

Analiza Twitter socijalne mreže

Majdandžić, Elenora

Undergraduate thesis / Završni rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Polytechnic of Međimurje in Čakovec / Međimursko veleučilište u Čakovcu**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:110:299595>

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

Download date / Datum preuzimanja: **2024-12-21**



Repository / Repozitorij:

[Polytechnic of Međimurje in Čakovec Repository -
Polytechnic of Međimurje Undergraduate and
Graduate Theses Repository](#)





MEĐIMURSKO VELEUČILIŠTE U ČAKOVCU
STRUČNI PRIJEDIPLOMSKI STUDIJ RAČUNARSTVO

Elenora Majdandžić, 0313022646

Analiza Twitter socijalne mreže

Završni rad

Čakovec, rujan 2024



MEĐIMURSKO VELEUČILIŠTE U ČAKOVCU
STRUČNI PRIJEDIPLOMSKI STUDIJ RAČUNARSTVO

Elenora Majdandžić, 0313022646

Analiza Twitter društvene mreže

Analysis of the Twitter Social Network

Završni rad

Mentor:

Željko Knok, prof.

Čakovec, rujan 2024.



MEĐIMURSKO VELEUČILIŠTE U ČAKOVCU

PRIJAVA TEME I OBRANE ZAVRŠNOG/DIPLOMSKOG RADA

Stručni prijediplomski studij:

Računarstvo Održivi razvoj Menadžment turizma i sporta

Stručni diplomski studij Menadžment turizma i sporta:

Pristupnik: Elenora Majdandžić, JMBAG: 0313022646
(ime i prezime)

Kolegij: Baze podataka I
(na kojem se piše rad)

Mentor: mr.sc. Željko Knok
(ime i prezime, zvanje)

Naslov rada: Analiza Twitter socijalne mreže

Naslov rada na engleskom jeziku: Twitter social network analysis

Članovi povjerenstva: 1. mr.sc. Ivan Hegeduš, predsjednik
(ime i prezime, zvanje)
2. Nenad Breslauer, v.pred., član
(ime i prezime, zvanje)
3. mr.sc. Željko Knok, mentor
(ime i prezime, zvanje)
4. Marija Miščančuk, v.pred., zamjenski član
(ime i prezime, zvanje)

Broj zadatka: 2023-RAČ-7

Kratki opis zadatka: _____

U današnje vrijeme društvene mreže imaju izuzetno važnu ulogu, jer se gotovo sve oglašava putem socijalnih mreža. Potrebno je u ovom radu analizirati Twitter socijalnu mrežu, kroz

utjecaj korisnika, njihove odnos prema određenoj temi, a u cilju pravilne percepcije događaja i trendova. Pored ostalog važno je blagovremeno donošenja odluka. U pojedinim

situacijama izuzetno je važna procjena razvoja događaja i trendova. Analizu provesti kroz alate otvorenog koda. Alati: Okruženje za rad Google Colab, Python, Teltar API, Grafčki

prikaz društvenih mreža, Analiza jedne osobine kroz Python biblioteku matplotlib

Datum: 5.9.2024.

Potpis mentora: Željko Knok

Predgovor

Upućujem zahvalu svim suradnicima koji su svojim savjetima i podrškom doprinijeli realizaciji ovog rada. Također zahvaljujem svom mentoru na stručnom vođenju i usmjeravanju prilikom nejasnoća prilikom istraživanja i pisanja.

Sažetak

U ovom su radu spojene tehnologije i grane računarskih znanosti koje tvore jedan sustav obrade podataka i iskorištavanja istih, sa svrhom predviđanja ili prepoznavanja uzoraka.

Za realizaciju ovog rada upotrijebljen je programski jezik Python i popratne biblioteke koje on sadrži, a matplotlib i pandas pripadaju nekim od značajnijih. Iako je ovaj rad prvotno zamišljen kao analiza samo društvene mreže Twitter, uključene su i druge društvene mreže koje sadrže identične parametre kao i Twitter, odnosno uključena je i platforma X. Razlog leži u ograničenom pristupu podacima. Službeni pristup skupovima podataka (eng. *dataset*) naziva se Twitter API. Međutim, besplatni pristup podacima vrlo je oskudan i nije bio pogodan za izradu rada. Također postoji i plaćena razina pristupa, a cijena ovisi o količini podataka i o drugim specifičnostima.

Kako je već spomenuto, u ovom radu iskorišten je skup podataka koji je treniran s pomoću Python koda. Korisnik s pomoću korisničkog sučelja bira količinu i vrstu skupa podataka te bira i željeni algoritam. Moguće je birati među pet algoritama, a rezultat se prikazuje s pomoću histograma.

Ključne riječi: *prepoznavanje uzoraka, društvene mreže, python, skup podataka*

Abstract

In this paper, technologies and branches of computer science are combined to form a single data processing system, utilizing the data for prediction or pattern recognition purposes. For the implementation of this work, the Python programming language and its libraries were used. Some of the most notable libraries include matplotlib and pandas. Although this paper was initially conceived as an analysis exclusively of the Twitter social network, other social networks with identical parameters to Twitter, or rather, the X platform, were also included. The reason for this was limited access to data. The official access to datasets is through the Twitter API. However, the free access to data is very limited and was not suitable for the development of this paper. There is also a paid level of access, with the price depending on the amount of data and other specific factors.

As already mentioned, in this paper, a dataset trained using Python code was utilized. The user, through the user interface, selects the amount and type of dataset, as well as which algorithm they wish to use. It is possible to choose between five algorithms, and the final result is displayed using a histogram.

Keywords: *pattern recognition, social networks, Python, dataset.*

Popis korištenih kratica

Twitter API – Twitter aplikacijsko programsko sučelje (eng. *Application Programming Interface*)

DBSCAN – eng. *Density-Based Spatial Clustering of Applications with Noise*

CSV – vrijednosti odvojene zarezima (eng. *Comma-Separated Values*)

Sadržaj

1.	Uvod	1
2.	Alati	2
2.1.	Python	2
2.2.	<i>Matplotlib</i>	2
2.3.	<i>Pandas</i>	3
2.4.	Skup podataka.....	3
3.	Raspoznavanje uzoraka.....	4
3.1.	Vrste metoda	5
3.2.	Značajke.....	6
4.	Algoritmi.....	7
4.1.	Šume nasumičnog odlučivanja	7
4.2.	Linearna regresija	8
4.3.	<i>K-Nearest neighbors</i>	9
4.4.	<i>K-Means clustering</i>	10
4.5.	Stablo odluke	11
5.	Python kodovi	12
5.1.	Učitavanje skupa podataka	12
5.2.	Predproces skupa podataka.....	12
5.3.	Izračunavanje suma	13
5.4.	Vizualizacija suma	14
7.1.	Pokretanje algoritma.....	15
6.	Princip rada.....	16
7.	Rezultat	18
8.	Zaključak.....	20
9.	Izjava o autorstvu.....	21
10.	Literatura.....	22
11.	Popis ilustracija	23

1. Uvod

Društvene mreže čine svakodnevicu modernog čovjeka te pripadaju internetskim uslugama koje se najčešće može vidjeti kao platforme ili web-stranice. Postoji stotine servisa, a kao najpopularnije moguće je izdvojiti one koje većina svakodnevno koristi, a to su Facebook, Twitter (X), YouTube, Instagram i Tik Tok.

