COLLEGE OF ENGINEERING
Since 1982

An Autonomous Institute

(Affiliated to Visvesvaraya Technological University, Belagavi
Approved By AICTE, New Delhi,
Recognized by UGC under 2(f) & 12(B)
Accredited by NBA and NAAC)

**A MINI PROJECT REPORT**
ON

# PARKINSON'S DISEASE DETECTION USING MACHINE LEARNING

Submitted in partial fulfillment of requirements for the award of 6th semester,

**BACHELOR OF ENGINEERING**

**IN**

**ARTIFICIAL INTELLIGENCE & MACHINE LEARNING**

Submitted By:
**CICIYA SEBASTIN**
**1MJ20AI010**

Under the Guidance of
**Ms. Anusha S**
Assistant Professor, Department of Computer Science & Engineering,

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE & MACHINE LEARNING**
**MVJ COLLEGE OF ENGINEERING**
**BENGALURU-67**
**ACADEMIC YEAR 2022-23**

# MVJ COLLEGE OF ENGINEERING

**Near ITPB, Whitefield, Bengaluru-67**

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE & MACHINE LEARNING**

An Autonomous Institute
(Affiliated to Visvesvaraya Technological University, Belagavi
Approved By AICTE, New Delhi,
Recognized by UGC under 2(f) & 12(B)
Accredited by NBA and NAAC)

# CERTIFICATE

This is to certify that the mini-project work, entitled **"PARKINSON'S DISEASE DETECTION USING MACHINE LEARNING"** is a bonafide work carried out by '**CICIYA SEBASTIN(1MJ20AI010)**' in partial fulfillment for the award of the degree of Bachelor of Engineering in ARTIFICIAL INTELLIGENCE & MACHINE LEARNING during the academic year 2022-23. It is certified that all the corrections/suggestions indicated for Internal Assessment have been incorporated in the Report. The mini-project report has been approved as it satisfies the academic requirements.

**Signature of Guide**
**Ms. Anusha S**

**Signature of the HOD**
**Dr. Kiran Babu T S**

Name of examiners:
1.
2.

Signature with date:

# MVJ COLLEGE OF ENGINEERING

**Whitefield, Near ITPB, Bengaluru-67**

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE & MACHINE LEARNING**

COLLEGE OF ENGINEERING
Since 1982
An Autonomous Institute
(Affiliated to Visvesvaraya Technological University, Belagavi
Approved By AICTE, New Delhi,
Recognized by UGC under 2(f) & 12(B)
Accredited by NBA and NAAC)

## DECLARATION

I, **CICIYA SEBASTIN** hereby declare that the entire work titled "**PARKINSON'S DISEASE DETECTION USING MACHINE LEARNING"** embodied in this mini project report has been carried out by us during the 6$^{th}$ semester of BE degree at MVJCE, Bengaluru under the esteemed guidance of **Ms. Anusha S**, Assistant Prof, Dept. of Computer Science and Engineering, MVJCE. The work embodied in this dissertation work is original and it has not been submitted in part or full for any other degree in any University.

CICIYA SEBASTIN
1MJ20AI010

Place:

Date:

# ABSTRACT

Parkinson's Disease is a nervous system disorder and is progressive in nature. It affects body movements. Symptoms begin with a hardly noticeable tremor in body parts. The disorder results in stiffness and a slowing down of movements. It’s a debilitating neurodegenerative disease and cannot be diagnosed through blood tests. Parkinson’s disease mostly affects people above 60 and is one of the most common diseases among war veterans. Currently, available systems for PD detection are high-priced and less efficient. Hence, there was a need for a faster and cheaper diagnostic tool. The project uses ML algorithms to analyze the variations in voice patterns to predict the existence of Parkinson's Disease in patients. These data of varying frequencies can be fed to the model and the results can be compared to find the people who are affected by the PD and display the result in an application that can be used by the patient to get the result.

# ACKNOWLEDGEMENT

With gratitude, I acknowledge all those whose guidance and encouragement served as a beacon of light and crowned our effort with success.

I am thankful to the **Management of MVJ College of Engineering Bengaluru** for their continuous support and encouragement in carrying out the mini-project work.

I am thankful to our **Dr. M. Brindha,** Principal, MVJ College of Engineering, Bengaluru for her constant inspiration and for providing all the facilities that were needed throughout the mini-project work.

I also like to express our sincere gratitude to **Dr. M A Lourdu Antony Raj**, Controller of Examinations, MVJ College of Engineering, Bengaluru, for his persistent guidance.

I am thankful to our **Dr. Kiran Babu T S**, HOD of Computer Science Engineering, MVJ College of Engineering, Bengaluru, for his constant support and providing all the facilities that were needed throughout the mini-project work.

