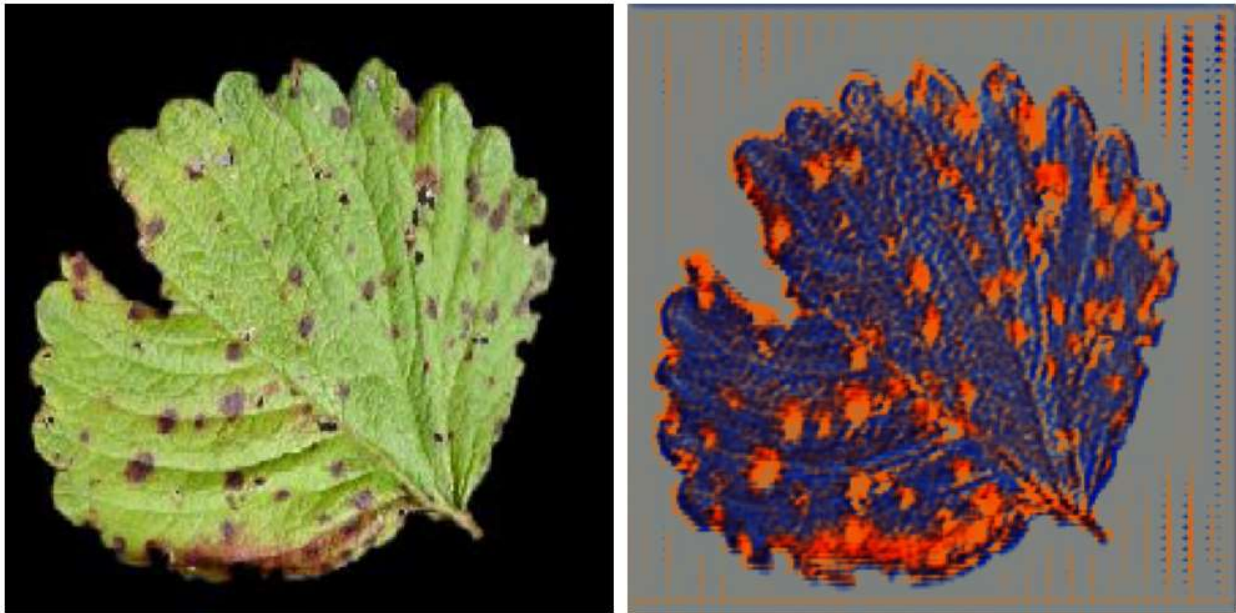
Usredotočeni su na predviđanje parametara tla, predviđanje prinosa usjeva, otkrivanje bolesti biljaka. Postoje razne primjene koje zahtijevaju računalni vid s kojim pregledavaju različite slike usjeva kako bi se pratila kvaliteta usjeva i procjena prinosa. Ne treba izostaviti da je vrlo značajna primjena i u stočarskoj industriji, predviđanje plodnosti, dijagnosticiranje poremećaja prehrane, prepoznavanje aktivnosti životinja na temelju podataka dobivenih senzorima itd.

2.5.1. Predviđanje prinosa

Predviđanje prinosa jedna je od najvažnijih tema u modernoj poljoprivredi. Vrlo točan i precizan model može pomoću na primjer farmerima da donesu bitne odluke o tome što saditi i na što se najviše fokusirati prema zahtjevima tržišta. Predviđanje prinosa nije toliko jednostavan i beznačajan zadatak. Neki bitni činjenici su okoliš, genotipske i fenotipske karakteristike usjeva i slično [15]. To zahtijeva temeljno razumijevanje odnosa između interaktivnih čimbenika i prinosa.

2.5.2. Detekcija bolesti

Bolesti usjeva predstavljaju veliku prijetnju sustavima poljoprivredne proizvodnje koji pogoršavaju kvalitetu i količinu prinosa. Izvješća o gubicima prinosa zbog biljnih bolesti vrlo su česta na farmi. Bolesti usjeva također predstavljaju velik rizik za sigurnost hrane na globalnoj razini. Pravovremena reakcija na bolesti vrlo su bitna za gospodarstvo. Nepravovremenom reakcijom mogu se izazvati razne bolesti poput bakterija, gljivica, virusa i drugih. Prisutnost bolesti može se sastojati od pjegavosti lista i ploda, uvenuća i promjene boje [14]. Slika 2.12. prikazuje jedan takav primjer bolesti gdje se vidi značajna razlika u boji i pojavljivanje pjega na listu. Tehnologija je uznapredovala te je time u mogućnosti otkriti bolesti biljaka prije nego simptomi postanu vidljivi. Najveća detekcija bolesti biljaka dolazi preko računalnog vida. Klasifikatorima slika potreban je veliki skup podataka zdravih i oboljelih biljaka.



Slika 2.12. Prepoznavanje bolesti biljke [16]

2.5.3. Prepoznavanje korova

Kao rezultat njihove plodne proizvodnje sjemena i dugovječnosti, korovi najčešće rastu i vrlo brzo se šire na velike dijelove polja. Korovi često nastaju prije usjeva što im omogućuje nesmetan rast, a time otežava i negativno utječe na rast usjeva. Kako bi se suzbio rast korova potrebno je primijeniti herbicide. Međutim, pokazalo se da su velike količine herbicida skupe i štetne za okoliš. Sve je izglednije da će velika uporaba herbicida vrlo vjerojatno učiniti korov otpornijim, što rezultira zahtjevnijom i skupljom kontrolom korova [14]. U posljednjih nekoliko godina tehnologija sa ciljem suzbijanja korova je uznapredovala. Koriste se razni senzori, kamere, kopnena vozila, dronovi i slično kako bi što bolje razlikovali korijenje od usjeva i kako bi što efikasnije suzbili širenje korova.

2.5.4. Sustav navodnjavanja

Poljoprivredni sektor glavni je potrošač dostupne slatke vode budući da se rast biljaka bazira na dostupnosti vode. Učinkovito upravljanje vodama može dovesti do poboljšanja kvalitete vode kao i smanjenja zagađenja. Nova istraživanja nude sustav navodnjavanja s promjenjivom stopom kako bi se postigla što bolja ušteda vode [14]. Time bi se voda uvelike sačuvala, jer ne bi dolazilo do jednake stope na svim poljima. Ako neko polje zahtjeva manje vode to se stopom može regulirati i neće biti potrebe za neizmjernom potrošnjom. Također, praćenje vode u tlu, uvjete rasta usjeva

u kombinaciji sa praćenjem vremenskim uvjeta može pomoći u pametnijem navodnjavanju polja i uštedi vode.