Twitter ili X socijalna je mreža namijenjena čitanju i slanju kraćih poruka (do 280 znakova). Takve poruke nazivaju se cvrkuti (eng. *tweet*). Također su društvene mreže izvor enormnog broja podataka koji služe za prikupljanje i obradu, stoga se ovaj rad, na temelju takvih podataka, dotiče jedne kompleksne znanstvene discipline u svijetu računarskih znanosti koja se naziva Raspoznavanje uzoraka (eng. *Pattern recognition*). Nadalje, u ovom radu moguće je vidjeti uporabu nekih algoritama i metoda strojnog učenja te je moguće blago se dotaknuti pojma umjetne inteligencije.

2. Alati

2.1. Python

Programski jezik Python vrlo se često opisuje kao programski jezik opće namjene. Jedan je od najpopularnijih programskih jezika današnjice zbog svoje jednostavnosti sintaksâ. Korištenje Pythona u ovom radu prikladno je zbog modula analitike nad podacima i manipulacije istim. Osim raznolikosti primjene, neke su prednosti Pythona – besplatna upotreba, kompatibilnost i brzo učenje istog. Platforme koje koriste programski jezik Python jesu Instagram, YouTube, Google i Spotyfi.

2.2. Matplotlib

Ova biblioteka služi vizualiziranju podataka u Pythonu. Matplotlib je *open-source* i besplatna je zamjena MATLAB-a. Sadrži API, tako da se histograme koji su kreirani unutar ove biblioteke može priključiti u aplikaciju s korisničkim sučeljem.

```
import matplotlib.pyplot as plt
```

Slika 1. Priključenje matplotlib biblioteke u Python kôd (izvor: autor)

2.3. Pandas

Pandas je softverska biblioteka napisana za programski jezik Python te služi za analizu i manipulaciju podacima. Indeks, Series i DataFrame neki su od glavnih objekata Pandas biblioteke nad kojima se obavljaju gotovo sve operacije. Objekt Index služi za pohranjivanje indeksa, Series služi za obradu jednodimenzionalnih podataka i pohranjivanje istih, dok objekt DataFrame služi za pohranu i obradu dvodimenzionalnih podataka.

```
import pandas as pd
```

Slika 2. Priključenje Pandas biblioteke u Python kôd (izvor: autor)

2.4. Skup podataka

Kako bi realizacija ovog rada uopće bila moguća, bilo je bitno izraditi odgovarajući skup podataka (eng. *dataset*) koji će se upotrebljavati. U ovom slučaju preuzet je skup podataka koji je bio javno dostupan te su takvi podatci korišteni za testiranje algoritama. Podatci koji se nalaze u spomenutom skupu podataka sadrže tisuću najpopularnijih utjecajnih osoba (eng. *influencers*) u 2022. godini na društvenoj mreži Instagram, YouTube i Tik Tok. Utjecajnici su kategorizirani po broju pratitelja (eng. *followers*) na tim trima društvenim mrežama, uključujući i poznate osobe koje su na njima također popularne.

3. Raspoznavanje uzoraka

Pojam raspoznavanja uzoraka jedna je vrlo kompleksna cjelina koja obuhvaća puno više nego što je spomenuto u ovom radu. Raspoznavanje uzoraka mali je dio koji doprinosi razvoju umjetne inteligencije i strojnog učenja koji još nisu postigli svoj vrhunac u svijetu moderne tehnologije. Raspoznavanje uzoraka omogućuje računalima visoku preciznost prilikom prepoznavanja kompleksnih podataka.

Svrha je raspoznavanja uzoraka u prepoznavanju strukture ili uzorka u podacima i njihovo svrstavanje u određene kategorije ili modele na temelju tih struktura. Prvi je korak prikupljanje podataka. Podatci koji se prikupljaju mogu biti različiti, bilo da se radi o prikupljenim numeričkim podacima, kao što je prikazano u ovom radu, ili su ipak podatci drugačijeg oblika – fotografije, audiozapisi, razni tekstovi ili složeniji signali.

Nakon procesa prikupljanja podataka, drugi korak je pretproces (eng. *preprocessing*), odnosno priprema podataka koji bi bili pogodni za obradu u algoritmima. To mogu biti različite tehnike čišćenja, proces normalizacije ili transformiranje podataka.

Poslije završetka prve dvije faze potrebno je izdvajanje značajki (eng. *Feature Extraction*), odnosno pretvorba sirovih podataka u skupine koje su pogodne za analizu. Ovaj je korak bitan zato što značajno povećava učinkovitost rada algoritama strojnog učenja.

Idući je korak grupiranje koje se izvršava kada su već poznate sve klase i kada se novi uzorci lako mogu svrstati u klasu kojoj pripadaju. Grupiranje (eng. *clustering*) koristi se kada se radi o klasama koje nisu poznate, a algoritam svrstava podatke na princip najveće podudarnosti.

Na samom kraju odvija se evaluacija ili procjena kako bi se utvrdilo koliko dobro odabrani algoritam obavlja raspoznavanje uzoraka.

3.1. Vrste metoda

Kako bi se točno odredilo koji će se algoritam koristiti prilikom treniranja podataka, važno je znati koje su to metode u raspoznavanju uzoraka. Moguće je koristiti sljedeće metode:

1. **Učenje s pojačanjem** – uz interakciju s okolinom, algoritmi su „nagrađeni“ ili „kažnjeni“ za donošenje odluka.
2. **Nenadzirano učenje** – podatci nisu unaprijed označeni, a algoritam u njima traži strukture. (*K-Means Clustering*, *DBSCAN*).
3. **Nadzirano učenje** – podatci su unaprijed označeni (*K-Nearest Neighbors*, *Naive Bayes*, *Support Vector Machines*).
4. **Polunadzirano učenje** – učenje radi poboljšanja točnosti, koristi unaprijed označene i neoznačene podatke.