I consider it as a privilege and honor to express our sincere gratitude to our guide **Ms. Anusha S,** Assistant Professor**,** Dept. of ARTIFICIAL INTELLIGENCE & MACHINE LEARNING, MVJ College of Engineering, Bengaluru for her encouragement that has been a constant source of motivation to us for successful completion of our project.

It’s also a great pleasure to express our deepest gratitude to all the other faculty members of our department for their cooperation and constructive criticism offered, which helped us a lot during our mini-project work.

Finally, I would like to thank all our family members and friends whose encouragement and support was invaluable.

Thank you.

# TABLE OF CONTENTS

# LIST OF FIGURES

# CHAPTER 1

## INTRODUCTION

Parkinson's disease is a neurodegenerative disorder that affects the nervous system and is characterized by a variety of motor symptoms such as tremors, rigidity, and bradykinesia. Early detection and diagnosis of Parkinson's disease can greatly improve the quality of life for affected individuals, but it can be a challenging task for healthcare professionals due to the complexity of the disease.

Machine learning techniques offer promising solutions for early detection and diagnosis of Parkinson's disease. By analyzing large amounts of data, such as medical records, brain imaging, and movement patterns, machine learning algorithms can identify patterns and relationships that may not be immediately apparent to human observers.

In this project, we aim to develop a machine learning model for the early detection of Parkinson's disease. We will use a vocal dataset consisting of various features such as minimum vocal frequency, maximum vocal frequency, shimmers, jitters and so on of individuals with and without Parkinson's disease. Our goal is to build a model that can accurately predict the presence of Parkinson's disease based on these features.

In this model, the huge data is collected from previously affected person and then by using machine learning algorithm will process the user input data with previous data to check he/she affected.

This project has the potential to make a significant impact in the field of Parkinson's disease diagnosis and treatment by providing healthcare professionals with a powerful tool for early detection and intervention.

# CHAPTER 2

# LITERATURE SURVEY

Title 1: Using Machine Learning to Diagnose Parkinson’s Disease from Voice Recordings

Author: Akshaya Dinesh, Jennifer

Year: 2017

Description: “The author tested the ML model using Microsoft Azure Machine Learning Studio and found the best-suited model to be Two-Class Boosted Decision Trees, an ensemble model of boosted regression trees made in a stepwise method. Another conclusion that was derived was that Spread1, spread2, and PPE in the dataset, had the strongest weights in labeling a patient as healthy or not. The results were that we successfully created a predictive model for PD from voice analysis. The best accuracy was found using Boosted Decision Trees and the accuracy achieved was 95%.”

Title 2: Prediction of Parkinson’s disease using Machine Learning Techniques

Author: Kirti Sharma

Year: 2018

Description: “The thesis focuses on the speech articulation difficulty symptoms of PD affected people and formulates the model using various machine learning techniques such as neural networks, decision tree, random forest and linear regression. The author evaluates the performance of these classifiers using various metrics i.e. accuracy, receiver operating characteristic curve (ROC),Sensitivity, precision, specificity.” The author found out that using Boruta feature selection and Random Forest algorithm gives an accuracy of 95%. This outperformed all the other ML algorithms.

Title 3: Voice-Based Detection of Parkinson’s Disease through Ensemble Machine Learning Approach

Author: Iqra Nissar1, Danish Raza Rizvi1, Sarfaraz Masood, Aqib Nazir Mir

Year: 2019

Description: “This paper aims to analyze the effect of feature type selection i.e. MFCC and TQWT on the efficiency of voice-based PD detection systems along with the use of an ensemble learning based classifier for this task. In this paper, the author performed the XGBoost algorithm using mRMR feature selection which outperformed all other models with a high accuracy of 95% with AUC value around.

Title 4: Parkinson’s Disease Diagnosis Using Machine Learning and Voice

Author: Timothy Wroge, Yasin O¨ zkanca,Cenk Demiroglu,Dong Si,David C

Year: 2019

Description: “This paper explores the effectiveness of using supervised classification algorithms, such as deep neural networks, to accurately diagnose individuals with the disease. The author‟s peak accuracy was of 85% using AVEC selected feature and Gradient Boosted Decision Tree exceeding the average clinical diagnosis accuracy of non-experts (73.8%) and average accuracy of movement disorder specialists.

# CHAPTER 3

# PROBLEM ANALYSIS

The problem statement for this study is to develop a robust and accurate machine learning model for the early detection of Parkinson's disease using sensor data, overcoming the challenges of limited data and interpretability, and improving the accuracy and reliability of diagnosis for better patient outcomes.

## CHAPTER 4

# EXISTING SYSTEM

The existing systems for Parkinson's disease detection involve a combination of clinical assessments, medical imaging techniques, and specific diagnostic tests. These methods are typically performed by healthcare professionals, such as neurologists or movement disorder specialists. Imaging techniques like magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomography (PET) scans can be used to help rule out other potential causes of symptoms and provide additional information about the brain structures and functions related to Parkinson's disease.

