**18CS3166S-Machine Learning**

**PROJECT BASED REPORT**

On

**PREDICTION OF PARKINSON’S DISEASE USING GAIT DATASET**

submitted in partial fulfillment of the requirement for the award of the degree of

**BACHELOR OF TECHNOLOGY**

In

**COMPUTER SCIENCE AND ENGINEERING**

By

T.N.S. Rama Charan (180030326)

N Naga Teja (180030346)

K Harshith (180030358)

Under the Esteemed Guidance of

Dr. Sagar Imambi

******

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**K L (Deemed to be) University**

Green Fields, Vaddeswaram, Guntur District – 522 502

**(2020-2021)**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**



**DECLARATION**

This is certify that the project based report entitled **“Parkinson Disease Prediction”** is a bona fide work done and submitted by **T.N.S. Rama Charan (180030326), N Naga Teja (180030346), K Harshith (180030358)** in partial fulfillment of the requirements for the award of the degree of BACHELOR OF TECHNOLOGY in Department of Computer Science Engineering, K L (Deemed to be University), Guntur District during the academic year 2019-2020.

**ACKNOWLEDGEMENT**

The success in this project would not have been possible but for the timely help and guidance rendered by many people. Our sincere thanks to all those who has assisted us in one way or the other for the completion of my project.

Our greatest appreciation to our guide **Dr. Sagar Imambi**, Professor, Department of Computer Science and Engineering which cannot be expressed in words for her tremendous support, encouragement and guidance for this project.

We express our gratitude to **Mr. V Hari Kiran,** Head of the Department for Computer Science and Engineering for providing us with adequate facilities, ways and means by which we are able to complete this project.

We thank all the members of teaching and non-teaching staff members, and also who have assisted me directly or indirectly for successful completion of this project.

Finally, we sincerely thank our friends and classmates for their kind help and co- operation during our work.

**INDEX**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **Title** | **Page No** |
| **1** | Introduction | 5 |
| **2** | Aim of the Project | 6 |
| **3** | Software Requirements | 6 |
| **4** | Hardware Requirements | 6 |
| **5** | Implementation | 7-8 |
| **6** | Source Code | 9-11 |
| **7** | Output Screenshots | 12-15 |
| **8** | Conclusion | 16 |
| **9** | References | 16 |

**INTRODUCTION:**

**Parkinson’s Disease:**

A disorder of the central nervous system that affects movement, often including tremors. Nerve cell damage in the brain causes dopamine levels to drop, leading to the symptoms of Parkinson's. Parkinson's often starts with a tremor in one hand. Other symptoms are slow movement, stiffness and loss of balance.

We predict if a person is effect by Parkinson or not by using gait dataset.

**Gait Dataset**

Dataset A (former NLPR Gait Database) was created on Dec. 10, 2001, including 20 persons. Each person has 12 image sequences, 4 sequences for each of the three directions,

**Symptoms**

Tremor: can occur at rest, in the hands, limbs, or can be postural

Muscular: stiff muscles, difficulty standing, difficulty walking, difficulty with bodily movements, involuntary movements, muscle rigidity, problems with coordination, rhythmic muscle contractions, slow bodily movement, or slow shuffling gait

Sleep: early awakening, nightmares, restless sleep, or sleep disturbances

Whole body: fatigue, dizziness, poor balance, or restlessness

Cognitive: amnesia, confusion in the evening hours, dementia, or difficulty thinking and understanding

Speech: difficulty speaking, soft speech, or voice box spasms

Nasal: distorted sense of smell or loss of smell

Urinary: dribbling of urine or leaking of urine

Mood: anxiety or apathy

Facial: jaw stiffness or reduced facial expression

Also common: blank stare, constipation, depression, difficulty swallowing, drooling, falling, fear of falling, loss in contrast sensitivity, neck tightness, small handwriting, trembling, unintentional writhing, or weight loss

**Treatments**

Medications like-Dopamine promoter, Antidepressant, Cognition-enhancing medication and Anti-Tremor

**AIM OF THE PROJECT:**

The aim of the project is to predict whether a person will get Parkinson disease or not.

**SOFTWARE REQUIRED:**

1.Python

2.Jupyter Notebook

**HARDWARE REQUIREMENTS:**

The hardware requirements that map towards the software are as follows:

RAM 8GB

Processor Intel® core™ i3-6006u CPU @ 2.00GHz 2.00 GHz.

System type: 64-bit operating system, x64-based processor.