3.2. Značajke

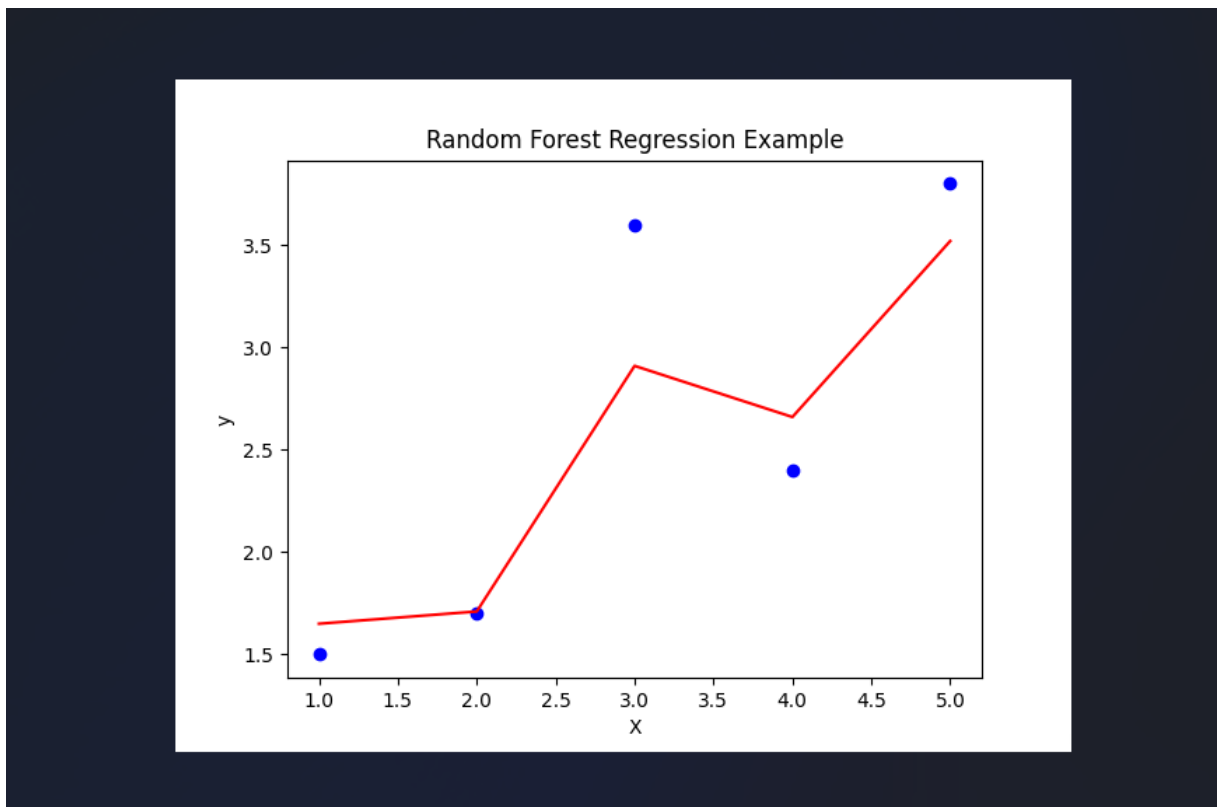
Izdvajanje značajki jedan je od procesa prilikom raspoznavanja uzoraka, kao što je već ranije spomenuto. Izdvajanje značajki omogućuje povećanje točnosti treniranog modela, izbacivanje nepotrebnih podataka te zadržavanje bitnih informacija i općenito povećavanje učinkovitost. To rezultira smanjenjem prekomjernog učenja (eng. *overfitting*) i efikasnijim i preciznijim radom modela.

1. **Numeričke značajke** – brojevi, mogu biti razni statistički podatci (standardna devijacija, prosjek, medijan).
2. **Tekstualna značajka** – riječi, fraze ili druge strukture u tekstu
3. **Slikovne značajke** – oblici, boje ili teksture unutar fotografije
4. **Kategorijske značajke** – podatci koji pripadaju određenim kategorijama. Kategorija se može nazivati „bojom“ i imati vrijednosti „zelena“, „crna“, „crvena“, a one se zatim pretvaraju u numerički oblik.

4. Algoritmi

4.1. Šume nasumičnog odlučivanja

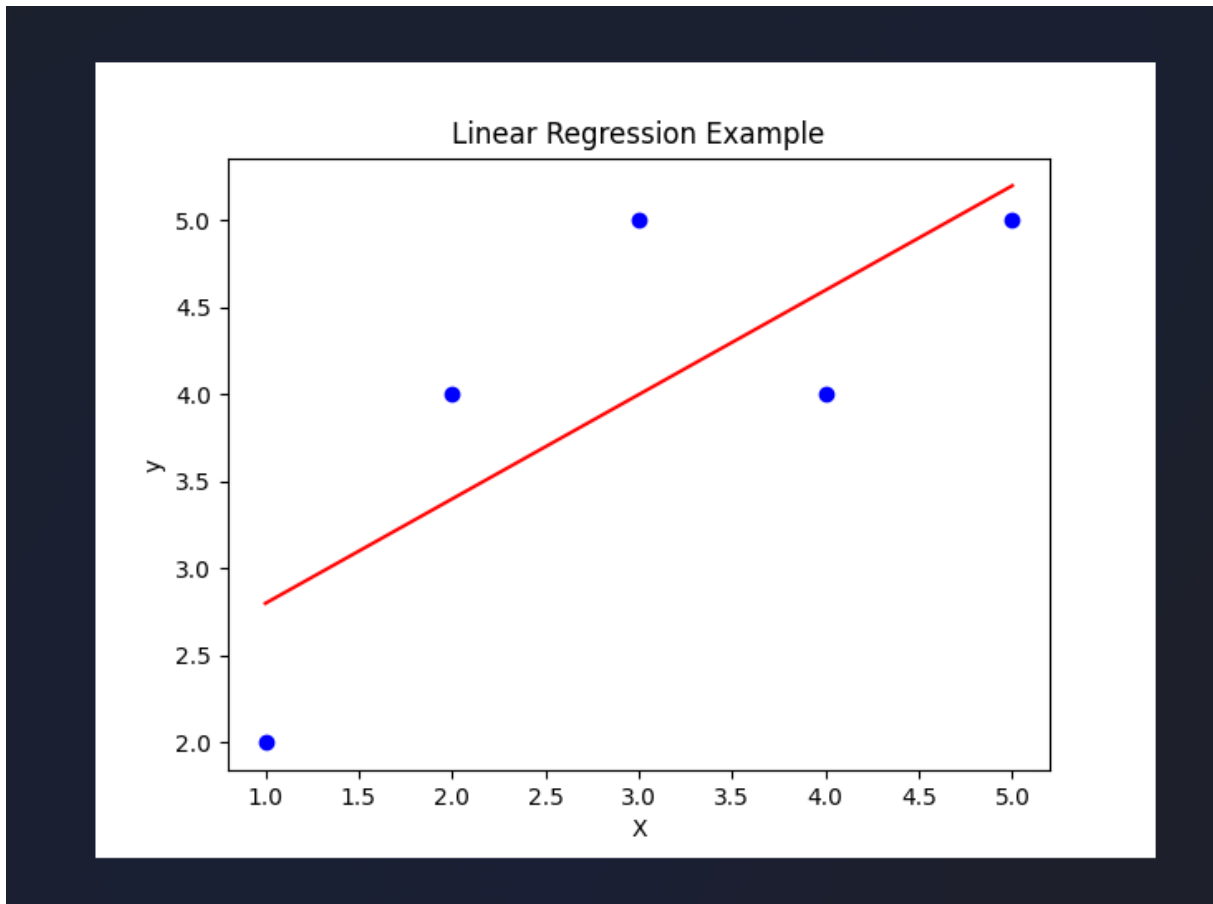
Šume nasumičnog odlučivanja (eng. *random forrest regression*) metoda je strojnog učenja kojoj je osnova stablo odluke (eng. *decision tree*). Metode koje se zasnivaju na izgrađivanju stabla zapravo su samo algoritmi s nadzorom. Najčešće se koriste za rješavanje kada je u pitanju problematika regresijske i klasifikacijske prirode.



