Univerzitet u Kragujevcu

Fakultet inženjerskih nauka

Seminarski rad iz predmeta

Veštačka inteligencija

Tema:

***ZOO***

Student: Predmetni professor:

Marta Ignjatović 601/2020 red. prof. dr Vesna Ranković

saradnik:dr Tijana Geroski

Sadržaj:

Uvod ............................................................................................................................................... 3

Razrada .......................................................................................................................................... 4

Zaključak ...................................................................................................................................... 10

Literatura ...................................................................................................................................... 11

*Uvod*

U ovom seminarskom radu, analizirala sam dataset koji sadrži informacije o različitim vrstama životinja koje se nalaze u zoo vrtu. Cilj mi je bio da istražim karakteristike svih zadatih životinja, identifikovati njihove uzorke i ponašanje, kao i klasifikovati ih u određene grupe na osnovu njihovih osobina.

Najviše sam se bazirala na kod i njegovu izradu dok sam podatke o samom zadatku dobila kroz zip file i u njemu podatke koje je bilo neophodno konvertovati u .csv kako bih imali pristup njima.

Sam kod sam pisala u Pythonu dok sam Matlab koristila kao okruženje koje mi je omogućilo pristupu grafičkog prikaza istih podataka.

*Razrada*

Cilj ovog zadatka je istraživanje i analiza podataka o različitim životinjama koje se nalaze u zoo vrtu. Neophodno je razumeti karakteristike ovih životinja, identifikovati uzorke u podacima i klasifikovati ih u određene grupe na osnovu njihovih osobina. Koristeći Python za obradu podataka i MATLAB za vizualizaciju, pristupamo datasetu koji sadrži informacije o 101 vrsti životinja, uključujući određene atribute poput prisustva krzna, perja, jaja, mleka, sposobnosti letenja, morskog života i dosta drugih.

Najpre samu izradu ovog seminarskog zadatka možemo razložiti na par celina.

1. Učitavanje podataka: Kao prvi korak uzećemo učitavanje dataset-a. Koristim Pzthon I pandas biblioteku za ovo a dataset sadrži informacije o 101 životinjskoj vrsti.
2. Analiza podataka: Nakon učitavanja, vršimo analizu dobijenih podataka kako bismo razumeli strukturu dataset-a. Ovo uključuje prikaz prvih nekoliko redova podataka, proveru tipova podataka, proveru nedostajućih vrednosti, kao i identifikovanje i uklanjanje duplikata.
3. Vizualizacija podataka: Koristimo Matlab za vizualizaciju podataka. Prikazujemo raspodelu klasa životinja, analiziramo osobine svake klase, kao i njihovu korelaciju između osobina.
4. Klasifikacija životinja: Koristimo algoritam odlučivanja (Decision Tree) za klasifikaciju životinja na osnovu njihovih osobina. Zatim delimo podatke na trening i test set, treniramo model na trening setu, a zatim testiramo performanse modela na test setu.

Nakon što smo detaljno razgraničili i rasčlanili sam zadatak i njegov tok i izradu, u daljem nastavku će se to i ispratiti kako kroz kod tako i kroz grafički prikaz.

import pandas as pd

# Definisemo kolone

columns = [

    'animal\_name', 'hair', 'feathers', 'eggs', 'milk', 'airborne', 'aquatic',

    'predator', 'toothed', 'backbone', 'breathes', 'venomous', 'fins', 'legs',

    'tail', 'domestic', 'catsize', 'class\_type'

]

# Ucitavamo podatke

data = pd.read\_csv('C:\\Users\\HP\\vestacka-i\\zoo.data', names=columns)

# Snimiju se podaci u CSV fajl

data.to\_csv('C:\\Users\\HP\\vestacka-i\\zoo.csv', index=False)

# Broj zivotinja po tipu klase

class\_counts = data['class\_type'].value\_counts().reset\_index()

class\_counts.columns = ['class\_type', 'count']

# Snimi brojanje u CSV fajl

class\_counts.to\_csv('C:\\Users\\HP\\vestacka-i\\class\_counts.csv', index=False)

Kod koji je gore opisan je python kod za učitavanje csv file kako bismo dalje mogli da koristimo sam file za učitavanje i klasifikaciju životinja. Ovaj kod nama daje file-ove (zoo, class\_counts) koje zapravo možemo da koristimo i učitavamo u daljem kodu za samu izradu.