### DISADVANTAGES:

- Limited accuracy
- Invasiveness
- Cost
- Time-consuming

# CHAPTER 5

# PROPOSED SYSTEM

The proposed system of Parkinson's disease detection using machine learning involves the development of a robust and accurate model for early detection of the disease using sensor data. The system includes the data collection where Vocal data will be collected from patients with and without Parkinson's disease. The collected sensor data will be preprocessed to remove noise and standardize the data for further analysis. Relevant features will be selected from the preprocessed sensor. Different machine learning algorithms, such as support vector will be trained on the selected features to develop a robust and accurate model for Parkinson's disease detection. The developed model will be evaluated on a separate test dataset to assess its accuracy and reliability in detecting Parkinson's disease. Model will be trained to accept and predict new data values. The model will be optimized to improve its performance and reduce the risk of overfitting.

## ADVANTAGES:

- Early detection
- Non-invasive
- Cost-effective
- Scalable
- Continuous learning and improvement of model

# CHAPTER 6

# SYSTEM REQUIREMENTS

## 6.1 Hardware Requirements:

- Processor **:** Pentium IV 2.4 GHz.
- Input Device **:** webcam, Keyboard, Monitor and Mouse.
- Storage **:** 20GB
- RAM **:** 4GB(min)

## 6.2 Software Requirements:

- Operating System **:** Windows 11
- Python 9.16
- Platform : pycharm.
- Libraries : cv2, dlib, cmake.
- OpenCV 3.1