Slika 3. Primjer random forests regression grafikona (izvor: autor)

4.2. Linearna regresija

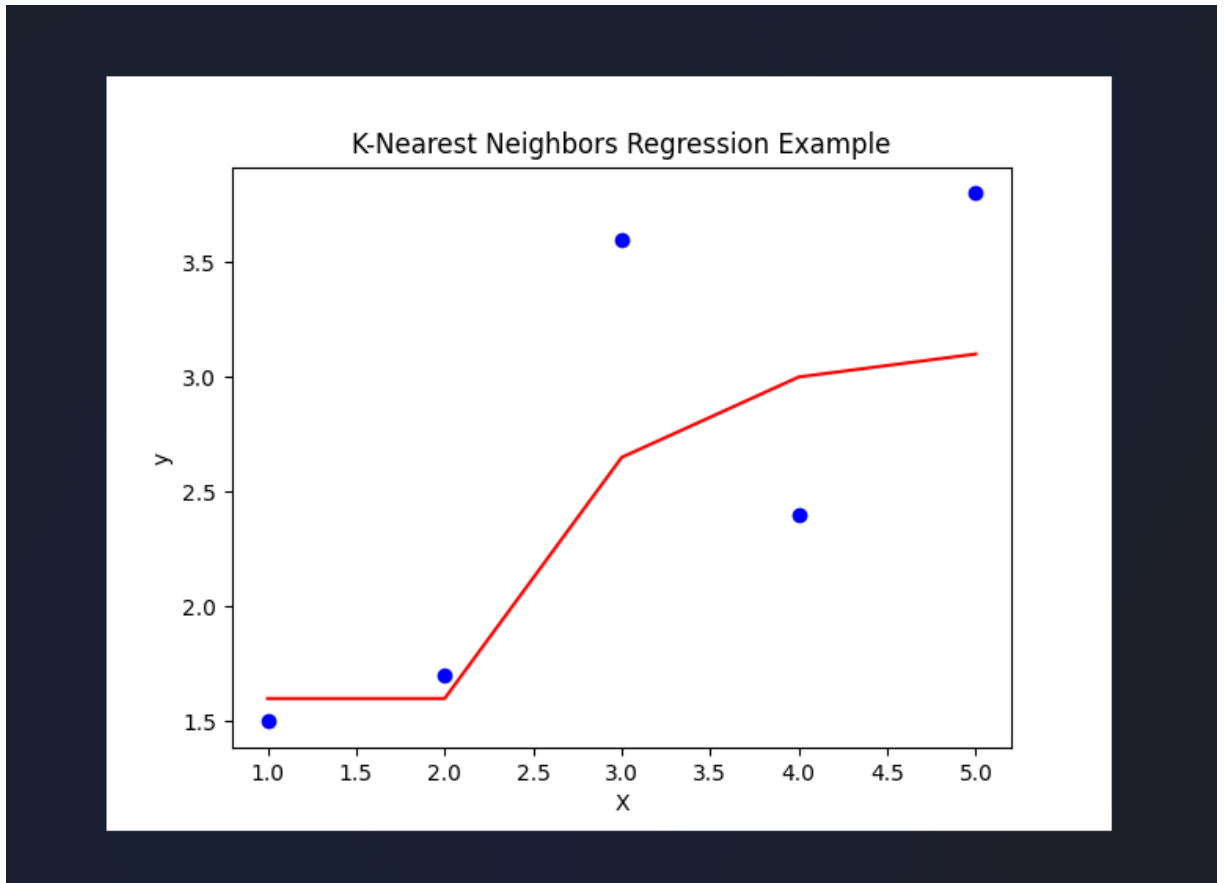
Linearna regresija statistička je metoda opisivanja linearnom vezom jedne ovisne varijable s pomoću jedne ili više varijabli koje će biti nazvane prediktorima. Naziv regresija dao je sir Francis Galton (<https://hrcak.srce.hr/file/450496>)



Slika 4. Primjer grafikona linearne regresije (izvor: autor)

4.3. K-Nearest neighbors

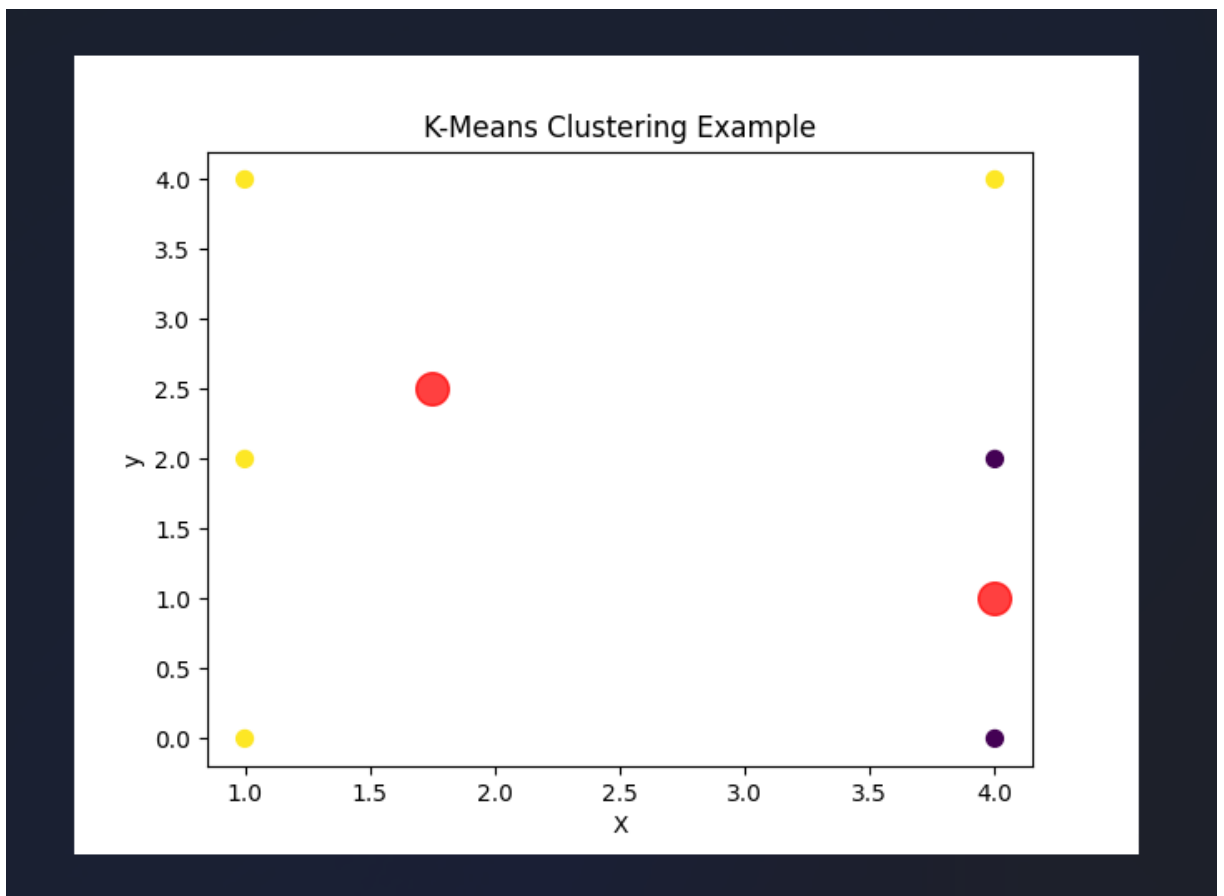
Algoritam k- najbližih susjeda pripada kategoriji najosnovnijih, ali ključnih algoritama za klasifikaciju u području strojnog učenja. Pripada domeni nadziranog učenja i nalazi široku primjenu u prepoznavanju uzoraka, prikupljanju podataka i otkrivanju neželjenih upada.



