**Innovative machine learning approach to predict onset for Parkinson's Disease**

Submitted in partial fulfilment of the requirements of the degree of

**Bachelor of Engineering (B.E.)**

in

**COMPUTER ENGINEERING**

by

**Zeal Bipin Tailor EU1202095**

**Priyadarshani Tapas Sarangi EU2202006**

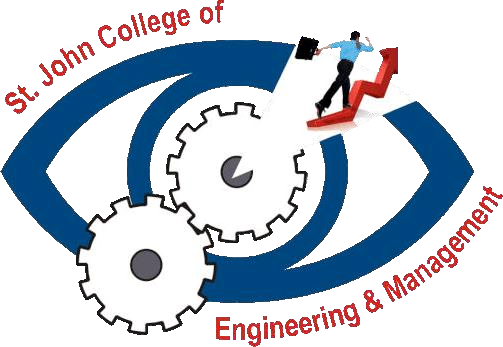
**Krutisha Rajesh Shirke EU2202005**

**Alroy Louis lobo EU1202113**

Under the guidance of

## Ms. Deepika Nadar

Assistant Professor



**Department of Computer Engineering**

**St. John College of Engineering and Management, Palghar**

**University of Mumbai**

2022–2023

**CERTIFICATE**

This is to certify that the B.E. project entitled  **“ Innovative machine learning approach to predict onset for Parkinson's”**  is a bonafide work of  **“ Zeal Bipin Tailor ”**

**(EU1202095)** ,  **“Priyadarshani Tapas Sarangi” (EU2202006)** ,  **“ Krutisha Rajesh Shirke” (EU2202005) “Alroy Louis Lobo” (EU1202113)**  submitted to University of Mumbai in partial fulfilment of the requirement for the award of the degree of  **“Bachelor of Engineering”**  in  **“Computer Engineering”**  during the academic year 2023–2024.

**Ms. Deepika Nadar**

Project Guide

**Dr. Nilesh T. Deotale Dr. G. V. Mulgund**

Head of Department sPrincipal

**B.E. Project Report Approval**

This project report entitled  ***Innovative machine learning approach to predict on set for Parkinson's Disease*** by  ***Zeal Bipin Tailor, Priyadarshani Tapas Sarangi,***  and  ***Krutisha Rajesh Shirke Alroy Louis Lobo***  is approved for the degree of  ***Bachelor of Engineering***  in  ***Computer Engineering***  from  ***University of Mumbai***  .

## Examiners

1. .
2. .

Date :

Place :

**Declaration**

We declare that this written submission represents our ideas in our own words and where others’ ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Signature

Zeal Bipin Tailor (EU1202095)

Signature

Priyadarshani Tapas Sarangi (EU2202006)

Signature

Krutisha Rajesh Shirke (EU2202005)

Signature

Alroy Louis Lobo(EU1202113)

Date:

Place:

## Abstract

Various medical databases contribute extensive data to identify illnesses, with a focus on Parkinson's disease, a prevalent neurological condition marked by movement impairment, reduced lifespan, and disability. Unfortunately, there is no cure, and its onset involves decreased dopamine synthesis due to brain cell degeneration. Dopamine is vital for communication among brain regions that regulate movement, resulting in a range of movement-related and non-motor symptoms. Recent research indicates that non-motor symptoms manifest years before motor symptoms, highlighting the imperative for early and precise diagnosis to effectively impede disease progression. Speech difficulties are present in approximately 90% of Parkinson's cases. In practical applications, machine learning (ML) is harnessed to process information and predict diseases at an early stage. Several machine learning approaches, including K-Nearest Neighbors , Naïve Bayes, Logistic Regression, and a hybrid methodology, are used in this work to predict the risk of Parkinson's disease based on user-provided input using a pre-established dataset.Emphasizing the crucial role of early-stage identification in patient recovery, this research underscores the significance of predictions facilitated by hybrid ML approaches in enhancing diagnostic precision and intervention strategies.