**IMPLEMENTATION:**

The implementation has 7 parts in it.,

1 Dataset Selection  
2 Data Preprocessing  
3 Data Visualization  
4 Splitting the Data  
 I Training  
 II Testing  
5 Object Creation  
6 Fitting the Data  
7 visualizing the Metrics

**1 Dataset Selection**

we used gaits data set for the prediction

**2** **Data Preprocessing**

Removed unnecessary columns

Filled Null values

Removed Noisy Data

Removed improper values

**3 Data Visualization**

In visualization we are trying to check for any missing data.

We are using heatmap for this process.

**4 Splitting the Data**

Separating data into training and testing sets is an important part of evaluating data mining models. By using similar data for training and testing, you can minimize the effects of data discrepancies and better understand the characteristics of the model. **I Training**:

Training data is the data you use to train an algorithm or machine learning model to predict the outcome you design your model to predict. **II Testing:**

Test data is used to measure the performance, such as accuracy or efficiency, of the algorithm you are using to train the machine.

**5 Object Creation**

We use some methods to classify our problem for those classification we use some methods to apply those methods.

In this project we have created objects for Decision tree and Random forest.

**6 Fitting the Data**

Model fitting is a measure of how well a machine learning model generalizes to similar data to that on which it was trained.

During the fitting process, you run an algorithm on data for which you know the target variable, known as “labeled” data, and produce a machine learning model.

**7 visualizing the Metrics**

Accuracy prediction

R2 score

confusion matrix

**CODE:**

***Using Decision Tree***

import numpy as np

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

import numpy as np

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 DecisionTreeRegressor

from sklearn.metrics import accuracy\_score

from sklearn.metrics import confusion\_matrix

from sklearn.metrics import classification\_report

import math

from sklearn.metrics import mean\_squared\_error,r2\_score

ds=pd.read\_csv(r"F:\Project\_4\_Section\_S01.csv",header=0,encoding='unicode\_escape')

plt.figure(figsize=(7,7))

sns.heatmap(ds.isnull())

ds=ds.fillna(ds.mean())

plt.figure(figsize=(7,7))

sns.heatmap(ds.isnull())

X=ds.drop(['class'],axis=1)

y=ds['class']

x\_train,x\_test,y\_train,y\_test=train\_test\_split(X,y,test\_size=0.2,random\_state=10)

tree=DecisionTreeClassifier(max\_depth=4,random\_state=10)

tree.fit(x\_train,y\_train)

y\_pred=tree.predict(x\_test)

print("Accuracy:",accuracy\_score(y\_test,y\_pred))

print("confusion matrix :",confusion\_matrix(y\_pred,y\_test))

print("classification report :\n",classification\_report(y\_pred,y\_test))

print(tree.predict(x\_test.head(5)))

***Using Random Forest Classifier***

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

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

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy\_score

from sklearn.metrics import confusion\_matrix

from sklearn.metrics import classification\_report

ds=pd.read\_csv(r"F:\Project\_4\_Section\_S01.csv",header=0,encoding='unicode\_escape')

#print(ds.isnull().sum())

plt.figure(figsize=(7,7))

sns.heatmap(ds.isnull())

ds=ds.fillna(ds.mean())

plt.figure(figsize=(7,7))

sns.heatmap(ds.isnull())

#print(ds.isnull().sum())

X=ds.drop(['class'],axis=1)

y=ds['class']

x\_train,x\_test,y\_train,y\_test=train\_test\_split(X,y,test\_size=0.2,random\_state=100)

#print(x\_test.head(2))

regressor=RandomForestClassifier(n\_estimators=100,random\_state=10)

model=regressor.fit(x\_train,y\_train)

y\_pred=model.predict(x\_test)

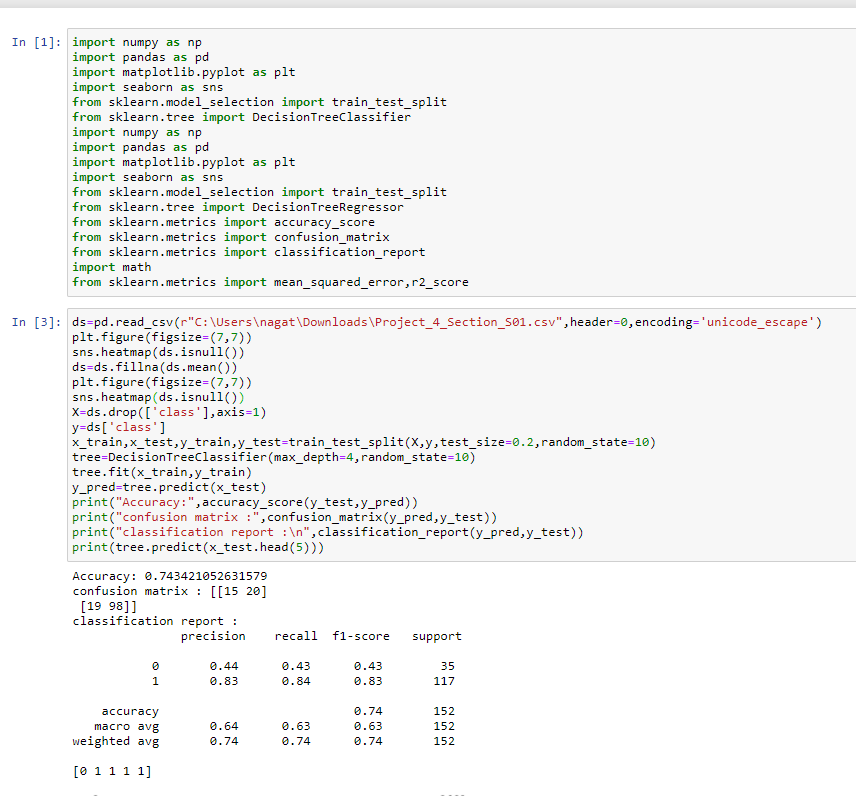
print("Accuracy:",accuracy\_score(y\_test,y\_pred))