Slika 5. Primjer grafikona algoritma k-najbližih susjeda (izvor: autor)

4.4. K-Means clustering

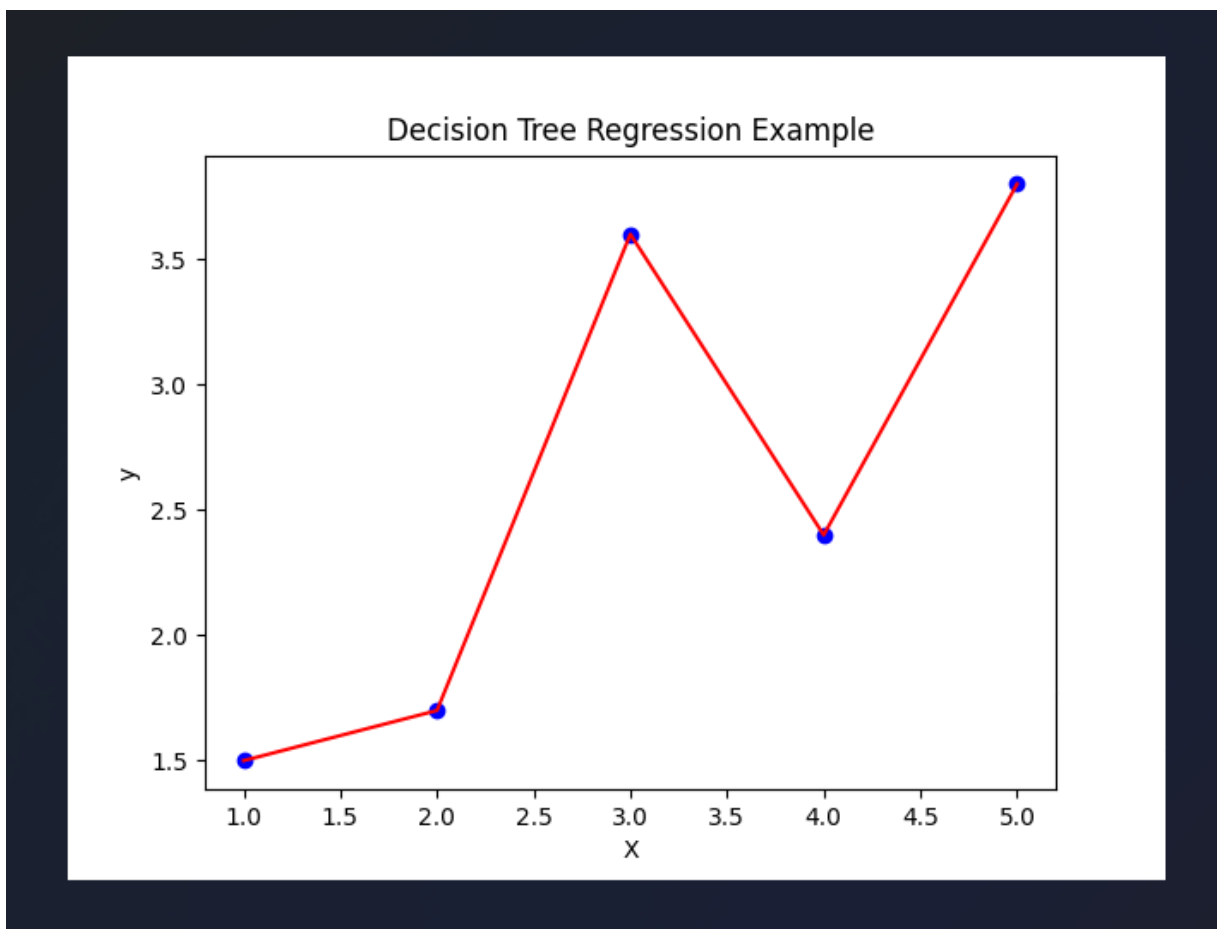
Grupiranje k-means metoda je koja se koristi za vektorsku kvantizaciju i potječe iz obrade signala. Cilj je ove metode podijeliti n- opažanja u k- klastera tako da svako opažanje pripada klasteru čija je sredina najbliža tom opažanju, pri čemu ta sredina služi kao prototip klastera. Kao rezultat, podatkovni prostor dijeli se na Voronoijeve ćelije.



Slika 6. Primjer K-means Clustering grafikona (izvor: autor)

4.5. Stablo odluke

Stablo odluke model je za regresiju ili klasifikaciju, organiziran u obliku stabla. Razdvaja skup podataka na sve manje podskupove, pri čemu se istovremeno i postupno gradi pripadajuće stablo odluke. Na kraju se dobiva stablo koje se sastoji od čvorova odluka i završnih čvorova, odnosno listova.



Slika 7. Primjer grafa Decision tree regression algoritma (izvor: autor)

5. Python kodovi

5.1. Učitavanje skupa podataka

Kôd 1 *Funkcija load_dataset učitava skup podataka iz CSV datoteke. Uklanja praznine oko naziva kolona i vraća DataFrame objekt.*

```
def load_dataset(file_path):
    df = pd.read_csv(file_path)
    df.columns = df.columns.str.strip() # Strip any
leading/trailing whitespace from column names
    df.name = file_path.split('/')[-1]
    return df
```

Izvor: autor

5.2. Pretproces skupa podataka

Kôd 2 *Funkcija preprocess_dataset pretvara vrijednosti u kolonama koje sadrže oznake 'M' (milijuni) i 'K' (tisuće) u odgovarajuće numeričke vrijednosti. Također, uklanja redove koji sadrže NaN vrijednosti.*

```
def preprocess_dataset(df):
    for col in df.columns:
        if df[col].dtype == 'object':
            df[col] =
pd.to_numeric(df[col].str.replace('M', 'e6').str.replace('K',
'e3').str.replace('N/A', '0').str.replace(',', ''),
errors='coerce')
    df = df.select_dtypes(include=['float64', 'int64'])
    df = df.dropna()
    return df
```