***Keywords—*** Parkison’s disease [PD] ;Motor Disorders ;Machine Learning

**Table of Contents**

Abstract v

List of Figures vii

List of Tables viii

List of Abbreviations 1

Chapter 1 Introduction 2

1.1 Motivation 4

1.2 Problem Statement 4

1.3 Objectives 4

1.4 Scope 4

Chapter 2 Review of Literature 5

2.1 Parkinson's Disease Prediction Using Machine

Learning Algorithm 5

2.2. Early Detection of Parkinson's Disease using Machine

Learning & Image Processing 5

2.3 A Deep Learning Based Method for Parkinson’s Disease

Detection Using Dynamic Features of Speech 5

2.4 Detection of Parkinson’s Disease by Employing

Boosting Algorithm 5

2.5 Application of Adaptive Back-Propagation Neural

Networks for Parkinson’s Disease Prediction 6

2.6 Early Detection of Parkinson’s Disease Using Deep

Learning and Machine Learning 6

**Chapter 1 Introduction 1**

1.1 Motivation 2

1.2 Problem Statement 3

1.3 Objectives 3

1.4 Scope 3

**Chapter 2 Review of Literature 5**

2.1 A review of the effect of traffic and weather

5

characteristics on road safety

2.2 . Accident analysis on national highway-3

5

between Indore to Dhamnod

2.3 Personal and behavioral predictors of automobile

5

crash and injury severity

2.4 Analysis and prediction of traffic fatalities

resulting from angle collisions including the effect of 6

vehicles

2.5 The independent contribution of driver, crash,

6

and vehicle characteristics to driver fatalities

**Chapter 3 Requirement Gathering and Planning** 7

3.1 Requirement Elicitation 7

3.2 Software Requirements 7

3.3 Functional Requirements 7

3.4 Non-Functional Requirements 7

**Chapter 4 Design**  8

4.1 DFD Diagram for Road Accident Analysis and

Hotspot Prediction using Clustering 8

|  |  |  |
| --- | --- | --- |
|  | 4.2 UML Diagrams | 9 |
|  | 4.2.1 Use Case Diagram | 9 |
|  | 4.2.2 Class Diagram | 10 |
|  | 4.2.3 Activity Diagram | 11 |
|  | 4.2.4 Sequence Diagram | 12 |
| **Chapter 5** | **Report on Present Investigation** | **14** |
|  | 5.1 Methodology/Proposed System | 14 |
|  | 5.1.1 Architecture/Block diagram of the System | 14 |
|  | 5.2 Implementation | 15 |
|  | 5.2.1 Algorithm/Flowchart | 15 |
|  | 5.2.2 Dataset | 17 |
|  | 5.2.3 GUI | 18 |
| **Chapter 6** | **Results and Discussion** | **20** |
| **Chapter 7** | **Conclusion** | **23** |
|  | **References** | **24** |
| **Appendix** | **Technologies Used** | **25** |
|  | **Publication** | **26** |
|  | **Acknowledgments** | **27** |

# List of Figures

**Figure Figure Name Page No.**

**No.**

4.1.1 Level 0 DFD for Road Accident Analysis and Hotspot

8

Prediction using Clustering

4.1.2 Level 1 DFD for Road Accident Analysis and Hotspot

8

Prediction using Clustering

4.1.3 Level 2 DFD for Road Accident Analysis and Hotspot

9

Prediction using Clustering

4.1 Use Case Diagram for Road Accident Analysis and Hotspot

9

Prediction using Clustering

4.2 Class Diagram for Road Accident Analysis and Hotspot

1. Prediction using Clustering

4.3 Activity Diagram for Road Accident Analysis and Hotspot

1. Prediction using Clustering

4.4 Sequence Diagram for Road Accident Analysis and Hotspot

1. Prediction using Clustering

4.5 Timeline/Gantt Chart for Road Accident Analysis and

1. Hotspot Prediction using Clustering

4.6 Work Breakdown Structure (WBS) Chart for Road Accident

1. Analysis and Hotspot Prediction using Clustering
   1. Architecture /Block diagram of the system 14
   2. GUI for Connecting our Model to Google Maps 18
   3. GUI for Select your Starting Point and Destination 18
   4. GUI for Model showing all Accident Prone areas and main

19 cause live

# List of Tables

**Table No. Table Name Page No.**

3.1 Dataset 18

3.2 Dataset of Accident Location 18

# List of Abbreviations

|  |  |
| --- | --- |
| **SJCEM** | St. John College of Engineering and Management |
| **CAS** | Collision Avoidance System |
| **WBS** | Work Breakdown System |
| **DFD** | Data Flow Diagram |
| **GUI** | Graphical User Interface |

**Chapter 1**

## Introduction

Parkinson's disease is a steadily progressing neurological disorder that affects the central nervous system and causes tremors linked with movement, difficulty speaking, and difficulty walking. For the best possible intervention, prompt detection is crucial. The disease manifests as a confluence of non-motor symptoms including mood swings, cognitive difficulties, pain, sensory dysfunction, and dysautonomia with motor symptoms like tremors, reduced movement speed, stiffness, and postural instability.[1].Since the symptoms of this disease and other neurological illnesses can be mistaken for one another, around 25% of PD patients are misdiagnosed. The main aim of the suggested method is to create a machine learning-based automated diagnostic tool that can precisely determine if Parkinson's disease is present and whether healthy controls are in good health[2].Artificial intelligence and machine learning are becoming more and more popular because they can automate accurate pattern recognition. About 90% of those with this disease exhibit motor function impairments[3].

It is a degenerative sickness that affects both motor and non-motor abilities. It is the second most common neurological ailment. Common movement disorders include tremors, challenges initiating or stopping motions, balance issues, and stiffness. A significant proportion of PD individuals experience cognitive deficits impacting memory, language, and processing skills during the chronic course of the illness.