print("confusion matrix :",confusion\_matrix(y\_pred,y\_test))

print("classification report :\n",classification\_report(y\_pred,y\_test))

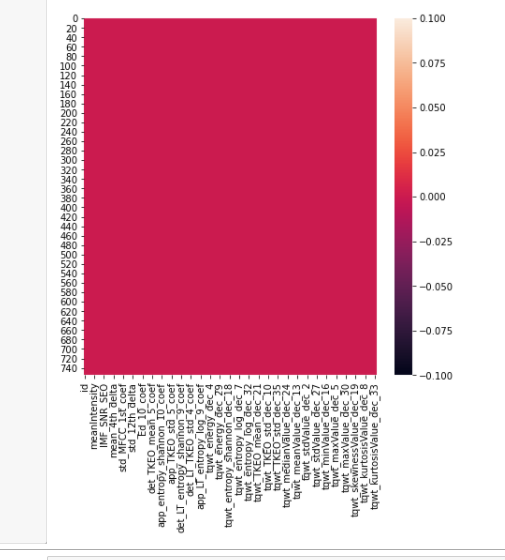
print(model.predict(x\_test.head(2)))

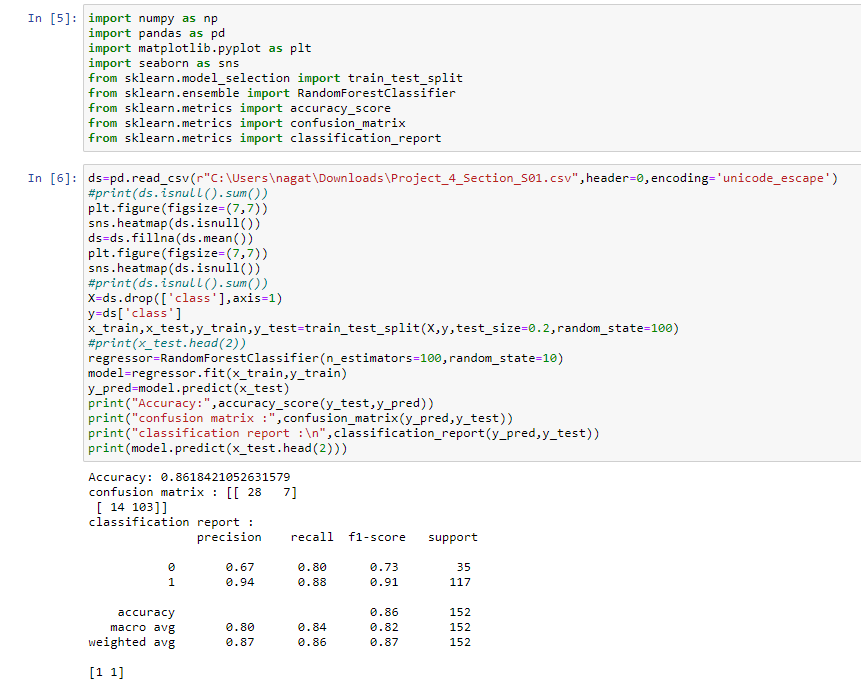
**SCREENSHOTS:**

****

A picture containing shape

Description automatically generated





A picture containing chart

Description automatically generated

A picture containing chart

Description automatically generated

**CONCLUSION:**

We get better accuracy score and better model for prediction using random forest classifier compared to decision tree as we can say forest classifier is a developed version of decision tree model.

This model is just for prediction, cannot be used by doctors or scientists for base.

**REFERENCES:**

https://www.kaggle.com/

https://towardsdatascience.com/

https://www.javatpoint.com/