2.5.5. Stočarska proizvodnja

Strojno učenje se može koristiti za praćenje raznih informacija vezanih za stoku. Pomaže optimizirati proizvodni sustav mlijeka, jaja, mesa i drugih stvari kroz točno predviđanje i procjenu poljoprivrednih parametara. Istraživači su razvili zanimljivo rješenje za identifikaciju svinja bez korištenja radio-frekvencijskih identifikacijskih oznaka, koje imaju ograničen raspon i zahtijevaju stresne postupke za životinje.

2.5.6. Raspoznavanje aktivnosti životinja

Poznato je da na produktivnost životinja može utjecati njihova zadovoljnost, sreća i slično. Tako je osmišljen algoritam koji pomoću raznih senzora dolazi do parametara potrebnih za očitavanje stanja životinje. Rezultati mogu biti stajanje, trčanje, jedenje, pijeње i drugo. Takvim se načinom može znati dosta više o samim životinjama i kako im ugoditi kako bi produktivnost bila bolja.

3. RASPOZNAVANJE AKTIVNOSTI ŽIVOTINJE NA TEMELJU PODATAKA NOSIVIH SENZORA

Cilj ovog završnog je razmatranje problema raspoznavanja aktivnosti životinja i usporedba performansi različitih klasifikacijskih modela. Pomoću prikupljenih podataka sa senzora, obradom podataka, uporabom raznih klasifikatora i algoritama, dolazi se do rezultata koji se detaljno obrađuju i komentiraju kroz razne tablice i dijagrame.

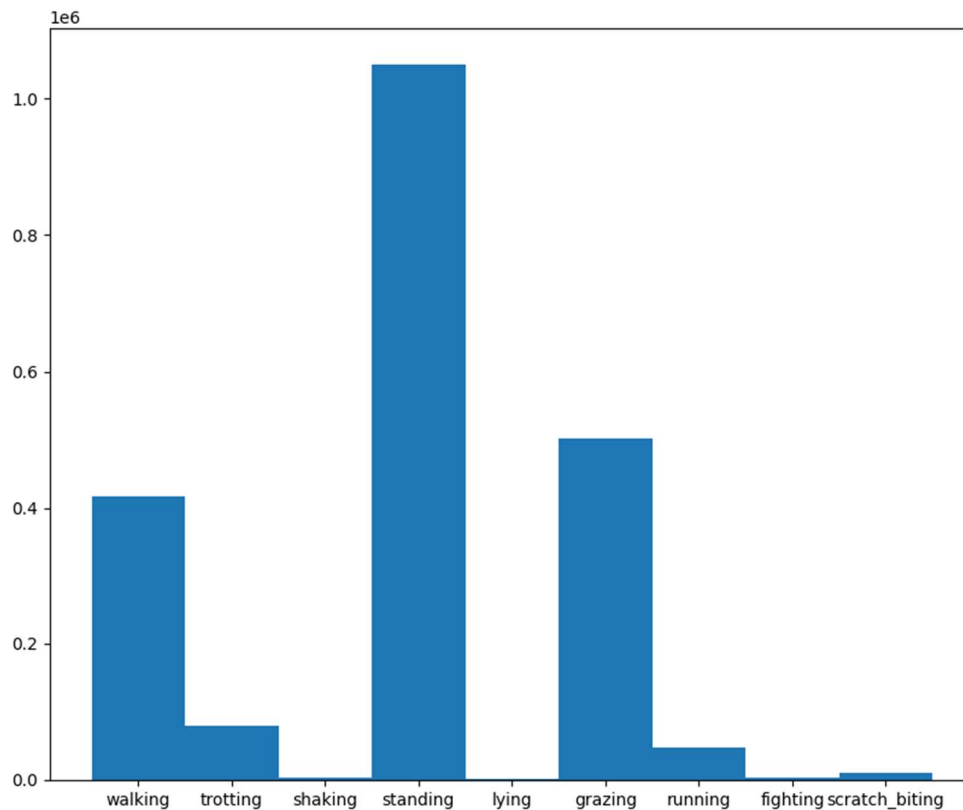
3.1. Priprema podatkovnog skupa

Kako bi se stvorio uspješan model strojnog učenja, imperativ je da se mora dobro obučiti, testirati i potvrditi prije nego što se primjeni u proizvodnji [17]. Prvi korak do uspješnog rada sa klasifikatorima i algoritmima je prikupljanje podataka. Podaci se mogu prikupiti iz raznih načina. Česti primjer prikupljanja podataka u poljoprivredi, kao i u ovom radu su senzori, žiroskopi, analizatori zvuka, senzori za znoj, slinu i slično. Podaci se pohranjuju u datoteke ekstenzije.csv. Slika 3.1. prikazuje jedan takav primjer datoteke.