Sledećim python kodom zapravo pokazali sam cilj zadatka.

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.model\_selection import train\_test\_split

from sklearn.tree import DecisionTreeClassifier

from sklearn.metrics import classification\_report

# Ucitaj podatke

columns = [

    'animal\_name', 'hair', 'feathers', 'eggs', 'milk', 'airborne', 'aquatic',

    'predator', 'toothed', 'backbone', 'breathes', 'venomous', 'fins', 'legs',

    'tail', 'domestic', 'catsize', 'class\_type'

]

data = pd.read\_csv('C:\\Users\\HP\\vestacka-i\\zoo.data', names=columns)

# Provera broja instanci i osobina

print(f"Broj instanci: {data.shape[0]}")

print(f"Broj osobina (kolona): {data.shape[1] - 2}")  # Oduzimamo 2 jer su 'animal\_name' i 'class\_type' kolone svaka za sebe

# Provera nedostajucih vrednosti

print("Nedostajuće vrednosti po kolonama:")

print(data.isnull().sum())

# Provera i uklanjanje duplikata

duplicates = data.duplicated(subset=['animal\_name'])

print(f"Broj duplikata: {duplicates.sum()}")

data = data.drop\_duplicates(subset=['animal\_name'])

# Prikaz prvih nekoliko redova podataka

print(data.head())

# Distribucija klasa

class\_counts = data['class\_type'].value\_counts()

class\_counts.plot(kind='bar')

plt.xlabel('Tip klase')

plt.ylabel('Broj zivotinja')

plt.title('Broj zivotinja po tipu klase')

plt.show()

# Analiziramo osobinu za svaku klasu

class\_groups = data.groupby('class\_type').mean()

print(class\_groups)

# Korelacija osobina

correlation\_matrix = data.drop(columns=['animal\_name', 'class\_type']).corr()

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm')

plt.title('Correlation between Features')

plt.show()

# Priprema podataka

X = data.drop(columns=['animal\_name', 'class\_type'])

y = data['class\_type']

# Podela podataka na trening i test set

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.3, random\_state=42)

# Kreiranje i treniranje modela

clf = DecisionTreeClassifier()