# CHAPTER 7

# IMPLEMENTATION

## 7.1 System architecture

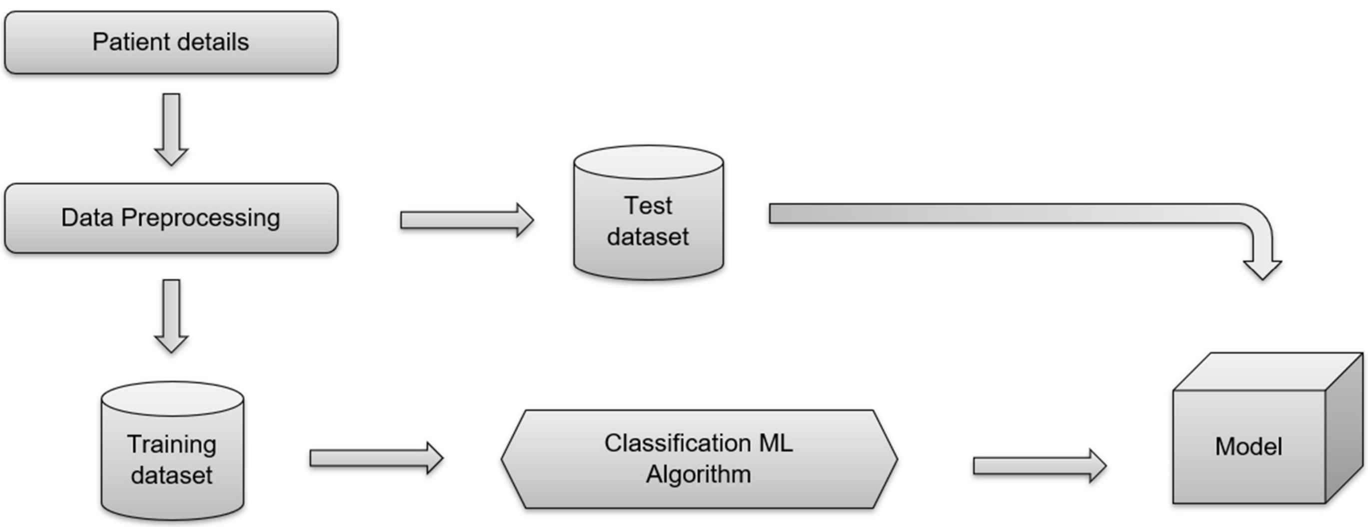



Fig 7.1: System Architecture

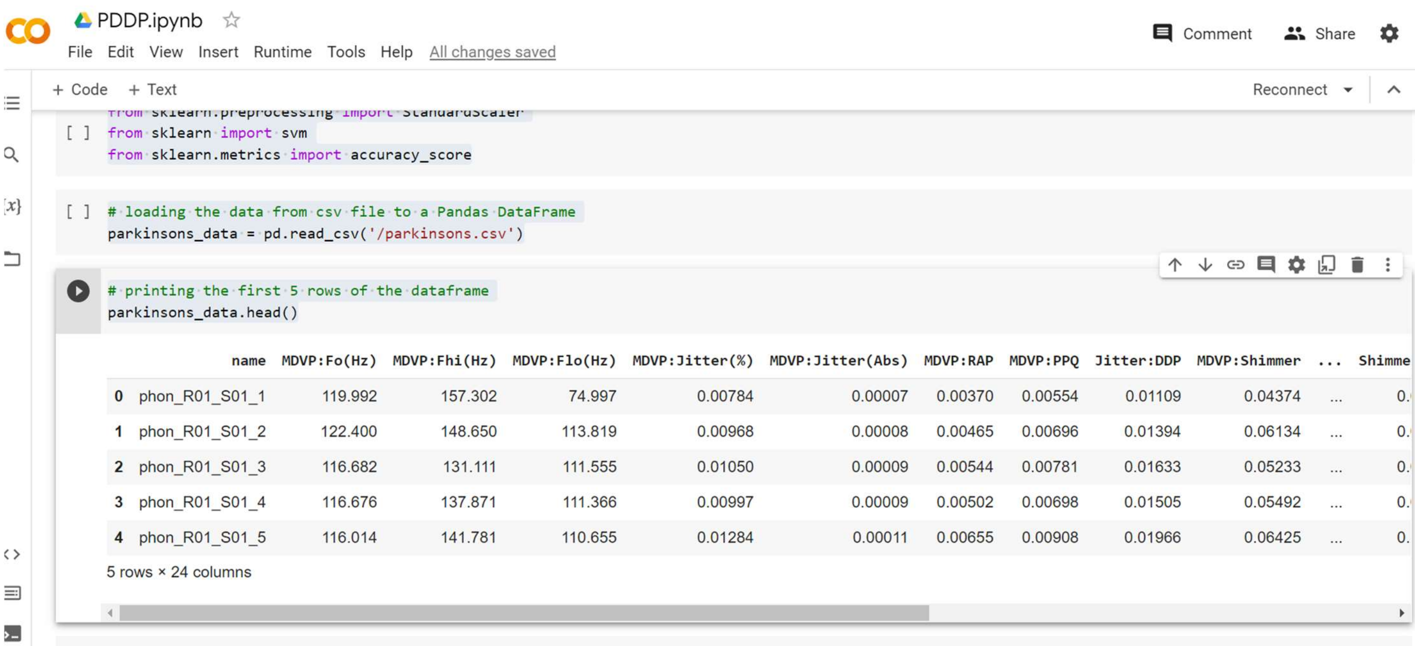
## 7.2 Implementation Code

```
import numpy as np
 import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn import svm
from sklearn.metrics import accuracy_score
# loading the data from csv file to a Pandas DataFrame
parkinsons_data = pd.read_csv('/parkinsons.csv')
# printing the first 5 rows of the dataframe
parkinsons_data.head()
parkinsons_data.shape
parkinsons_data.info()
parkinsons_data.isnull().sum()
parkinsons_data.describe()
parkinsons_data['status'].value_counts()
parkinsons_data.groupby('status').mean()
X = parkinsons_data.drop(columns=['name','status'], axis=1)
Y = parkinsons_data['status']
print(X)
print(Y)
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=2)
print(X.shape, X_train.shape, X_test.shape)
scaler.fit(X_train)
X_train = scaler.transform(X_train)
X_test = scaler.transform(X_test)
print(X_train)
model = svm.SVC(kernel='linear')
model.fit(X_train, Y_train)
X_train_prediction = model.predict(X_train)
training_data_accuracy = accuracy_score(Y_train, X_train_prediction)
print('Accuracy score of training data : ', training_data_accuracy)
X_test_prediction = model.predict(X_test)
test_data_accuracy = accuracy_score(Y_test, X_test_prediction)
print('Accuracy score of test data : ', test_data_accuracy)
```

```
input_data =
(119.99200,157.30200,74.99700,0.00784,0.00007,0.00370,0.00554,0.01109,0.04374,0.42600,0.0
2182,0.03130,0.02971,0.06545,0.02211,21.03300,0.414783,0.815285,-
4.813031,0.266482,2.301442,0.284654)

# changing input data to a numpy array
input_data_as_numpy_array = np.asarray(input_data)

# reshape the numpy array
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

# standardize the data
std_data = scaler.transform(input_data_reshaped)

prediction = model.predict(std_data)
print(prediction)

if (prediction[0] == 0):
  print("The Person does not have Parkinsons Disease")

else:
  print("The Person has Parkinsons")
```

# CHAPTER 8

# RESULT



PDDP.ipynb

File Edit View Insert Runtime Tools Help All changes saved

```
from sklearn.preprocessing import StandardScaler
from sklearn import svm
from sklearn.metrics import accuracy_score
```