| 1 | label | animal_ID | segment | timestam | ax | ay | az | axhg | ayhg | azhg | cx | cy | cz | gx | gy | gz | pressure | temp |
|----|---------|-----------|---------|----------|----------|---------|----------|----------|---------|---------|---------|---------|--------|----------|----------|---------|----------|-------|
| 2 | walking | G1 | 1 | 1 | 1.57538 | 4.34787 | -9.27514 | 5.61615 | 8.19897 | 10.2112 | -0.5685 | -0.7875 | 2.5245 | 10.5488 | -28.4756 | 91.9512 | NaN | 28.16 |
| 3 | walking | G1 | 1 | 6 | 1.47962 | 4.30477 | -9.31105 | 4.98546 | 8.64947 | 4.4749 | NaN | NaN | NaN | 9.5122 | -27.0732 | 88.1707 | NaN | 28.15 |
| 4 | walking | G1 | 1 | 11 | 1.36469 | 4.24492 | -9.42118 | 5.19569 | 9.25012 | 5.79634 | -0.585 | -0.783 | 2.526 | 9.57317 | -25.2439 | 83.9634 | NaN | 28.16 |
| 5 | walking | G1 | 1 | 16 | 1.21386 | 4.22816 | -9.59835 | 4.23463 | 8.82966 | 9.28016 | NaN | NaN | NaN | 11.2195 | -22.9878 | 79.2683 | NaN | 28.16 |
| 6 | walking | G1 | 1 | 21 | 1.07021 | 4.2952 | -9.67257 | 5.13562 | 7.89864 | 5.88644 | -0.5835 | -0.7815 | 2.5215 | 13.4756 | -19.878 | 73.8415 | NaN | 28.14 |
| 7 | walking | G1 | 1 | 26 | 0.921768 | 4.40054 | -9.73004 | 5.40592 | 9.49039 | 6.99766 | NaN | NaN | NaN | 15.5488 | -15.6098 | 67.6829 | NaN | 28.15 |
| 8 | walking | G1 | 1 | 31 | 0.730232 | 4.52744 | -9.72764 | 4.74519 | 12.2534 | 9.00986 | -0.5685 | -0.7575 | 2.5305 | 17.3171 | -10.9146 | 61.5244 | NaN | 28.16 |
| 9 | walking | G1 | 1 | 36 | 0.531513 | 4.69503 | -9.76834 | 6.6673 | 8.19897 | 6.72736 | NaN | NaN | NaN | 18.1098 | -6.28049 | 55.3049 | 1008.22 | 28.16 |
| 10 | walking | G1 | 1 | 41 | 0.304064 | 4.85305 | -9.78031 | 3.09339 | 8.64947 | 8.91976 | -0.5835 | -0.7665 | 2.5335 | 17.6829 | -2.9878 | 49.3902 | NaN | 28.15 |
| 11 | walking | G1 | 1 | 46 | 0.098162 | 4.9129 | -9.82102 | 4.26467 | 8.52933 | 8.4092 | NaN | NaN | NaN | 16.2195 | -0.79268 | 43.5976 | NaN | 28.16 |
| 12 | walking | G1 | 1 | 51 | -0.01197 | 4.90093 | -9.94551 | 0.360394 | 11.7729 | 8.13891 | -0.5835 | -0.7815 | 2.5275 | 14.2683 | 0.670732 | 38.1098 | NaN | 28.15 |
| 13 | walking | G1 | 1 | 56 | -0.07183 | 4.88178 | -9.94312 | 4.26467 | 8.22901 | 8.73956 | NaN | NaN | NaN | 12.622 | 2.0122 | 33.8415 | NaN | 28.14 |
| 14 | walking | G1 | 1 | 61 | -0.14844 | 4.88896 | -9.90721 | 5.04552 | 8.55937 | 5.52605 | -0.5925 | -0.7485 | 2.5275 | 10.7317 | 3.29268 | 30.9756 | NaN | 28.15 |
| 15 | walking | G1 | 1 | 66 | -0.20111 | 4.88417 | -9.77313 | 2.88316 | 11.2323 | 5.76631 | NaN | NaN | NaN | 8.10976 | 4.5122 | 29.6341 | NaN | 28.17 |
| 16 | walking | G1 | 1 | 71 | -0.29449 | 4.84108 | -9.64863 | 3.84421 | 10.2712 | 6.54716 | -0.573 | -0.768 | 2.523 | 5.12195 | 5.12195 | 29.2683 | NaN | 28.16 |
| 17 | walking | G1 | 1 | 76 | -0.35195 | 4.73573 | -9.46428 | 3.75411 | 10.3914 | 5.37588 | NaN | NaN | NaN | 1.46341 | 5.06098 | 29.8171 | 1008.2 | 28.19 |
| 18 | walking | G1 | 1 | 81 | -0.34477 | 4.64954 | -9.22965 | 3.33365 | 9.58048 | 6.90756 | -0.576 | -0.756 | 2.532 | -2.31707 | 4.5122 | 30.9756 | NaN | 28.15 |
| 19 | walking | G1 | 1 | 86 | -0.25379 | 4.52265 | -8.94234 | 4.2947 | 10.0009 | 7.68841 | NaN | NaN | NaN | -5.79268 | 3.59756 | 32.561 | NaN | 28.15 |
| 20 | walking | G1 | 1 | 91 | -0.14126 | 4.41251 | -8.63828 | 3.60394 | 7.32802 | 5.31582 | -0.5835 | -0.7725 | 2.5155 | -8.90244 | 2.13415 | 34.6341 | NaN | 28.15 |
| 21 | walking | G1 | 1 | 96 | -0.01197 | 4.30238 | -8.34619 | 4.89536 | 6.99766 | 6.06664 | NaN | NaN | NaN | -11.2805 | 0.121951 | 37.378 | NaN | 28.16 |
| 22 | walking | G1 | 1 | 101 | 0.057461 | 4.18985 | -7.99424 | 3.48381 | 8.19897 | 9.40029 | -0.5835 | -0.7725 | 2.5155 | -12.7439 | -2.86585 | 40.9756 | NaN | 28.18 |
| 23 | walking | G1 | 1 | 106 | 0.083797 | 4.07014 | -7.77397 | 4.53496 | 8.91976 | 10.7818 | NaN | NaN | NaN | -13.4146 | -6.46341 | 45.4268 | NaN | 28.15 |
| 24 | walking | G1 | 1 | 111 | 0.098162 | 3.94086 | -7.56807 | 3.18348 | 8.6795 | 7.80855 | -0.576 | -0.741 | 2.532 | -13.1707 | -10.3049 | 50.6707 | NaN | 28.15 |
| 25 | walking | G1 | 1 | 116 | 0.131681 | 3.79481 | -7.4316 | 2.07227 | 8.37917 | 11.0521 | NaN | NaN | NaN | -12.0732 | -13.9024 | 56.7073 | 1008.18 | 28.14 |
| 26 | walking | G1 | 1 | 121 | 0.136469 | 3.63679 | -7.27837 | 1.59174 | 7.62835 | 8.22901 | -0.5925 | -0.7575 | 2.5095 | -10 | -17.1951 | 63.2927 | NaN | 28.15 |
| 27 | walking | G1 | 1 | 126 | 0.124498 | 3.5099 | -7.1443 | 2.55279 | 8.6795 | 7.11779 | NaN | NaN | NaN | -8.04878 | -20.2439 | 70 | NaN | 28.17 |
| 28 | walking | G1 | 1 | 131 | 0.100556 | 3.4285 | -6.99346 | 3.84421 | 11.5326 | 7.68841 | -0.5775 | -0.7485 | 2.5305 | -6.52439 | -22.2561 | 76.2195 | NaN | 28.15 |
| 29 | walking | G1 | 1 | 136 | 0.083797 | 3.39019 | -6.8953 | 2.0122 | 8.25904 | 8.73956 | NaN | NaN | NaN | -6.09756 | -22.9268 | 81.2805 | NaN | 28.18 |
| 30 | walking | G1 | 1 | 141 | 0.069432 | 3.35667 | -6.85221 | 4.62506 | 8.4993 | 11.0221 | -0.5865 | -0.7575 | 2.5185 | -6.21951 | -21.7073 | 85.2439 | NaN | 28.17 |
| 31 | walking | G1 | 1 | 146 | 0.064643 | 3.35667 | -6.88572 | 1.38151 | 7.02769 | 11.0821 | NaN | NaN | NaN | -6.28049 | -19.1463 | 88.1098 | NaN | 28.17 |
| 32 | walking | G1 | 1 | 151 | 0.052672 | 3.36864 | -7.16585 | 4.0244 | 7.20789 | 8.34914 | -0.5895 | -0.7395 | 2.5305 | -5.85366 | -15.4878 | 89.1463 | NaN | 28.16 |
| 33 | walking | G1 | 1 | 156 | -0.01676 | 3.45244 | -7.64708 | 3.81417 | 5.76631 | 7.92868 | NaN | NaN | NaN | -3.41463 | -11.0366 | 87.9268 | 1008.17 | 28.18 |