Parkinson's disease (PD) symptoms emerge when the brain produces less than 60% of the neurotransmitter dopamine, crucial for transmitting movement-related nerve signals. Before basic motor signs appear, a prodromal period, lasting five to twenty years, includes non-specific symptoms affecting quality of life such as sleep abnormalities, loss of smell, and issues with rapid eye movement (REM)[4].

In this context, a hybrid approach incorporates a variety of machine learning techniques, including K-Nearest Neighbors (KNN), Naïve Bayes, and Logistic Regression .This amalgamation aims to enhance the precision and efficacy in forecasting Parkinson's disease based on user input.

Parkinson's disease (PD) is estimated to impact 1-2 out of every 1,000 seniors, with an annual incidence of 15 per 100,000. Onset typically occurs between ages 65 and 70, affecting around 10 million people worldwide, more commonly in males. As there is currently no cure, early symptomatic treatment is recommended, relying on an early and accurate diagnosis. Achieving this is challenging due to the unique characteristics of each person's illness, resulting in clinical detection rates ranging from 65% to 93%, with an error rate of approximately 25%. Despite efforts, a reliable biomarker for laboratory diagnosis remains elusive.

In medicine, machine learning algorithms show promise for faster and more precise prognoses in various chronic illnesses. Significant research advancements in this field offer new opportunities for rapid and accurate diagnosis.

* 1. **Motivation**

The motivation behind developing early-stage Parkinson's disease detection models using machine learning, particularly through the utilization of K-Nearest Neighbors (KNN), Naive Bayes, and Logistic Regression, is rooted in the pressing need to improve the lives of those affected by this debilitating condition.

1. Early Intervention: Parkinson's disease is a progressive disorder, and early intervention is crucial to slow its progression and manage its symptoms effectively. By detecting the disease in its early stages, we can initiate treatment and support mechanisms sooner, significantly enhancing the quality of life for patients.
2. Challenges in Clinical Diagnosis: Parkinson's disease diagnosis is often based on clinical assessments, which can be subjective and prone to error. Machine learning offers the potential for more objective and accurate diagnostic tools, reducing misdiagnoses and providing clinicians with valuable support.
3. Customized Treatment: Identifying the disease at an early stage allows for the tailoring of treatment plans to the specific needs of each patient. Machine learning models can help predict disease progression and recommend personalized treatment strategies, optimizing therapeutic outcomes.
4. Large-Scale Screening: Machine learning algorithms can be applied to large datasets, enabling efficient and cost-effective screening of at-risk populations. This could lead to the identification of cases that might otherwise go undetected until advanced stages of the disease.
5. Research Advancements: By developing machine learning models for Parkinson's disease detection, we contribute to the growing body of research and knowledge in this field. This research can lead to a deeper understanding of the disease's underlying mechanisms and potentially uncover novel insights into its early indicators.
6. Accessible Healthcare: Machine learning models can be integrated into telemedicine and wearable technology, making early-stage Parkinson's disease detection more accessible to a broader population, especially in remote or underserved areas.

### 1.2 Problem Statement

Parkinson’s disease is difficult to diagnosis at the early stage so it becomes difficult to take preventive measures at the early stage which can lead to danger to the life. Many people are unaware of this condition, and through our efforts we hope to empower individuals with knowledge and resources for early detection and management. By using machine learning approach Parkinson’ is diagnostic at the early stage enabling individuals to take proactive measures to safeguard their health. Our project provide information on it’s early-stage symptoms, educational materials and preventive measures.

### 1.3 Objectives

The objectives of the project are as follows:

* To build a machine learning model to predict the likelihood of Parkinson's disease in individuals based on their medical and demographic data.
* To provide materials that highlight common signs and symptoms for the public to be aware of.
* To help the patient to known about the disease at the early stage so that they get necessary health support at the earliest.
* To build a user-friendly interface for healthcare professionals and patients to interact with the system.

### 1.4 Scope

The system is build to identify the Parkinson disease at the early stage using machine learning. for real – time monitoring of individuals at risk, enabling early intervention and personalized healthcare.

Using machine learning algorithms K- nearest neighbors(KNN), Naïve Bayes and Logistic Regression with high accuracy prediction.

Considering the seriousness of the disease our system will provide educational material and suggest exercises and diet including physical therapy. Which will help the patients to improve the overall quality of life for people with Parkinson.

**Chapter 2**

## Review of Literature

**2.1 A Deep Learning Based Method for Parkinson’s Disease Detection Using Dynamic Features of Speech**

This assessment examines current developments in voice analysis-based deep learning applications for Parkinson's disease identification. To identify early indicators of Parkinson's disease, the focus is on extracting variables linked to articulation, phonation, and prosody, such as the Vowel Articulation Index and Formant Centralization Ratio. Artificial Neural Networks and Deep Neural Networks are two deep learning approaches used in the classification problem. Convolutional Neural Nets emphasize invariance by capturing voice-to-unvoiced segment transitions. The long-distance contextual information included in dynamic speech characteristics is addressed by Recurrent Neural Networks , specifically Bidirectional Long Short-Term Memory , which improves the accuracy of PD identification.