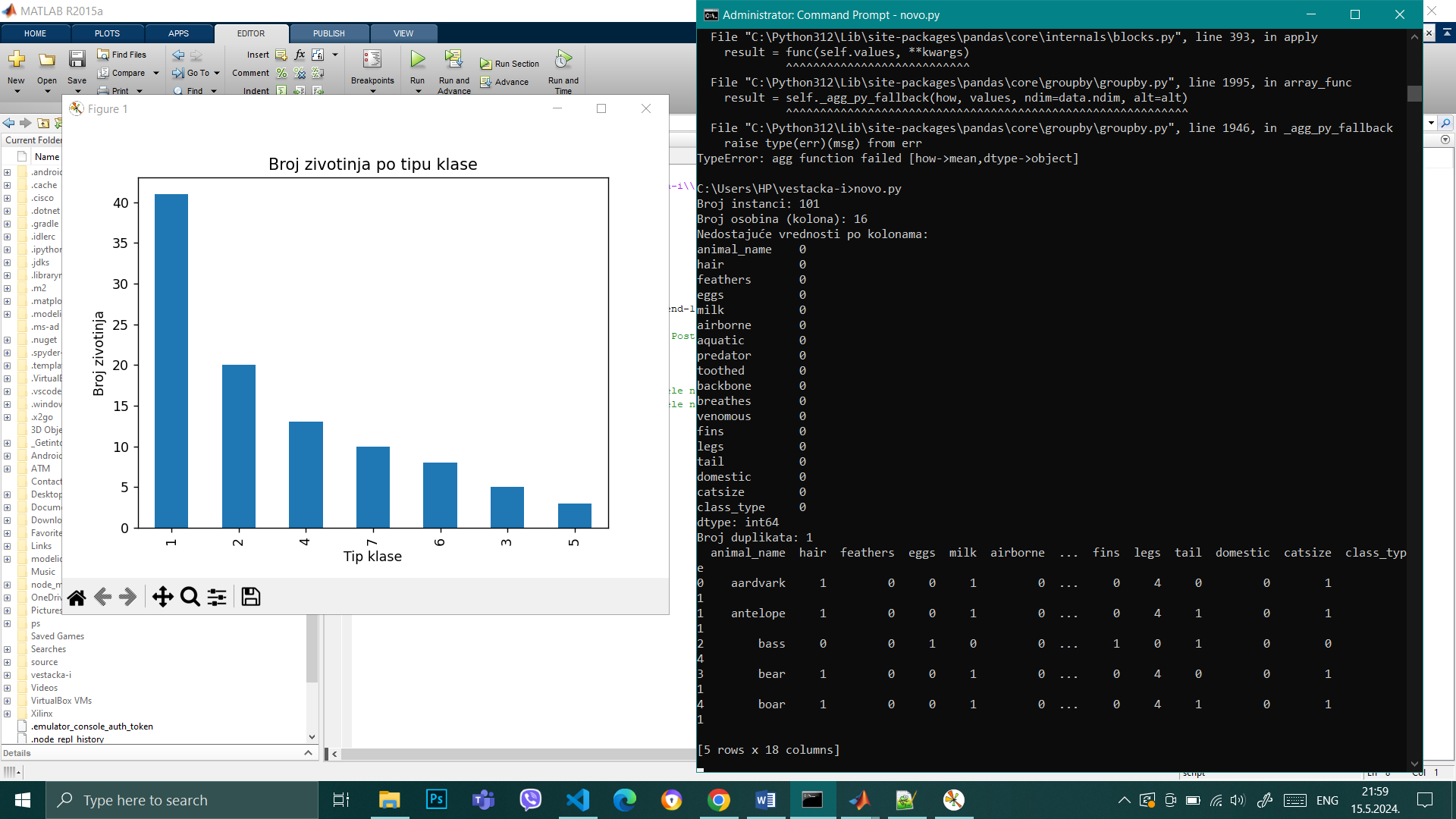
clf.fit(X\_train, y\_train)

# Predikcija i evaluacija

y\_pred = clf.predict(X\_test)