Slika 3.1. Prikaz podatkovnog skupa

Na slici su vidljivi razni podaci, prvi red prikazuje oznake (engl. *Label*) koje govore koji je položaj odnosno radnja životinje. Na slici je vidljiva oznaka hodanja (engl. *Walking*), dok postoje još i stajanje, ležanje, trčanje i druge. Zatim je vidljiva identifikacijska oznaka životinje. Slika 3.2. prikazuje omjer oznaka životinje, odnosno omjer izlaznih varijabli.



Slika 3.2. Histogram izlaznih veličina jednog podatkovnog skupa

Nakon toga slijede razni podaci očitani sa raznih senzora. Neki od njih su ax , ay , $axhg$, $ayhg$, tlak, temperatura itd. Kako bi model mogao učiti potrebno je prvo pripremiti podatkovni skup, na način da se razdvoje ulazne i izlazne karakteristike kao X i Y skupovi podataka. Za izlaznu karakteristiku uzima se prvi stupac, oznaka. Oznaka će na osnovu raznih podataka senzora predstavljati ponašanje životinje u određenim trenucima. Za ulaznu karakteristiku uzimaju se od trećeg do zadnjeg stupca, odnosno stupci podataka očitanih sa senzora. Mogući problem na koji se nailazi, kao što je i ovdje slučaj je moguć veliki broj NaN vrijednosti, odnosno vrijednosti koje možda nisu dobro očitane. Takve vrijednosti mogu dovesti do netočnih i nepreciznih rezultata te je takve podatke potrebno ukloniti, što je više moguće. Jedno od rješenja je da se uklone stupci i reci sa više od pedeset posto NaN vrijednosti time bi rezultati bili točniji. U ovom radu postoji više ovakvih podatkovnih skupova koji predstavljaju različite životinje. Podaci prikupljeni sa koze (engl. *Goat*, G), podaci prikupljeni sa ovce (engl. *Sheep*, S). S obzirom da postoji više skupova sa iste životinje podijeljeni su brojevima, kao na primjer $G1$ i $G2$ što predstavljaju dva različita

podatkovna skupa koze. Svi su podatkovni skupovi vrlo veliki, s preko dva milijuna podataka što dovodi do značajnih razlika u rezultatima eksperimenata.

3.2. Metodologija testiranja i opis eksperimenta

Primjenom metode zadržavanja testirani su skupovi i dobiveni rezultati. Metom zadržavanja podatkovni skup je podijeljen na omjer 70-30 što predstavlja omjer broja instanci unutar skupa za treniranje i skupa za testiranje. Prvi eksperiment predstavlja eksperiment gdje se model trenira na jednoj životinji, zatim testira na podacima iste životinje koje model ranije nije vidio, a kasnije na još dvije različite životinje. Prvi uzeti skup je skup G1 koji predstavlja podatke jedne koze. On sadrži oko dva milijuna podataka koji su podijeljeni na trening i test skupove. Prvo model trenira na 70 posto podataka tog skupa, a zatim testira na ostatku. Zbog točnosti rezultata i ovisnosti o nasumičnoj podjeli skupa, ovaj bi proces trebalo ponoviti više puta. U ovom slučaju ovaj proces se ponavlja 30 puta i podaci se zapisuju u tablicu. Drugi podatkovni skup je skup G2 koji predstavlja istu vrstu, kozu, ali drugu životinju. Ovaj skup također ima nešto više od dva milijuna podataka. Iz tog skupa, uzeta je jednaka veličina testnog skupa kao što je bio i u prethodnom skupu. Točnije, testni skup skupa G1 iznosi oko 63500 podataka, tako je i iz skupa G2 nasumično izvučeno također oko 63500 podataka i na njima se vršio test. Zadnji, treći slučaj je slučaj skupa S1, ponovljeno je isto što je učinjeno na skupu G2. Proces testiranja i za ova dva skupa se također izvodio 30 puta i rezultati su zapisani u tablicu. Drugi eksperiment donosi malo drukčiju postavu podataka. U drugom eksperimentu su uzeta tri skupa koze, G1, G2, G3 i spojeni u jedan skup. Taj spojeni skup je uzet kao skup za treniranje i testiranje. Jednakom metodom razdvojen je na omjer 70-30 trening i test skupa. Model je istreniran na većem dijelu skupa te je zatim testiran na ostatku. Takav skup je puno veći od skupova prethodnog testiranja, on sadrži oko šest i pol milijuna podataka. Što bi značilo da je skup za testiranje velik približno dva milijuna podataka. Nakon treniranja i testiranja ovog skupa, ponovno dolazi do spajanja, ovaj put spajanja dva skupa ovaca, S1 i S2. Nakon što su skupovi spojeni također je nasumično uzet jednak broj podataka kao u prošlom skupu koji je iznosio oko dva milijuna podataka te je na njemu izvršeno testiranje. Kao i u prvom eksperimentu ovaj proces testiranja je proveden 30 puta i rezultati su zapisani u tablicu.

3.3. Usporedba rezultata testiranja

Nakon izvršene metode zadržavanja rezultati su upisani u tablicu. Pomoću tih podataka dobivene su neke statističke vrijednosti kao što su srednja vrijednost, standardna devijacija, minimum, maksimum i medijan. Pomoću tih podataka uspoređeni su i objašnjeni radovi svih klasifikatora,

njihove prednosti i mane. Također, rezultati su prikazani u grafičkom obliku pomoću okvirne plohe.