**2.2. Early Detection of Parkinson's Disease using Machine Learning & Image Processing**

This research tackles the difficulty of precise Parkinson's Disease (PD) prediction, a challenge exacerbated by frequent overlaps with other neurological symptoms, leading to a misdiagnosis rate exceeding 25%. The system in question utilizes machine learning techniques in conjunction with weighted Magnetic Resonance Imaging (MRI) data obtained from the Parkinson's Progression Markers Initiative. By employing image registration for midbrain slice alignment and a hybrid SVM-Random Forest approach, the study achieves enhanced accuracy and specificity, offering a promising avenue for improved PD diagnosis.

#### 2.3. Parkinson's Disease Prediction Using Machine Learning Algorithm

This section describes how Parkinson's disease affects the substantia nigra pars compacta's capacity to produce dopamine, which leads to motor symptoms such bradykinesia, stiffness, rest tremor, and postural instability. It also exhibits non-motor symptoms, such as cognitive impairment. It underscores the pivotal role of machine learning (ML) in computer science, emphasizing its application in automating tasks and learning from input data. The passage briefly mentions image normalization techniques and various computer vision applications, showcasing ML's broad range of applications beyond PD research.

#### 2.4. Deep learning-based Parkinson’s disease classification using vocal feature sets

This work uses deep learning to analyze premotor variables, including olfactory loss, imaging indicators, and rapid eye movement, with the particular goal of identifying Parkinson's disease early. The deep learning model performs better than twelve other approaches, with an average accuracy of 96.45% on a dataset including 401 patients with early-stage of this disease and 183 healthy persons. Additionally, the model reveals feature importance through Boosting, providing valuable insights for enhanced detection. The ultimate goal is to facilitate early intervention and access to disease-modifying therapy, thereby improving patient outcomes.

#### 2.5. Early detection of Parkinson's disease using machine learning

This study investigates the early detection of Parkinson's Disease (PD) by employing ensemble machine learning, with a particular focus on boosting algorithms. Assessing four algorithms on the UCI PD dataset, the emphasis is on accurate and early diagnosis. Results show satisfactory performance across various parameters. Following hyperparameter tuning, Light GBM achieves the highest accuracy at 93.39%, while XGBoost and Gradient Boosting surpass 90%. Adaboost, with tuning, reaches a maximum accuracy of 87.22%. This analysis underscores the potential of boosting algorithms for effective early PD detection.

**2.6. Detection of Parkinson’s Disease by Employing Boosting Algorithm**

This review emphasizes the underexplored realm of audio impairment in early Parkinson's Disease (PD) detection compared to other data types. It highlights the efficacy of an enhanced SVM model for genetic data classification. The study contrasts deep learning approaches, including ensemble models on phonation data and traditional decision trees on speech features. Proposing KNN, logistic regression, random forest regression, and SVM for PD classification on audio data, preliminary findings favor the K nearest neighbor model, achieving 91.83% accuracy and 0.95 sensitivity. These results contribute to advancing telemedicine-driven PD detection.

#### 2.7. Machine Learning for the Diagnosis of Parkinson’s Disease

This survey encompasses various Parkinson's Disease (PD) detection methods, incorporating voice signals, speech analysis, gait, and multimodal datasets. Utilizing SVM, RF, CNN, and CNN-BLSTM, studies explore deep learning techniques like ResNet and AE for high-accuracy early PD detection. Addressing a gap in existing methods, the survey introduces hyperparameter-optimized Adaptive Bi-Directional Neural Networks (ABPNN) for swift and accurate PD prediction.

#### 2.8. Classification of Parkinson Disease Based on Patient’s Voice Signal Using Machine Learning

#### Since its first description in 1817, this disease has been well recognized as a common neurological condition that primarily affects the elderly. Recent studies show that 89% of PD patients have abnormal speech patterns that are linked to a drop in dopamine levels as a result of substantia nigra cell injury. Six different machine learning algorithms are used in this work to identify this disease based on human speech signals: Random Forest, K-Nearest Neighbor, Extreme Gradient Boosting, Logistic Regression, and Decision Tree. Extracting crucial characteristics is a necessary step in reducing the complexity of the dataset.Notable accuracy values are found by analyzing the voice intensity and spectrum: 91% for SGD-Classifier, 95% for XGB-Classifier, 91% for Logistic Regression, 97% for Random Forest, 95% for KNN, and 95% for Decision Tree. Interestingly, Random Forest has the best accuracy, indicating that it might be a useful diagnostic tool. The study emphasizes how important it is to investigate more traits in PD patients in order to improve its relevance in the medical field.