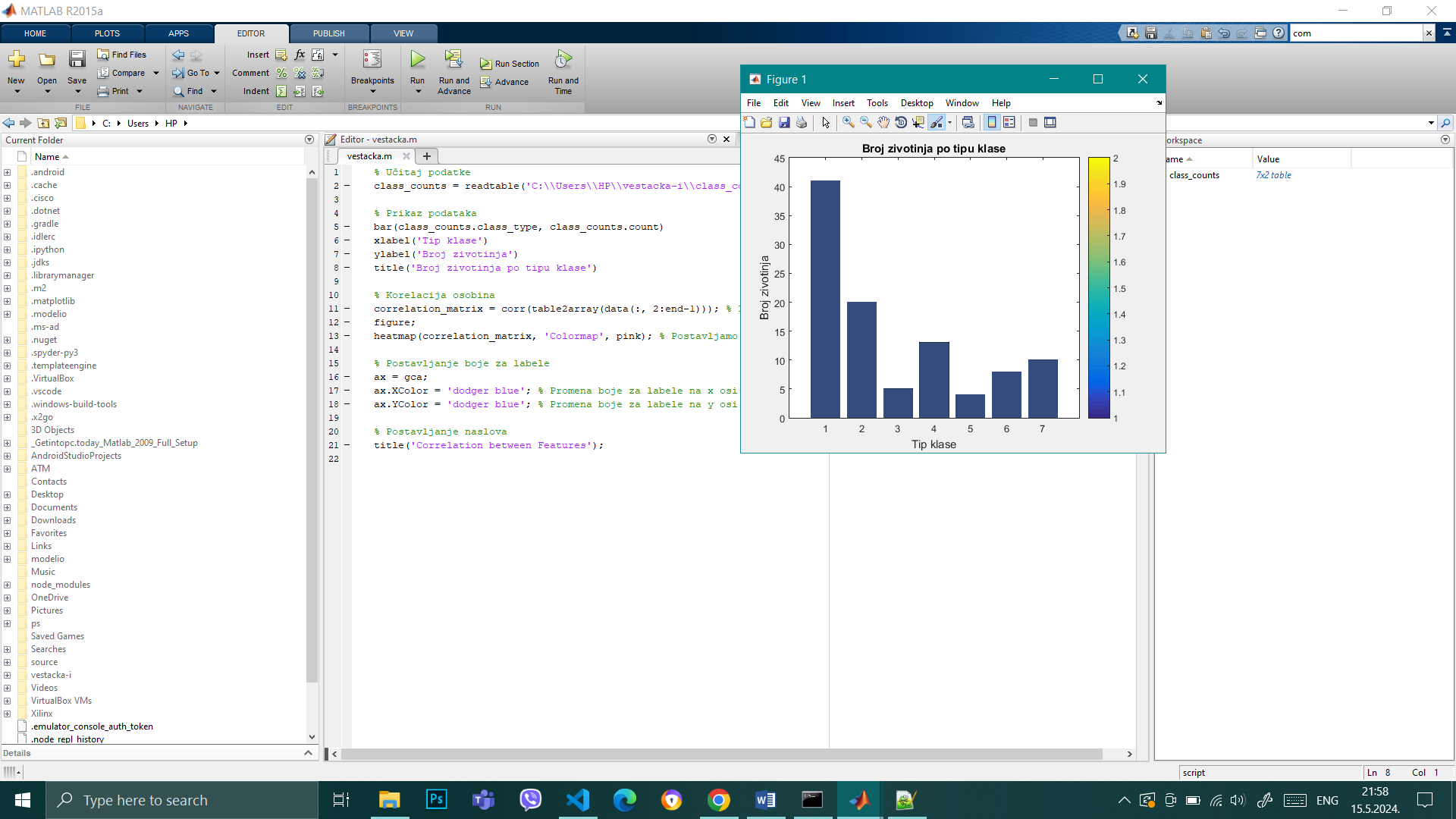
print(classification\_report(y\_test, y\_pred))

Izlaz iz Python koda u cmd-u zajedno sa grafičkim prikazom je prikazan na slici ispod (slika 1).



*Slika 1*

Slika 2 je grafički prikaz u Matlabu zajedno sa kodom koji je takođe napisan u Matlabu.



*Slika 2*

*Zakljucak:*

Analiza podataka o životinjama u zoo vrtu omogućila nam je dublje razumevanje karakteristika i ponašanja algoritama veštačke inteligencije. Kroz kombinaciju Python-a za njenu obradu podataka i korišćenje Matlab-a za njenu vizualizaciju, uspešno smo istražili dataset, identifikovane uzorke i klasifikovali životinje na osnovu njihovih osobina. Ovaj pristup nam je omogućio dalje istraživanje i razumevanje bitnosti neuronskih mreža u globalu. Takođe povezivanje dva totalno različita jezika kako bismo ne samo istražili i analizirali podatke, već i prikaz istih kroz grafiku koju nam pruza Matlab.

*Literatura:*

1. <http://moodle.fin.kg.ac.rs/course/view.php?id=989>
2. Vesna Ranković, Veštačka inteligencija, skripta, Mašinski fakultet, Kragujevac, 2008.
3. Vesna Ranković, Inteligentno upravljanje, Mašinski fakultet, Kragujevac, 2008.
4. https://archive.ics.uci.edu/dataset/111/zoo
5. Miroslav Jocković, Zoran Ognjanović, Stevan Stankovski, Veštačka inteligencija, inteligentne mašine i sistemi, Beograd, 1997.
6. Dr Marko Marković, M.Sc Ivan Pantelić, M.Sc Jelena Kaljević, dr Biljana Tešić, Poslovni fakultet Valjevo, Univerzitet Singidunum
7. Subašić, Fazi logika i neuronske mreže, Tehniĉka knjiga, 1997.
8. Russell, Norvig, Artificial Intelligence A Modern Approach, Second Edition, Prentice Hall, 2003.
9. Jackson, Introduction to Expert Systems, Third Edition, Addison-Wesley, 1999.
10. Devedţić, Ekspertni sistemi za rad u realnom vremenu, Institut Mihajlo Pupin, 1994