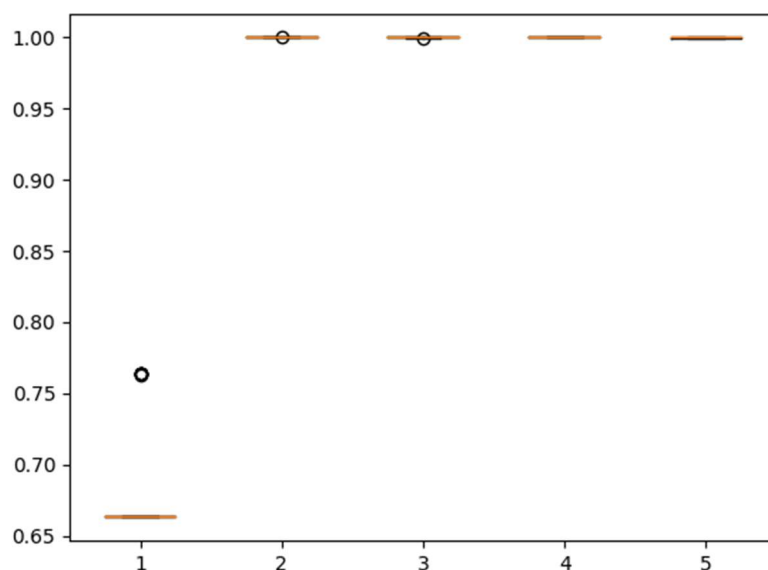
3.3.1. Testiranje na skupu podataka tri različite životinje

U prvom eksperimentu, rezultati prvog testiranja u kojem je došlo do treniranja i testiranja na istoj životinji su bili očekivano dobri, dok su testiranja na ostalim životinjama bili znatno slabiji. Tablica 3.1. prikazuje prosječne vrijednosti rezultata testiranja pri kojem je testirano i trenirano na istoj životinji, odnosno na G1 podatkovnom skupu, kao što je rečeno ranije. Rezultati su poprilično dobri, izuzev algoritma logističke regresije. Logistička regresija je kao što je rečeno, jedan od najjednostavnijih i najbržih algoritama koji dosta loše radi sa velikom količinom podataka, kao što je ovdje slučaj. Ostali algoritmi pokazali su vrlo zavidne rezultate, srednja vrijednost im je oko 1, što bi značilo 100% točnosti testiranja. Mala, ali i zanemariva odstupanja dolaze kod algoritama stablo odluke i 5 najbližih susjeda gdje je odstupanje svega 0.0001. Iz tablice je vidljiva velika točnost algoritma najbližih susjeda. S obzirom da se testira na istoj životinji na kojoj je model trenirao vidljivi su slični podaci što olakšava tom algoritmu da dobije točan rezultat. Ali, već se kod algoritma pet najbližih susjeda vidi mala odstupanja jer se u pet susjeda češće pronade krivi rezultat, za razliku od algoritma jedan najbliži susjed.

Tablica 3.1. Prosječne vrijednosti prvog testiranja

| | LR | 1NN | 5NN | RF | DT |
|------------------------------|-----------|------------|------------|-----------|-----------|
| Srednja vrijednost | 0.6833 | 1.0000 | 1.0000 | 1.0000 | 0.9999 |
| Medijan | 0.6632 | 1.0000 | 1.0000 | 1.0000 | 0.9999 |
| Minimum | 0.6632 | 1.0000 | 0.9996 | 1.0000 | 0.9997 |
| Maksimum | 0.7641 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Standardna devijacija | 0.0408 | 0.0000 | 0.0001 | 0.0346 | 0.0001 |

Slika 3.3. prikazuje prosječne rezultate testiranja, ali u grafičkom obliku pomoću kutijastog dijagrama. Na slici nisu vidljive glavne karakteristike kutijastog dijagrama iz razloga što je oscilacija ponavljanja vrlo mala, pa na slici nisu vidljive razlike prvog i trećeg kvartila, ili minimuma i maksimuma. Vidljiva je samo narančasta crta koja predstavlja medijan i koja vrijednošću odgovora maloprije spomenutim rezultatima.



Slika 3.3. Grafički prikaz prvog testiranja

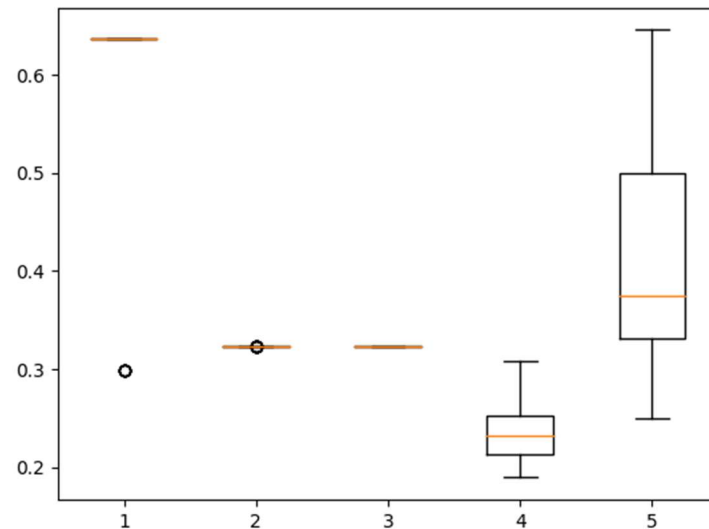
Drugo testiranje u kojem je testirano na različitoj životinji donosi puno lošije rezultate. Algoritam logistička regresija pokazuje se kao najbolji u ovom testiranju. Uz brzinu i jednostavnost, velika prednost algoritma logistička regresija je mala varijanca, odnosno malo odstupanje u rezultatima. Dok su ostali algoritmi napravili ogroman pad u točnosti rezultata, logistička regresija je tek za malo lošija od prvog testiranja. U tablici 3.2. vidljivi su rezultati drugog testiranja. 1 najbliži susjed i 5 najbližih susjeda pokazuju približno jednake rezultate što dovodi do zaključka da je poprilično jednaka raspodjela susjeda kada se gleda 1 ili 5 najbližih. Nasumična šuma u ovom testiranju daje najlošije rezultate, dok jedno stablo odluke daje bolje rezultate, razlika je u tome što stablo odluke ima velike oscilacije u rezultatima.