**2.9. Parkinson's Disease Detection through Vocal Biomarkers and Advanced Machine Learning Algorithm**

#### The literature survey highlights various strategies for Parkinson's disease (PD) detection. Wang et al. (2020) excel in early detection, achieving 96.45% accuracy with a deep-learning method for premotor PD stages. Prashanth et al. (2018) use patient questionnaires, achieving over 95% accuracy in distinguishing early PD from healthy individuals. Braga et al. (2019) employ Random Forest and SVM on spontaneous speech with notable precision. Zhao et al. (2018) introduce a gait-centric machine learning approach, outperforming existing methods in assessing PD severity. Yang et al. (2022) develop PD-ResNet, a ResNet-based model showing high accuracy and efficacy in early PD classification. Rocha et al. (2014) explore RGB-D cameras for gait analysis, revealing promising discriminatory capabilities. These studies collectively represent diverse and impactful advancements in PD detection.

**Chapter 3**

## Requirements Analysis

### 3.1 Hardware Requirements

**Operating System:**  Windows 10 Home Single Language

**Processor:**  Intel(R) Core(™) i5-7200U CPU @ 2.50GHz 2.70 GHz

**Hard Disk Storage:** 1 TB

**Installed RAM:** 8.00 GB (7.87 GB usable)

### 3.2 Software Requirements

1. . Jupyter Notebook IDE
2. . Different libraries
3. . 64- bit Windows OS / macOS
4. . Anaconda
5. . Dataset: The dataset in the form of (.csv) or (.xml)
6. . Python

### 3. 3 Functional Requirements

1. . Collect dataset of essay from various sources [8].
2. . Extracted data stored in .csv file format.
3. . Process data using Python libraries.

### 3 . 4 Non-Functional Requirements

Python workspace should be always updated with all the libraries that are required for running the process. Appropriate dataset should be given as an input.

**Chapter 4**

## Design

### 4.1 Dataflow Diagrams (DFDs)

#### 4.1.1 Level 0 DFD



Figure 4.1: Level 0 of Data Flow Diagram of System

A data flow diagram (DFD) maps out the flow of information for any process or system. It uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination.

#### 4.1.2 Level 1 DFD

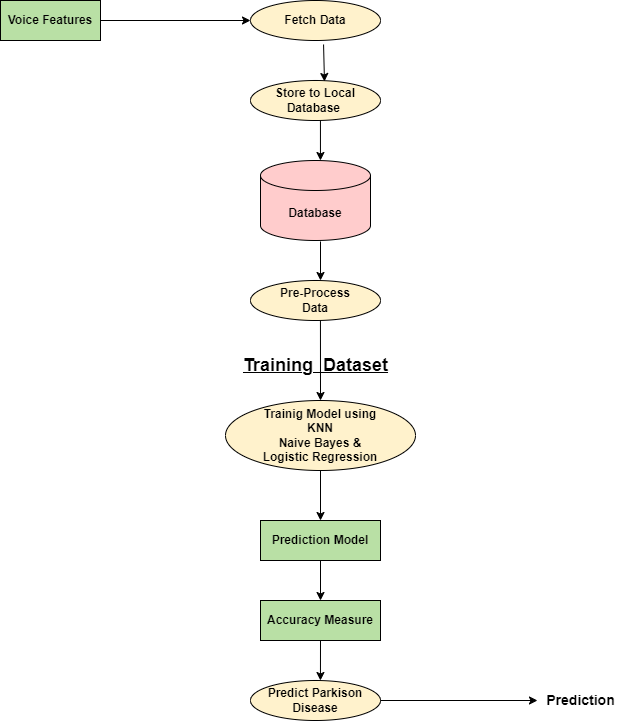


Figure 4.2: Level 1 of Data Flow Diagram of System

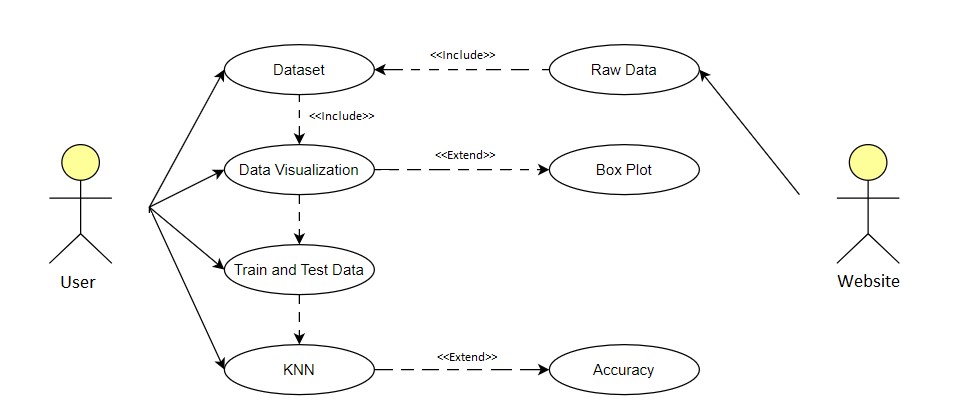
A data flow diagram (DFD) maps out the flow of information for any process or system. It uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination.