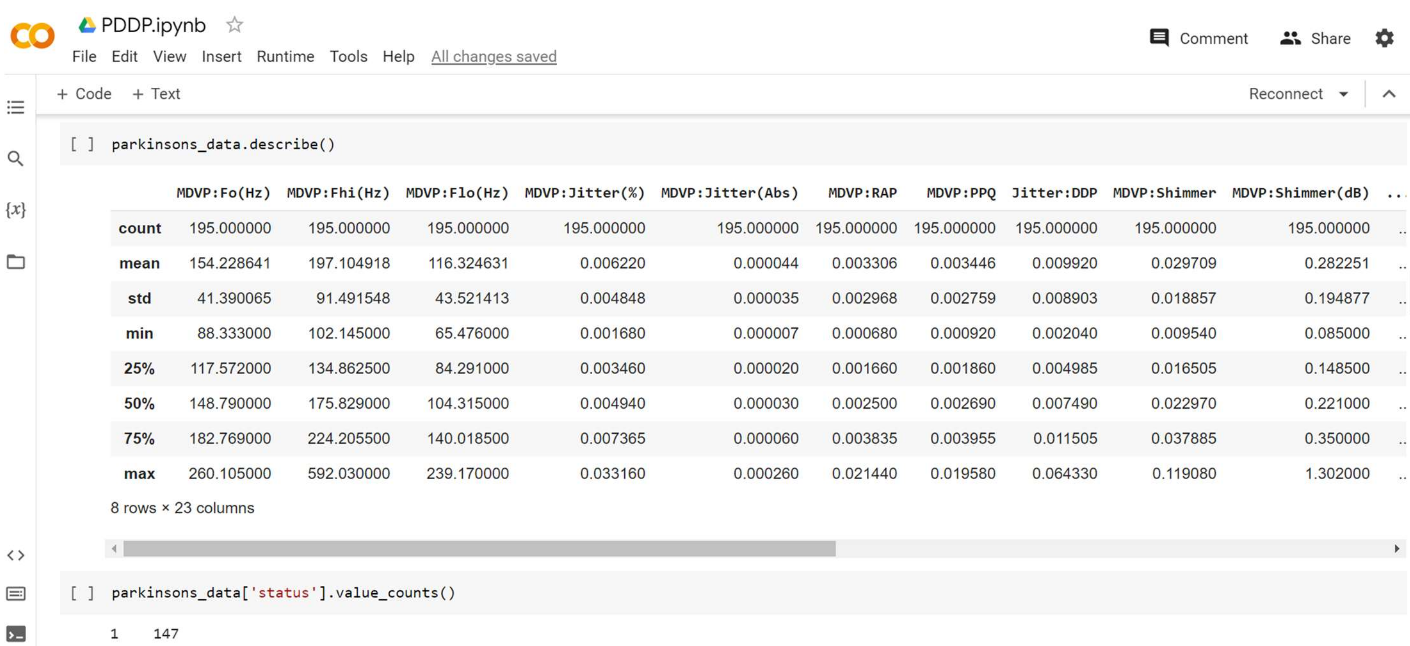
```
# loading the data from csv file to a Pandas DataFrame
parkinsons_data = pd.read_csv('/parkinsons.csv')
```

```
# printing the first 5 rows of the dataframe
parkinsons_data.head()
```

| | name | MDVP:Fo(Hz) | MDVP:Fhi(Hz) | MDVP:Flo(Hz) | MDVP:Jitter(%) | MDVP:Jitter(Abs) | MDVP:RAP | MDVP:PPQ | Jitter:DDP | MDVP:Shimmer | ... | Shimme |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | phon_R01_S01_1 | 119.992 | 157.302 | 74.997 | 0.00784 | 0.00007 | 0.00370 | 0.00554 | 0.01109 | 0.04374 | ... | 0. |
| 1 | phon_R01_S01_2 | 122.400 | 148.650 | 113.819 | 0.00968 | 0.00008 | 0.00465 | 0.00696 | 0.01394 | 0.06134 | ... | 0. |
| 2 | phon_R01_S01_3 | 116.682 | 131.111 | 111.555 | 0.01050 | 0.00009 | 0.00544 | 0.00781 | 0.01633 | 0.05233 | ... | 0. |
| 3 | phon_R01_S01_4 | 116.676 | 137.871 | 111.366 | 0.00997 | 0.00009 | 0.00502 | 0.00698 | 0.01505 | 0.05492 | ... | 0. |
| 4 | phon_R01_S01_5 | 116.014 | 141.781 | 110.655 | 0.01284 | 0.00011 | 0.00655 | 0.00908 | 0.01966 | 0.06425 | ... | 0. |

5 rows × 24 columns

**Fig 8.1**: Code snippets on google colab 1



PDDP.ipynb

File Edit View Insert Runtime Tools Help All changes saved

```
parkinsons_data.describe()
```