Tablica 3.2. Prosječne vrijednosti drugog testiranja

| | LR | 1NN | 5NN | RF | DT |
|------------------------------|--------|--------|--------|--------|--------|
| Srednja vrijednost | 0.5692 | 0.3233 | 0.3234 | 0.2357 | 0.3994 |
| Medijan | 0.6368 | 0.3233 | 0.3234 | 0.2280 | 0.3753 |
| Minimum | 0.2986 | 0.3233 | 0.3233 | 0.1896 | 0.2504 |
| Maksimum | 0.6368 | 0.3234 | 0.3234 | 0.3084 | 0.6458 |
| Standardna devijacija | 0.1375 | 0.0000 | 0.0000 | 0.0346 | 0.1056 |

U drugom eksperimentu kao što je rečeno, pojavljuju se veće oscilacije što dovodi do boljeg iscrtavanja okvirne plohe. Takav primjer je vidljiv na slici 3.4. gdje nasumična šuma i stablo

odluke daju velike oscilacije u 30 ponavljanja što dovodi do izduživanja okvirne plohe. Na slici su vidljive sve glavne karakteristike koje su spomenute ranije.



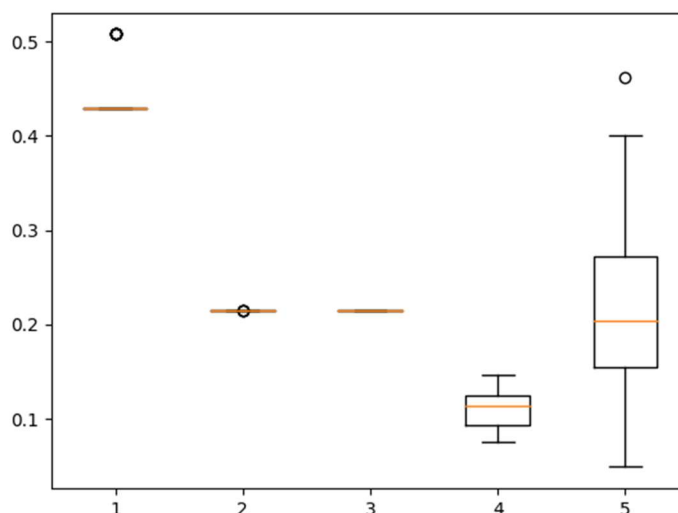
Slika 3.4. Grafički prikaz drugog testiranja

Treće testiranje donosi rezultate testirane na drugačijoj vrsti životinje nego na kojoj je model trenirao (Tablica 3.3.). Rezultati su očekivano loši. Logistička regresija kao i u prošlom testiranju sa vrlo malim padom drži najbolje i najtočnije rezultate. Ponovno je vidljiva razlika između nasumične šume i stabla odluke. Stabla odluke vrlo su jednostavna u usporedbi sa slučajnom šumom. Stablo odluke kombinira neke odluke, dok slučajna šuma kombinira nekoliko stabala odluke. To je dug i spor proces. Dok je stablo odluke brže i radi odlično sa velikom količinom podataka, kao što je tu slučaj.

Tablica 3.3. Prosječne vrijednosti trećeg testiranja

| | LR | 1NN | 5NN | RF | DT |
|------------------------------|-----------|------------|------------|-----------|-----------|
| Srednja vrijednost | 0.4444 | 0.2155 | 0.2155 | 0.1100 | 0.2128 |
| Medijan | 0.4285 | 0.2155 | 0.2155 | 0.1129 | 0.2042 |
| Minimum | 0.4285 | 0.2155 | 0.2155 | 0.0714 | 0.0493 |
| Maksimum | 0.5077 | 0.2155 | 0.2155 | 0.1463 | 0.4624 |
| Standardna devijacija | 0.0322 | 0.0000 | 0.0000 | 0.0194 | 0.0966 |

Slika 3.5. predstavlja grafički prikaz slijedećeg testiranja daje rezultate slične prethodnom. Vrlo malo odstupanja kod algoritama logističke regresije i najbližih susjeda. Dok su razlike u rezultatima vidljive na algoritmima nasumične šume i stabla odluke.



Slika 3.5. Grafički prikaz trećeg testiranja

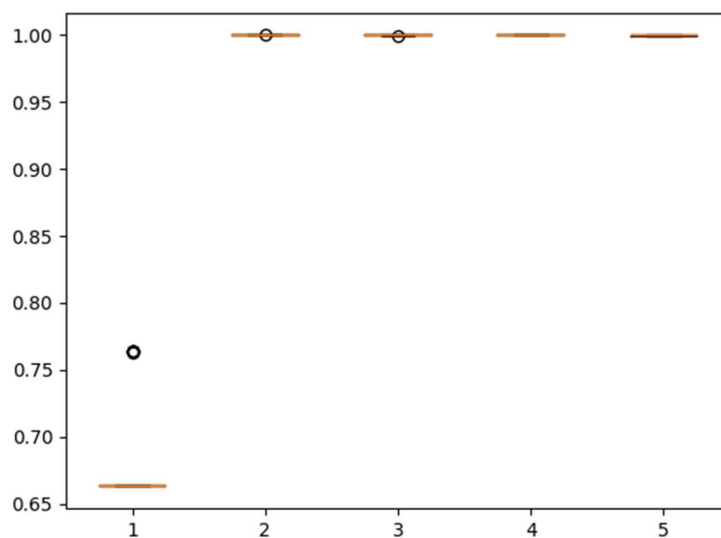
3.3.2. Testiranje na spojenim skupovima podataka različitih životinja

U drugom eksperimentu prvi podatkovni skup za testiranje bio je podatkovni skup načinjen od tri skupa iste životinje, koze. Rezultati su slični kao u prvom eksperimentu gdje je testirano i trenirano na istoj životinji (Tablica 3.4.). U ovom slučaju se samo radi o puno većem skupu podataka nego u prošlom eksperimentu. Logistička regresija ponovno pokazuje najlošije rezultate, što je bilo očekivano.

Tablica 3.4. Prosječne vrijednosti prvog testiranja eksperimenta spojenih podatkovnih skupova

| | LR | 1NN | 5NN | RF | DT |
|------------------------------|--------|--------|--------|--------|--------|
| Srednja vrijednost | 0.6833 | 1.0000 | 1.0000 | 1.0000 | 0.9999 |
| Medijan | 0.6632 | 1.0000 | 1.0000 | 1.0000 | 0.9999 |
| Minimum | 0.6632 | 1.0000 | 0.9996 | 1.0000 | 0.9997 |
| Maksimum | 0.7641 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Standardna devijacija | 0.0408 | 0.0000 | 0.0001 | 0.0346 | 0.0001 |

Slika 3.6. prikazuje grafički prikaz prosječnih rezultata. Na slici su vidljivi vrlo precizni rezultati sa vrlo malo odstupanja u 30 ponavljanja.