### 

### 4.2 UML Diagrams

#### 4.2.1 Use case Diagram

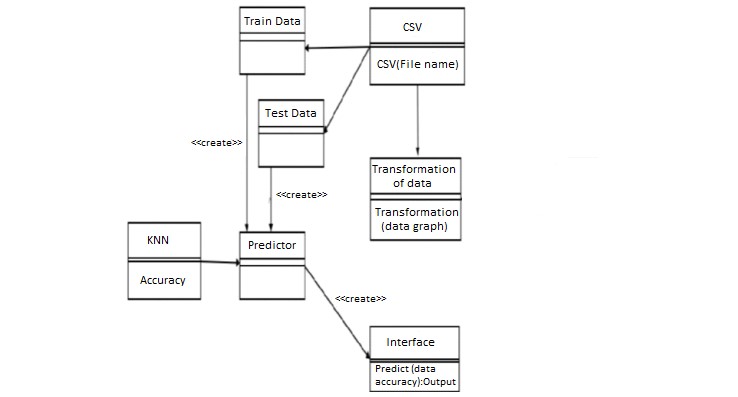
Figure 4.4: Use case diagram for System



Use-case diagrams describe the high-level functions and scope of a system. These diagrams also identify the interactions between the system and its actors. The use cases and actors in use-case diagrams describe what the system does and how the actors use it, but not how the system operates internally.

### Class Diagram

Figure 4.5: Class Diagram of System



Class diagrams are the blueprints of your system or subsystem. You can use class diagrams to model the objects that make up the system, to display the relationships between the objects, and to describe what those objects do and the services that they provide. Class diagrams are useful in many stages of system design.

### Activity Diagram

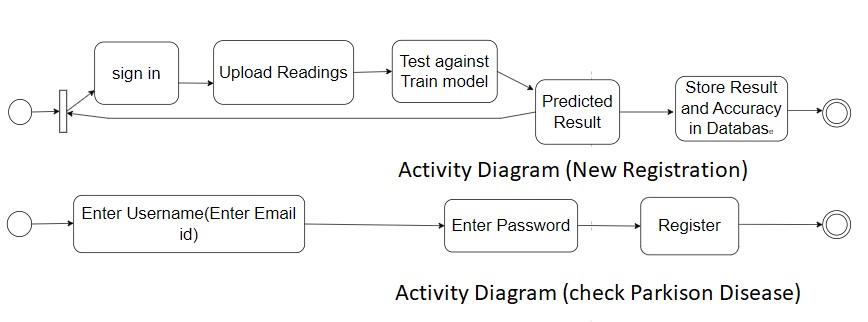


Figure 4.6: Activity diagram for Parkison Disease Prediction System

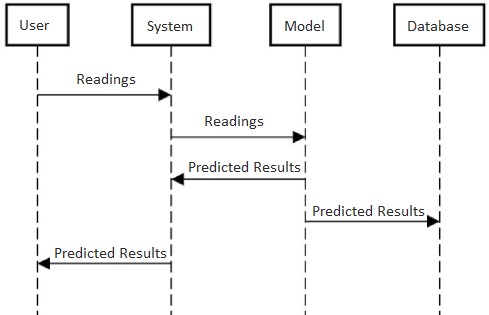
An activity diagram is a type of Unified Modelling Language (UML) flowchart that shows the flow from one activity to another in a system or process. It's used to describe the different dynamic aspects of a system and is referred to as a 'behaviour diagram' because it describes what should happen in the modelled system.

### 

### Sequence Diagram

### 

Figure 3.1.5: Sequence Diagram of System



#### 

#### Timeline/Gantt Chart

Figure 4.8: Timeline for Road Accident Analysis and Hotspot Prediction using Clustering

#### 4.2.6 Work Breakdown Structure (WBS) Chart

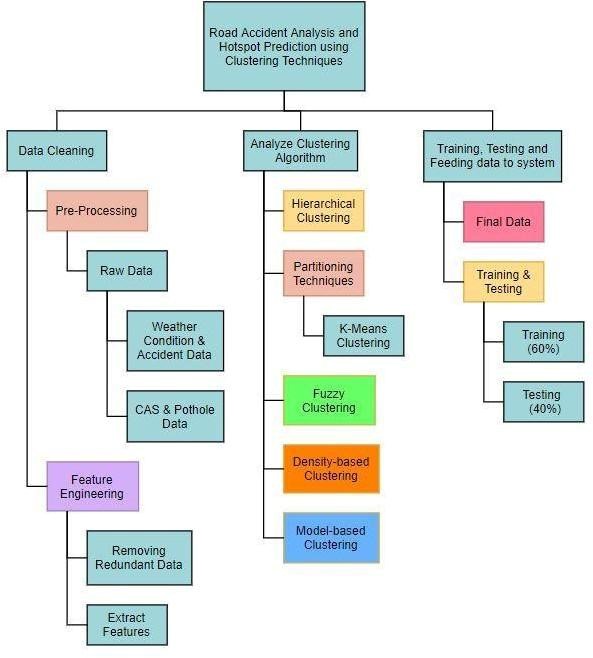


Figure 4.9: WBS Chart for Road Accident Analysis and Hotspot Prediction using Clustering