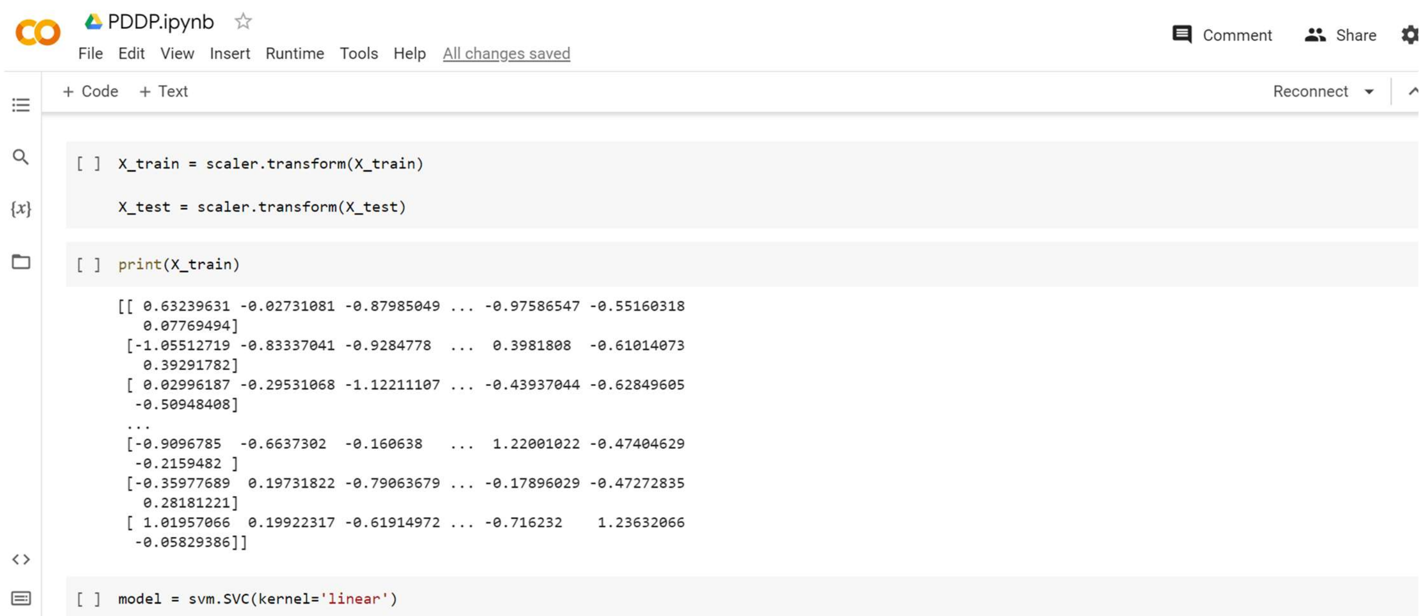
| | MDVP:Fo(Hz) | MDVP:Fhi(Hz) | MDVP:Flo(Hz) | MDVP:Jitter(%) | MDVP:Jitter(Abs) | MDVP:RAP | MDVP:PPQ | Jitter:DDP | MDVP:Shimmer | MDVP:Shimmer(dB) | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | 195.000000 | .. |
| mean | 154.228641 | 197.104918 | 116.324631 | 0.006220 | 0.000044 | 0.003306 | 0.003446 | 0.009920 | 0.029709 | 0.282251 | .. |
| std | 41.390065 | 91.491548 | 43.521413 | 0.004848 | 0.000035 | 0.002968 | 0.002759 | 0.008903 | 0.018857 | 0.194877 | .. |
| min | 88.333000 | 102.145000 | 65.476000 | 0.001680 | 0.000007 | 0.000680 | 0.000920 | 0.002040 | 0.009540 | 0.085000 | .. |
| 25% | 117.572000 | 134.862500 | 84.291000 | 0.003460 | 0.000020 | 0.001660 | 0.001860 | 0.004985 | 0.016505 | 0.148500 | .. |
| 50% | 148.790000 | 175.829000 | 104.315000 | 0.004940 | 0.000030 | 0.002500 | 0.002690 | 0.007490 | 0.022970 | 0.221000 | .. |
| 75% | 182.769000 | 224.205500 | 140.018500 | 0.007365 | 0.000060 | 0.003835 | 0.003955 | 0.011505 | 0.037885 | 0.350000 | .. |
| max | 260.105000 | 592.030000 | 239.170000 | 0.033160 | 0.000260 | 0.021440 | 0.019580 | 0.064330 | 0.119080 | 1.302000 | .. |

8 rows × 23 columns

```
parkinsons_data['status'].value_counts()
```

```
1    147
0     48
```