Izvor: autor

5.3. Izračunavanje suma

Kôd 3 Funkcija `sum_columns` prolazi kroz odabrane datoteke i računa sume za svaku numeričku kolonu koja ne sadrži ključne riječi kao što su 'name', 'category', itd.

```
def sum_columns():
    dataset_sums = {}
    for file_path, var in selected_files.items():
        if var.get():
            df = load_dataset(file_path)
            file_sums = {}
            for col in df.columns:
                if should_skip_column(col):
                    continue
                if df[col].dtype == 'object':
                    df[col] =
pd.to_numeric(df[col].str.replace('M', 'e6').str.replace('K',
'e3').str.replace('N/A', '0').str.replace(',', ''), errors='coerce')
                if df[col].dtype in ['float64', 'int64']:
                    file_sums[col] = df[col].sum()
            dataset_sums[os.path.basename(file_path)] = file_sums
    for widget in sum_selection_frame.wininfo_children():
        widget.destroy()
    global selected_sum_vars
    selected_sum_vars = {}
    for dataset, sums in dataset_sums.items():
        dataset_label = ttk.Label(sum_selection_frame,
text=f"Dataset: {dataset}", font=('Arial', 10, 'bold'))
        dataset_label.pack(anchor='w')
        sum_vars = {}
        for col, total in sums.items():
            var = tk.BooleanVar()
            sum_vars[col] = (total, var)
            cb = ttk.Checkbutton(sum_selection_frame,
text=f"{col}: {total}", variable=var)
            cb.pack(anchor='w')
        selected_sum_vars[dataset] = sum_vars
```

Izvor: autor

5.4. Vizualizacija suma

Kôd 4 Funkcija `visualize_sums` prikazuje histogram za odabrane sume kako bi korisnik mogao vizualno vidjeti vrijednosti suma za svaku kolonu.

```
def visualize_sums(sums):  
    plt.figure(figsize=(10, 6))  
    plt.bar(sums.keys(), sums.values())  
    plt.xlabel('Columns')  
    plt.ylabel('Sum')  
    plt.title('Sums of Selected Columns')  
    plt.xticks(rotation=45, ha='right')  
    plt.tight_layout()  
    plt.show()
```

Izvor: autor

3.3. Pokretanje algoritma

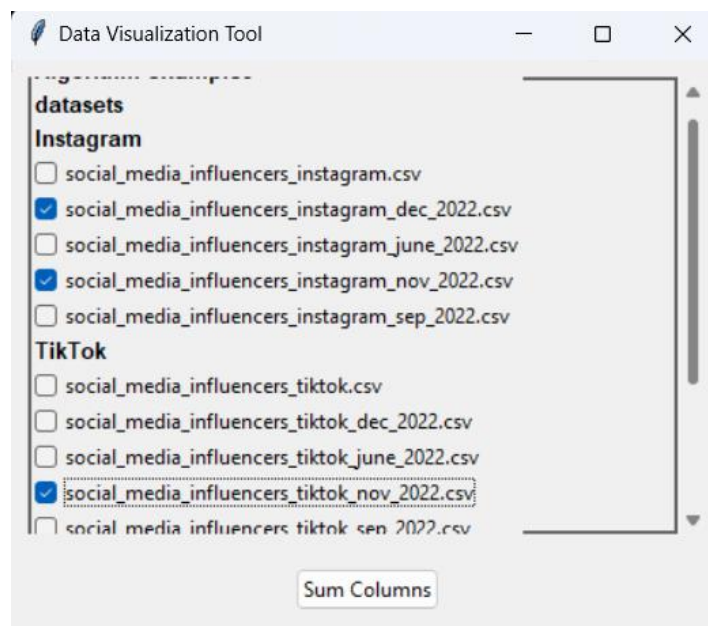
Kôd 5 Funkcija `run_algorithm` omogućava korisniku odabir algoritma za vizualizaciju suma. Podržani algoritmi uključuju "Linear Regression", "K-Means Clustering", "Decision Tree Regression", "Random Forest Regression" i "KNN Regression".

```
def run_algorithm():
    selected_sums = {col: sum_value for dataset, sum_list in
selected_sum_vars.items() for col, (sum_value, var) in
sum_list.items() if var.get()}
    if not selected_sums:
        messagebox.showerror("Error", "Please select at
least one sum to use")
        return
    visualize_sums(selected_sums)
    algo = algo_combobox.get()
    if algo == 'Linear Regression':
        visualize_linear_regression(selected_sums)
    elif algo == 'K-Means Clustering':
        visualize_kmeans(selected_sums)
    elif algo == 'Decision Tree Regression':
        visualize_decision_tree_regression(selected_sums)
    elif algo == 'Random Forest Regression':
        visualize_random_forest_regression(selected_sums)
    elif algo == 'KNN Regression':
        visualize_knn_regression(selected_sums)
    else:
        messagebox.showerror("Error", f"Algorithm '{algo}'
not implemented yet")
```

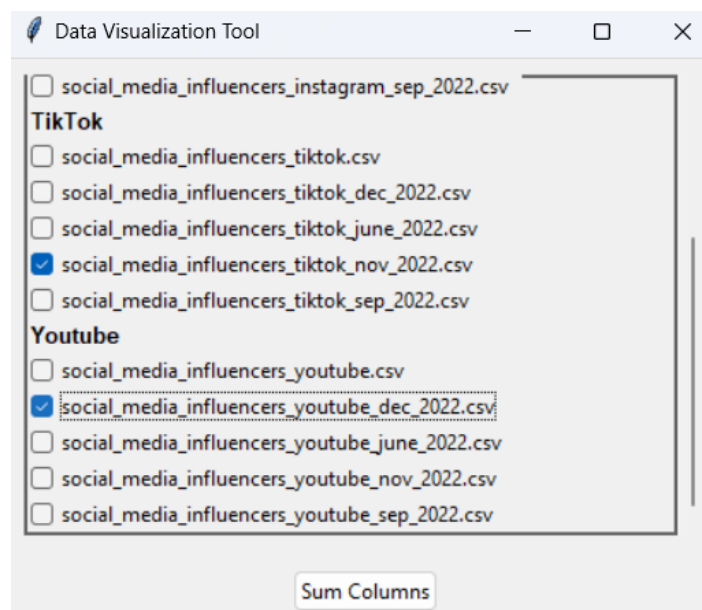
Izvor: autor

6. Princip rada

Praktični dio ovog rada načinjen je tako da se prilikom pokretanja Python programa korisniku prikaže interaktivno korisničko sučelje te korisnik odabire koje skupove podataka i koliko njih želi odabrati. Korisniku su na raspolaganju tri društvene mreže koje sadrže podatke sakupljene tijekom četiriju mjeseci. Kada korisnik odabere željene skupove, klikom na gumb „Sum Columns“ pokreće se numeričko zbrajanje kolona u skupu.