**Chapter 5**

**Report on Present Investigation**

#### 5.1 Methodology/Proposed System

In Figure , the delineation illustrates the systematic gathering of clinical, demographic, and medical data from diverse sources, encompassing individuals both afflicted with and without Parkinson's disease. The gathered

data undergoes cleaning and preprocessing, utilizing features such as MDVP, NHR, HNR, Status, RPDE, D2, and PPE for analysis. After that, the dataset is split into subsets for training and testing, and the K-Nearest Neighbors,

##### 5.1.1 Architecture/Block diagram of the System

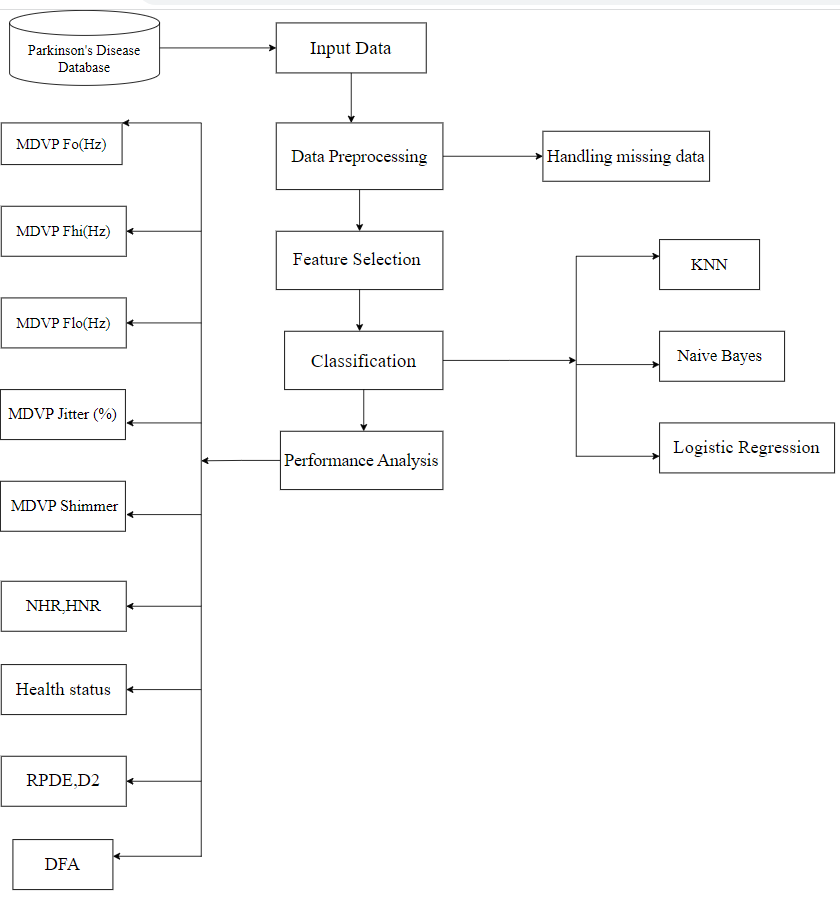


Figure 5.1: Architecture/Block Diagram for Parkinsion’s Disease Detection

#### 5.2 Implementation

**Input Data:** Raw data or observations that are sent to a system or algorithm for processing, analysis, or learning are referred to as input data. It is the starting set of facts that a system or model use to carry out a certain function.

**Data Preprocessing:** Preparing raw data for analysis or modeling include cleaning and formatting it. This might entail managing missing values, eliminating anomalies, scaling or normalizing data, and transforming categorical variables into numerical representations.

**Feature Selection:** Selecting a subset of pertinent characteristics (variables, attributes) from a broader collection of features is known as feature selection. It expedites the training process, lowers overfitting, and enhances model performance.

**Classification :** The objective of classification is to assign predefined groups or labels to incoming data, representing a form of supervised learning. For the algorithm to make predictions or decisions about new, unseen data, it must initially learn from a set of labeled instances.

**Performance Analysis:** Performance analysis is the process of assessing a model's correctness and efficacy. The evaluation metrics employed typically depend on the nature of the problem, whether it involves binary or multiclass classification. Indicators including accuracy, precision, recall, F1 score, and the area under the ROC curve (AUC-ROC) are frequently included in these measurements.

**Managing Missing Data:** Managing partial or absent values in a dataset entails developing methods. Typical ways include filling in missing data using more sophisticated techniques like predictive modeling, or eliminating missing values and imputing values using statistical methods.