**Fig 8.2**: Code snippets on google colab 2



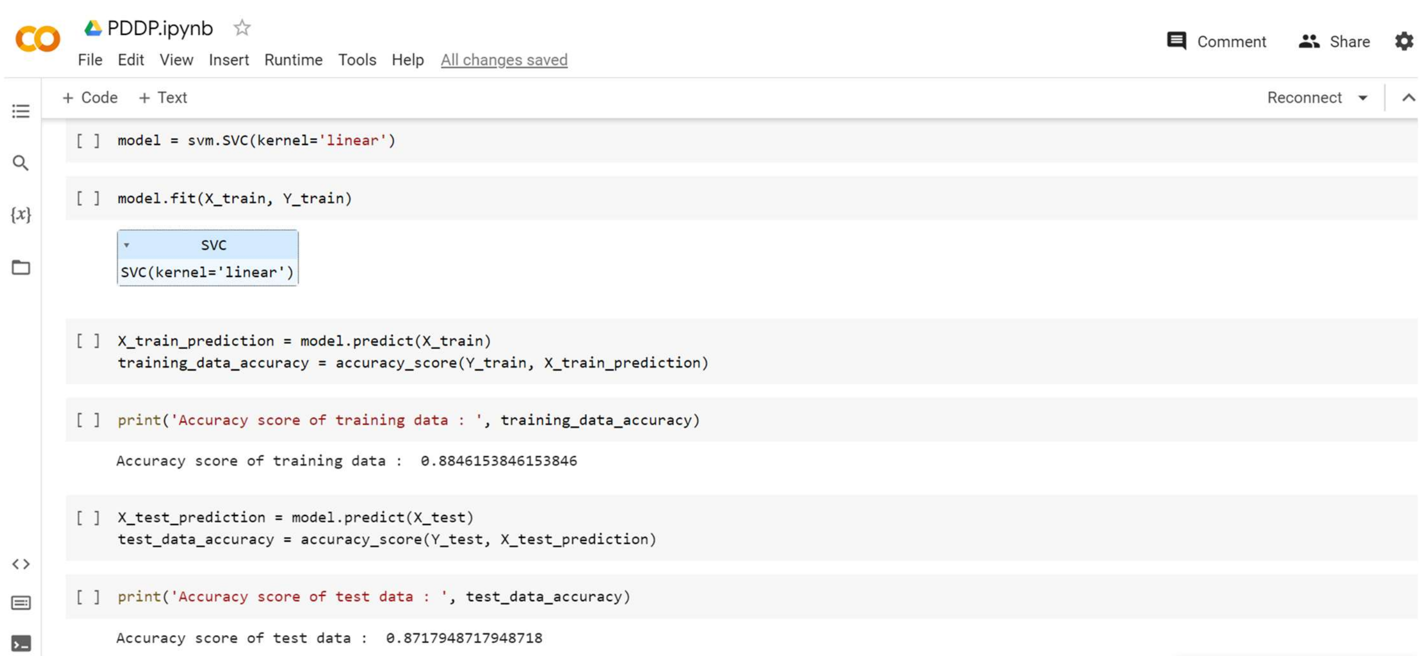
```
X_train = scaler.transform(X_train)

X_test = scaler.transform(X_test)
```

```
print(X_train)
```

```
[[ 0.63239631 -0.02731081 -0.87985049 ... -0.97586547 -0.55160318
   0.07769494]
 [-1.05512719 -0.83337041 -0.9284778  ...  0.3981808  -0.61014073
   0.39291782]
 [ 0.02996187 -0.29531068 -1.12211107 ... -0.43937044 -0.62849605
  -0.50948408]
 ...
 [-0.9096785  -0.6637302  -0.160638   ...  1.22001022 -0.47404629
  -0.2159482 ]
 [-0.35977689  0.19731822 -0.79063679 ... -0.17896029 -0.47272835
   0.28181221]
 [ 1.01957066  0.19922317 -0.61914972 ... -0.716232    1.23632066
  -0.05829386]]
```

```
model = svm.SVC(kernel='linear')
```

**Fig 8.3**: Code snippets on google colab 3



```
model = svm.SVC(kernel='linear')
```

```
model.fit(X_train, Y_train)
```

```
SVC
SVC(kernel='linear')
```

```
X_train_prediction = model.predict(X_train)
training_data_accuracy = accuracy_score(Y_train, X_train_prediction)
```

```
print('Accuracy score of training data : ', training_data_accuracy)
```