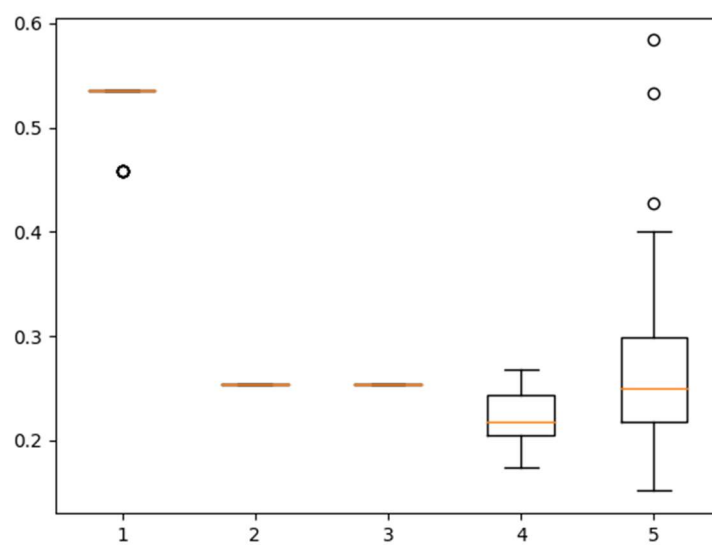
Slika 3.6. Grafički prikaz prvog testiranja eksperimenta spojenih podatkovnih skupova

Tablica 3.5. prikazuje slične omjere rezultata kao što je bio slučaj i u prvom eksperimentu, kada se testiralo na različitoj životinji nego na kojoj je trenirano. Algoritmi najbližih susjeda ponovno pokazuju vrlo slične rezultate. U ovom eksperimentu jedino dolazi do razlike u omjeru kvalitete nasumičnih šuma i stabla odluke. U prethodnom eksperimentu stablo šuma odradilo je puno bolji posao, rezultati su bili dosta točniji od algoritma nasumičnih šuma. U ovom eksperimentu također daje bolje rezultate, ali je razlika puno manja jer se radi o skupu sa puno više podataka.

Tablica 3.5. Prosječne vrijednosti drugog testiranja eksperimenta spojenih podatkovnih skupova

| | LR | 1NN | 5NN | RF | DT |
|------------------------------|-----------|------------|------------|-----------|-----------|
| Srednja vrijednost | 0.5200 | 0.2539 | 0.2539 | 0.2244 | 0.2821 |
| Medijan | 0.5354 | 0.2539 | 0.2539 | 0.2209 | 0.2495 |
| Minimum | 0.4584 | 0.2539 | 0.2539 | 0.1744 | 0.1518 |
| Maksimum | 0.5354 | 0.2539 | 0.2539 | 0.2683 | 0.5838 |
| Standardna devijacija | 0.0313 | 0.0000 | 0.0000 | 0.0244 | 0.1031 |

Slika 3.7. prikazuje grafički prikaz rezultata drugog testiranja. Zaključak je isti kao u prethodnom eksperimentu. Stvara se velika razlika u točnosti algoritama koji trenira i testira na istoj životinji od drugog primjera gdje se testira na drugačijoj životinji. U ovom primjeru je vidljiva razlika samo u algoritmu stablo odluke, koji je pokazao lošije rezultate u usporedbi sa prošlim eksperimentom kada je bilo manje podataka.



Slika 3.7. Grafički prikaz drugog testiranja eksperimenta spojenih podatkovnih skupova

4. Zaključak

U poljoprivredi dolazi sve veća potreba za strojnim učenjem u svrhu olakšavanja nekih poslova, donošenja bitnih odluka, rad s velikom količinom podataka. U budućnosti će se strojno učenje početi koristiti u puno većoj mjeri nego sada i zato je vrlo bitno da su potrošači takvih usluga svjesni s čime se susreću, što im to strojno učenje može donijeti. U svrhu toga ovaj rad pokazuje kvalitetu rada raznih algoritama na problemu raspoznavanja aktivnosti životinje. Korištena je velika količina podataka. Algoritmi su se susretali sa raznim problemima, od same količine podataka, do raznih nepoznatih vrijednosti, drugačijih vrsta životinja, nesrazmjera među brojem instanci koje pripadaju pojedinoj klasi (neuravnoteženost u podacima) i slično. Rezultati su bili razni, zapisani su u tablicu i kasnije detaljno objašnjeni i obrazloženi.

Zaključci su doveli do toga da ako se testira na istoj životinji na kojoj je algoritam treniran da će rezultati biti zavidni i vrlo korisni. Dok do problema dolazi kada se na testiranju nađe skup životinje na kojoj nije testirano. Preciznost algoritama značajno opada i rezultati su lošiji. Ovakva vrsta rada ima mjesta za napretkom. Postoje razni drugi algoritmi, razne druge metode uporabe podataka s kojima bi bilo isto tako korisno provjeriti i testirati kako se ponašaju u raznim okolnostima. Posebice se ovdje, radi količine dostupnih podataka, nameću postupci dubinskog učenja. Moglo bi se testirati na još mnoštvo drugih životinja, kojima bi možda podaci bili slični što bi dovelo do novih problema za algoritme i novih zaključaka. U konačnici, mogli bi se razmotriti i neki postupci za uravnoteživanje podatkovnih skupova koji bi mogli doprinijeti kvaliteti rezultata.