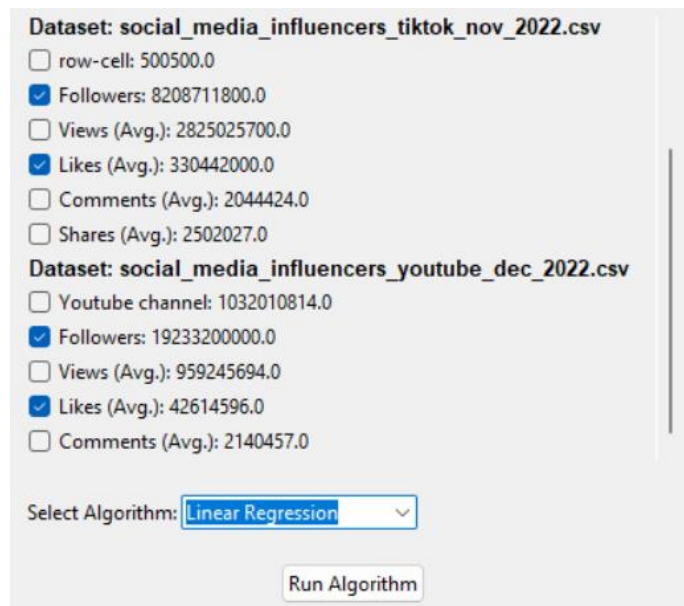
Slika 8. Odabir skupa podataka (izvor: autor)



Slika 9. Odabir skupa podataka (izvor: autor)

Kao primjer, u ovom su testu odabrana četiri podskupa podataka iz svih triju društvenih mreža. Odabrani su podatci prikupljeni u studenom i prosincu 2022. godine. Prvo, kako bi se odabrani podatci mogli zbrajati, bilo je neophodno obaviti pretproces podataka. Neke su od numeričkih kolona u skupu sadržavale oznaku M (milion), K (tisuću), oznaku „na entu“ ili samo praznu kolonu „null“ te kao takve nisu bile pogodne za zbrajanje i bilo je potrebno pretvoriti ih u odgovarajuće numeričke vrijednosti.

Nadalje, kolonama s oznakom M program je dodavao šest nula, dok je kolonama s oznakom K dodavao tri nule. Kolone sa oznakom „na entu“ program je zamijenio nulom. Korištena je i funkcija `coerce()` koja služi za automatsku pretvorbu dvaju brojeva različitog tipa u jedan tip prije nego što se izvede matematička operacija. Na primjer, ako postoji tip podatka „*int*“ i tip podatka „*float*“, „*int*“, broj će se pretvoriti u *float* broj jer je on u ovom slučaju šireg raspona. Nakon odabira gumba „*Sum Columns*“ i obavljenog pretprocesa, na korisničkom sučelju prikazuje se opcija odabira kolona koje želimo trenirati za korisnikove potrebe. Kolone su također numeričkog tipa. Pri odabiru kolona odabiru se srodne kolone kako bi testiranje bilo smisljeno. Na primjer, odabrane su kolone naziva pratitelji (eng. *followers*) i oznaka „sviđa mi se“ (eng. *likes*) kao zajedničku dodirnu točku tih triju skupova. Zatim, kada korisnik odabere kolone, slijedi odabir željenog algoritma.

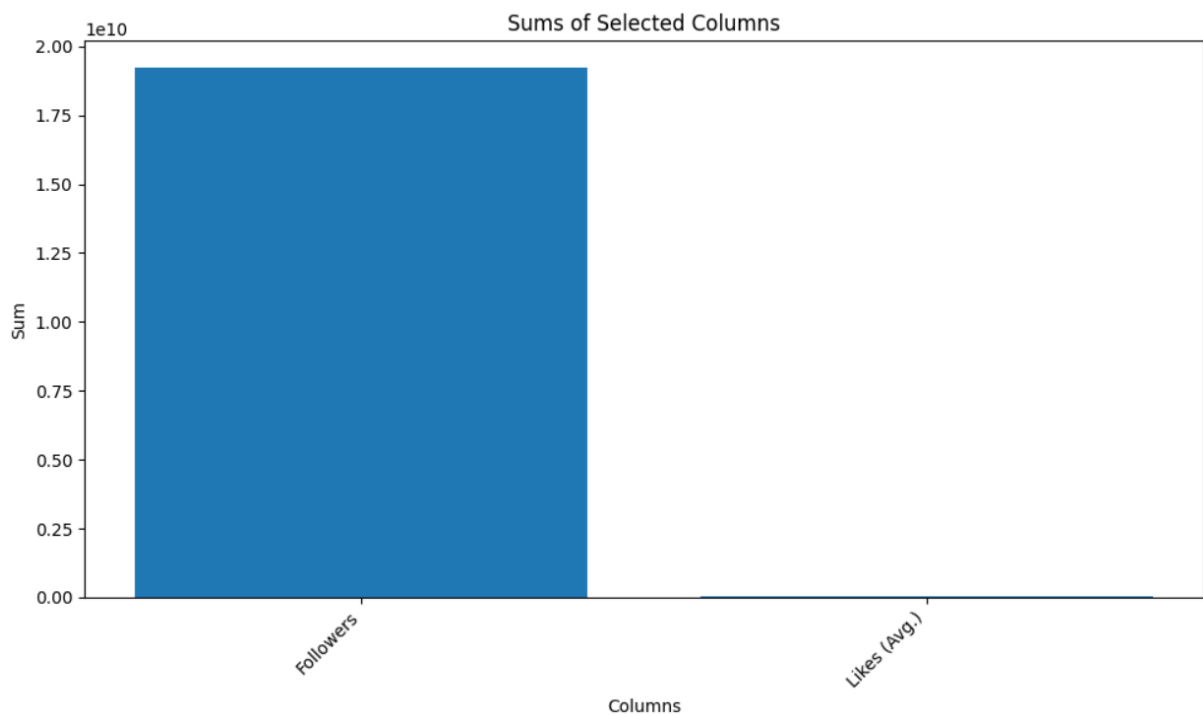


Slika 10. Odabir algoritma (izvor: autor)