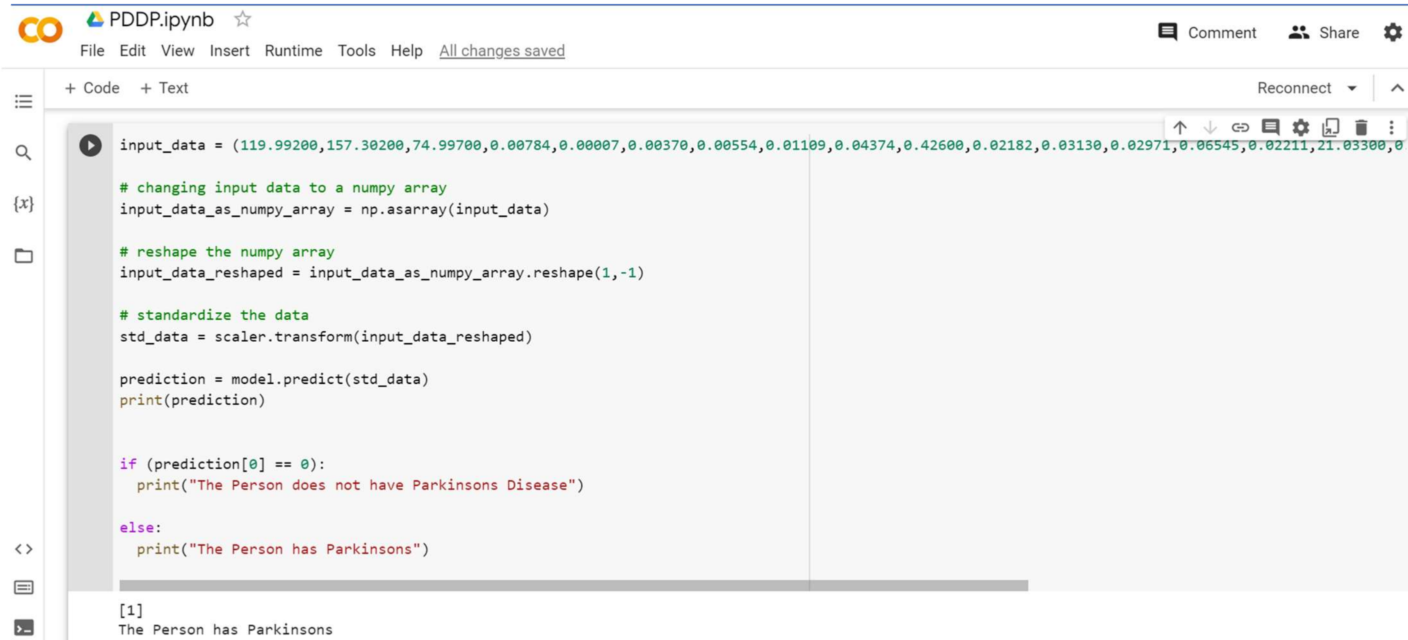
```
Accuracy score of training data :  0.8846153846153846
```

```
X_test_prediction = model.predict(X_test)
test_data_accuracy = accuracy_score(Y_test, X_test_prediction)
```

```
print('Accuracy score of test data : ', test_data_accuracy)
```

```
Accuracy score of test data :  0.8717948717948718
```

**Fig 8.4**: Code snippets on google colab 4




```
input_data = (119.99200,157.30200,74.99700,0.00784,0.00007,0.00370,0.00554,0.01109,0.04374,0.42600,0.02182,0.03130,0.02971,0.06545,0.02211,21.03300,0

# changing input data to a numpy array
input_data_as_numpy_array = np.asarray(input_data)

# reshape the numpy array
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

# standardize the data
std_data = scaler.transform(input_data_reshaped)

prediction = model.predict(std_data)
print(prediction)


if (prediction[0] == 0):
  print("The Person does not have Parkinsons Disease")

else:
  print("The Person has Parkinsons")

[1]
The Person has Parkinsons
```

**Fig 8.5:** Result

# CHAPTER 9

# APPLICATIONS

1. Monitoring disease progression: Machine learning algorithms can be used to monitor disease progression in patients with Parkinson's disease, allowing for early intervention and adjustment of treatment plans.

2. Drug development: Machine learning algorithms can be used to identify potential drug targets and develop new drugs for the treatment of Parkinson's disease.

3. Clinical trials: Machine learning algorithms can be used to identify potential participants for clinical trials and to analyze the results of clinical trials, improving the accuracy and reliability of clinical trial data.

4. Health Insurance Planning: ML predictions can assist insurance companies in assessing the risk of insuring individuals with Parkinson’s, leading to better pricing and coverage decisions.

5. Telemedicine: ML-based diabetes prediction can be integrated into telemedicine platforms, enabling remote consultations and providing healthcare professionals with valuable insights for remote patient management.

6. Remote patient monitoring: ML algorithms can be used to monitor patients with Parkinson's disease remotely, using wearable sensors to collect data on their symptoms and disease progression. This can reduce healthcare costs by reducing the need for frequent in-person appointments.

## CHAPTER 10

# FUTURE SCOPE

The purpose of this system is to predict the presence of Parkinson’s disease accurately in a person using the voice of the concerned person at a very low cost. It also allows users to efficiently predict the output for the given input. This also greatly reduces the possibility of doubt and potential risk of wrong prediction. This is a user-friendly project that would be suitable for all users.

# CONCLUSION

The purpose of this system is to predict the presence of Parkinson's disease accurately in a person using the voice of the concerned person at a very low cost. It also allows users to efficiently predict the output for the given input. This also greatly reduces the possibility of doubt and potential risk of wrong prediction. This is a user-friendly project that would be suitable for all users.

# REFERENCES

[1] http://dspace.srmist.edu.in/jspui/bitstream/123456789/44901/1/P11907.pdf
[2] https://www.sciencedirect.com/science/article/abs/pii/S0957417422012507
[3] https://ijcrt.org/papers/IJCRT2203530.pdf