LITERATURA:

- [1] IBM Cloud Education, Machine Learning, 2020., dostupno na <https://www.ibm.com/cloud/learn/machine-learning> [pristupljeno: 31.8.2021.]
- [2] Microsoft, Windows Machine Learning, What is a machine learning model?, 2021., dostupno na <https://docs.microsoft.com/en-us/windows/ai/windows-ml/what-is-a-machine-learning-model> [pristupljeno: 31.8.2021]
- [3] E. G. Learner-Miller, Introduction to Supervised Learning, Amherst, 2014. Detector Performance Analysis Using ROC Curves, 2016., dostupno na <https://www.mathworks.com/help/phased/ug/detector-performance-analysis-using-roc-curves.html> [pristupljeno: 29.8.2021]
- [4] S. Kaushik, An Introduction to Clustering and different methods of clustering, 2016., dostupno na <https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/> [pristupljeno: 25.8.2021]
- [5] M. Waseem, How To Implement Classification In Machine Learning?, 2021., dostupno na <https://www.edureka.co/blog/classification-in-machine-learning/> [pristupljeno: 28.8.2021.]
- [6] Detector Performance Analysis Using ROC Curves, 2016., dostupno na <https://www.mathworks.com/help/phased/ug/detector-performance-analysis-using-roc-curves.html> [pristupljeno: 29.8.2021]
- [7] J. Brownlee, What is a Confusion Matrix in Machine Learning, 2020., dostupno na <https://machinelearningmastery.com/confusion-matrix-machine-learning/> [pristupljeno: 31.8.2021.]
- [8] T. G. Mesevage, Machine Learning Classifiers – The Algorithms & How They Work, 2020., dostupno na <https://monkeylearn.com/blog/what-is-a-classifier/> [pristupljeno: 30.8.2021.]
- [9] K. Taunk, S. De, S. Verma, Machine learning classification with k-nearest neighbours, 2021., dostupno na https://www.researchgate.net/publication/340693569_A_Brief_Review_of_Nearest_Neighbor_Algorithm_for_Learning_and_Classification [pristupljeno: 26.8.2021.]
- [10] D. Jurafsky, J. H. Martin, Speech and Language Processing, Prentice Hall, United States 2009.
- [11] P. Tan, M. Steinbach, A. Karpatne, V. Kumar, Introduction to data mining, Addison – Wesley Longman Publishing Co., Inc., United States, 2005.

- [12] O. Mbaabu, Introduction to Random Forest in Machine Learning, 2020., dostupno na <https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/> [pristupljeno: 30.8.2021.]
- [13] Lakshanagy, 4 Ways to Evaluate your Machine Learning Model: Cross-Validation Techniques (with Python code), 2021., dostupno na <https://www.analyticsvidhya.com/blog/2021/05/4-ways-to-evaluate-your-machine-learning-model-cross-validation-techniques-with-python-code/> [pristupljeno: 31.8.2021.]
- [14] D. Levinger, VP Dev and Cloud Operations, Six Steps to Master Machine Learning with Data Preparation, 2018., dostupno na <https://www.kdnuggets.com/2018/12/six-steps-master-machine-learning-data-preparation.html> [pristupljeno: 30.8.2021.]
- [15] L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, D. Bochtis, Machine Learning in Agriculture: A Comprehensive Updated Review, sv. 21, str. 10-14, 2021.
- [16] David D. Luxton, Artificial Intelligence in Behavioral and Mental Health Care, Elsevier Academic Press, San Diego, Unites States, 2016.
- [17] Konstantinos Koutroumbas, Sergios Theodoridis, Pattern Recognition 4th Edition, Academic Press, Inc., Unites States, 2008.
- [18] Generic online animal activity recognition on collar tags, dostupno na <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:76131/tab/2> [pristupljeno: 10.8.2021.]
- [19] Scikit-learn, Machine Learning in Python, <https://scikit-learn.org/stable/> [pristupljeno: 12.8.2021.]
- [20] S. Asiri, Machine LEarning Classifiers, 2018., dostupno na <https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623> [pristupljeno: 25.8.2021.]
- [21] K. Dokic, From machine learning to deep learning in agriculture – the quantitative review of trends, IOP Conf. Series: Earth and Environmental Science, sv. 614, str. 1-5, United Kingdom, 2020.

SAŽETAK

U teorijskom dijelu rada opisan je problem raspoznavanja aktivnosti životinja uzimajući velike i različite podatkovne skupove. Kroz razne algoritme i klasifikatore cilj je bio testirati rad klasifikatora pod raznim problemima. U eksperimentalnom je dijelu bilo potrebno obraditi podatke kako bi bili spremni za daljnji rad. Zatim je naslijedilo treniranje modela na određenoj količini podataka koja je podijeljena algoritmom na omjer 70-30, odnosno omjer trening i test skupa. Nakon treniranja uslijedila su tri testiranja, testiranje na istoj i na dvije različite životinje. Rezultati su bili očekivani te detaljno obrazloženi. Drugi eksperiment se razlikovao od prvog po tome što je korišteno puno više podataka. Podatkovni skup za rad je bio spoj svih skupova iste životinje. Dobivena su dva takva velika skupa. Na prvom se vršilo treniranje na isti način kao u prvom eksperimentu. Te su na kraju dobiveni rezultati testiranja oba skupa.

Ključne riječi: klasifikacija, nosivi senzori, podatkovni skup, raspoznavanje aktivnosti životinja, strojno učenje

TITLE

Wearable sensor data based animal activity recognition

ABSTRACT

The theoretical part describes the problem of recognizing the activities of animals, by using big amount of data sets. Through different algorithms and classifiers, the main goal was to test operation of classifiers under different problems. The first problem was working with huge and unknown data in a data set. In practical part, it was necessary to process the data in order to be ready for further work. This way was followed by the training of the model on a certain amount of data which was divided by the algorithm into a ratio of 70-30, ie. the ratio of training and test set. Training was followed by three tests, testing on the same animal, and then on two different ones. The results were expected and explained in detail. The second experiment differed from the first one, because much more data was used in second one. The work data set was a combination of all sets from the same animal. It gives two big sets of data. The first one was trained in the same way as in the first experiment. In the end, the test results of both sets were obtained.

Keywords: classification, wearable sensors, data set, recognition of animal activities, machine learning

ŽIVOTOPIS

Ivan Kikić rođen u Vinkovcima 4.4.1999. s prebivalištem u Vinkovcima i boravištem u Osijeku. Nakon završene osnovne škole Vladimira Nazora Vinkovci, upisuje Tehničku školu Ruđera Boškovića u Vinkovcima. Po završetku srednje škole upisuje na Fakultet elektrotehnike, računarstva i informacijskih tehnologija u Osijeku na Preddiplomski sveučilišni studij Računarstvo 2018. godine na kojem i trenutno studira završavajući treću godinu preddiplomskog studija.

Ivan Kikić

PRILOZI

Na CD-u:

1. Završni rad “Raspoznavanje aktivnosti životinje na temelju podataka s nosivih senzora” u *.docx* formatu
2. Završni rad “ Raspoznavanje aktivnosti životinje na temelju podataka s nosivih senzora” u *.pdf* formatu
3. Izvorni kod eksperimentalnog dijela rada