7. Rezultat

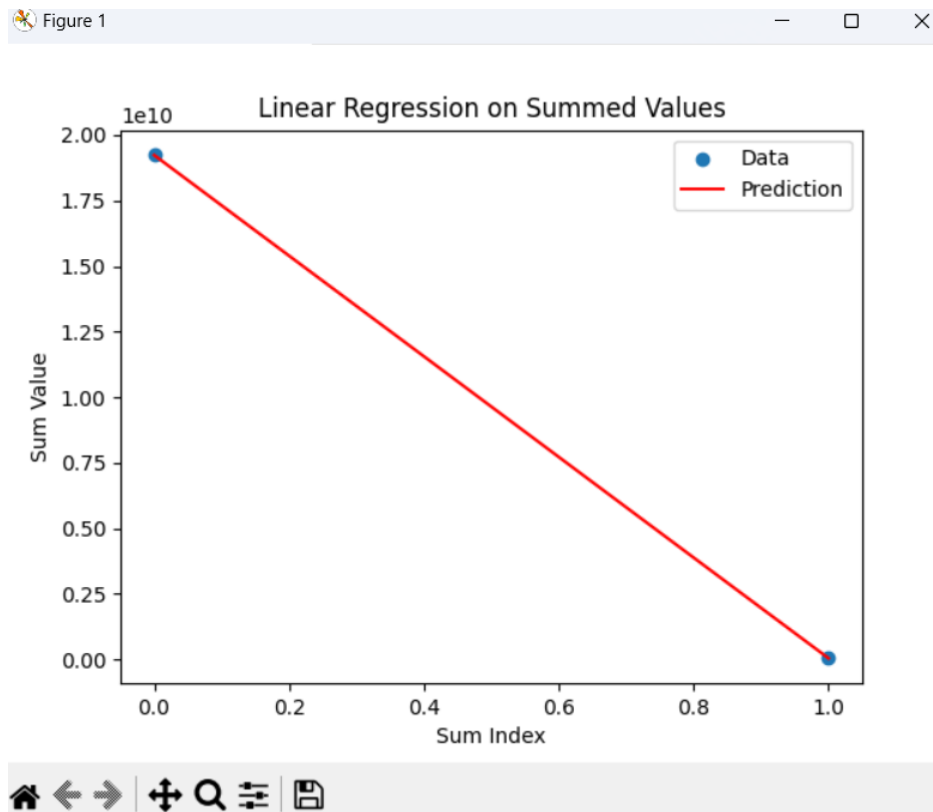
Korisnik može odabrati među pet ponuđenih algoritama. Za potrebe ovog testa odabrani algoritam je prvi ponuđeni, odnosno algoritam linearne regresije.

Prvi prozor koji se prikazuje vizualizacija je za odabrane sume kako bi korisnik mogao vizualno vidjeti vrijednosti suma za svaku kolonu.



Slika 11: Vrijednost suma (izvor: autor)

Na samom kraju treniranja skupa korisniku se prikazuje prozor koji prikazuje vizualizaciju odabranog algoritma, u ovom testu linearne regresije.



Slika 12. Linearna regresija rezultat (izvor: autor)

8. Zaključak

Izrada ovog završnog rada rezultirala je istraživanjem i uporabom područja računarskih znanosti koje su u velikom usponu. S pomoću različitih alata i metoda tijekom obrade ove teme dosegnete su osobne mogućnosti koje će biti od izrazite koristi u budućnosti odabranog zanimanja ili prilikom izrade budućih zadataka slične tematike.

Ovaj rad pogodan je za nadograđivanje, odnosno prilagođavanje potrebama korisnika i uvjetima tržišta. Raspon je upotrebe ovakvih sustava širok, ali mogu se izdvojiti područja kao što je medicina, financije, biometriji ili drugi sigurnosni sustavi.

Naposljetku, neke su od nadogradnji, koje se mogu dodati, drugačiji skupovi podataka koji ne moraju nužno biti numeričkog oblika. Mogu se dodati neki od algoritama te također postoji mogućnost boljeg vizualnog prikaza korisničkog sučelja. Sustav je prigodan za razna znanstvena istraživanja ili treniranje podataka i raspoznavanje uzoraka koji se nalaze svuda oko nas.

9. Izjava o autorstvu

MEĐIMURSKO VELEUČILIŠTE U ČAKOVCU

Bana Josipa Jelačića 22/a, Čakovec

IZJAVA O AUTORSTVU

Završni/diplomski rad isključivo je autorsko djelo studenta te student odgovara za istinitost, izvornost i ispravnost teksta rada. U radu se ne smiju koristiti dijelovi tuđih radova (knjiga, članaka, doktorskih disertacija, magistarskih radova, internetskih i drugih izvora) bez pravilnog citiranja. Dijelovi tuđih radova koji nisu pravilno citirani, smatraju se plagijatom i nezakonitim prisvajanjem tuđeg znanstvenog ili stručnoga rada. Sukladno navedenom studenti su dužni potpisati izjavu o autorstvu rada.

Ja, ELCORA MAJDAUŽIĆ (ime i prezime studenta) pod punom moralnom, materijalnom i kaznenom odgovornošću, izjavljujem da sam isključivi autor/ica završnog/diplomskog rada pod naslovom

ANALIZA TWITTER DRUŠTVENOG MREŽE

te da u navedenom radu nisu na nedozvoljeni način (bez pravilnog citiranja) korišteni dijelovi tuđih radova.

Student/ica:

Elcora Majdaušić

10.Literatura

Benšić, Grahovac, Mihalčić, Linearna regresija kroz primjer, Osječki matematički list, 2023. <https://hrcak.srce.hr/file/450496> (pristup 1. 9. 2024.)

Ethem Alpaydin (2020.), *Introduction to Machine Learning*, MIT Press

Geeks for Geeks, K-Nearest Neighbor(KNN) Algorithm <https://www.geeksforgeeks.org/k-nearest-neighbours/> (pristup 2. 9. 2024.)

Geeks for Geeks, K means Clustering – Introduction, <https://www.geeksforgeeks.org/k-means-clustering-introduction/> (pristup 2. 9. 2024.)

Geeks for Geeks, Decision Tree, <https://www.geeksforgeeks.org/decision-tree/> (pristup, 2. 9. 2024.)

Geeks for Geeks, Random Forest Algorithm in Machine Learning, https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/?ref=header_outind (pristup 2. 9. 2024.)

Mislav Acman, Odabrane metode strojnog učenja i njihova primjena u molekularnoj biologiji, Zagreb, 2015. <https://repozitorij.pmf.unizg.hr/islandora/object/pmf%3A2188/datastream/PDF/view> (pristup 1. 9. 2024.)

Nenad Bolf, Strojno učenje, Zagreb, 2021. <https://hrcak.srce.hr/file/382926> (pristup 29. 8. 2024.)

Skup podataka preuzet: <https://www.kaggle.com/datasets/ramjasmaurya/top-1000-social-media-channels> (pristup 20. 3. 2024.)

11. Popis ilustracija

<i>Slika 1. Priključenje matplotlib biblioteke u Python kôd (izvor: autor)</i>	2
<i>Slika 2. Priključenje pandas biblioteke u Python kôd (izvor: autor)</i>	3
<i>Slika 3. Primjer random forests regression grafikona (izvor: autor)</i>	7
<i>Slika 4. Primjer grafikona linearne regresije (izvor: autor)</i>	8
<i>Slika 5. Primjer grafikona algoritma k-najbližih susjeda (izvor: autor)</i>	9
<i>Slika 6. Primjer K-means Clustering grafikona (izvor: autor)</i>	10
<i>Slika 7. Primjer grafa Decision tree regression algoritma (izvor: autor)</i>	11
<i>Slika 8. Odabir skupa podataka (izvor: autor)</i>	16
<i>Slika 9. Odabir skupa podataka (izvor: autor)</i>	16
<i>Slika 10. Odabir algoritma (izvor: autor)</i>	17
<i>Slika 11. Vrijednost suma (izvor: autor)</i>	18
<i>Slika 12. Linearna regresija rezultat (izvor: autor)</i>	19