**KNN (K-Nearest Neighbors):** It is a flexible supervised machine learning method that can handle problems with both regression and classification.. Its operation involves identifying the 'k' nearest data points in the feature space relative to a given data point, and subsequently predicting the majority class (in classification tasks) or the average (in regression tasks) of those proximate neighbors.

**Naive Bayes :** Operating on the principles of probabilistic classification and Bayes' theorem, the Naive Bayes algorithm hinges on the assumption that, given the class label, the descriptors used to portray instances are conditionally independent. Despite its straightforwardness, Naive Bayes demonstrates frequent success, particularly in tasks such as text categorization and other diverse problem domains. Statistically, logistic regression is used to solve binary and multiclass classification issues. It does this by estimating the likelihood that an instance will belong to a given class using a logistic function. Regression issues do not employ logistic regression; instead, classification problems do.

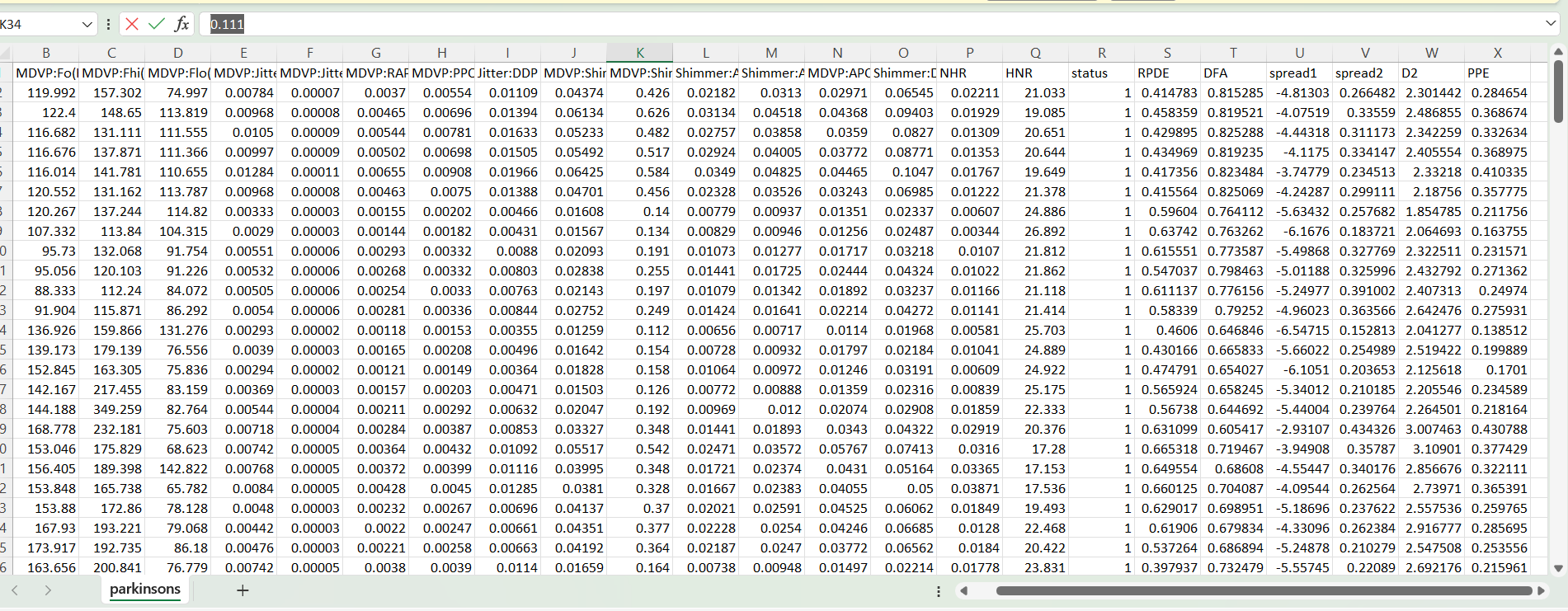
These are basic words in data analysis and machine learning, and they are important for creating and assessing models

##### 5.2.1 Algorithm/Flowchart

1. Load the dataset.
2. Initialize the value of k to a positive integer.
3. To find the class variable, iterate from starting to ending of all training data points.
4. Calculate the Euclidean distance between test data point and each data point in the training data. Here the author used Euclidean distance as the distance metric since it is the most popular method. The other metrics that c cosine, etc.
5. After that sort the resultant distances in ascending order.
6. Return the first k rows from the sorted list.
7. Find the most frequent class occurred from these rows.
8. Return the predicted class.

##### 5.2.2 Dataset

Table 3.1: Dataset of Parkinsion’s diease



## Chapter 6

### Results and Discussion

* After applying the above-mentioned model, the expected output is obtained. The snapshots attached below gives a clear idea how this model is implemented.



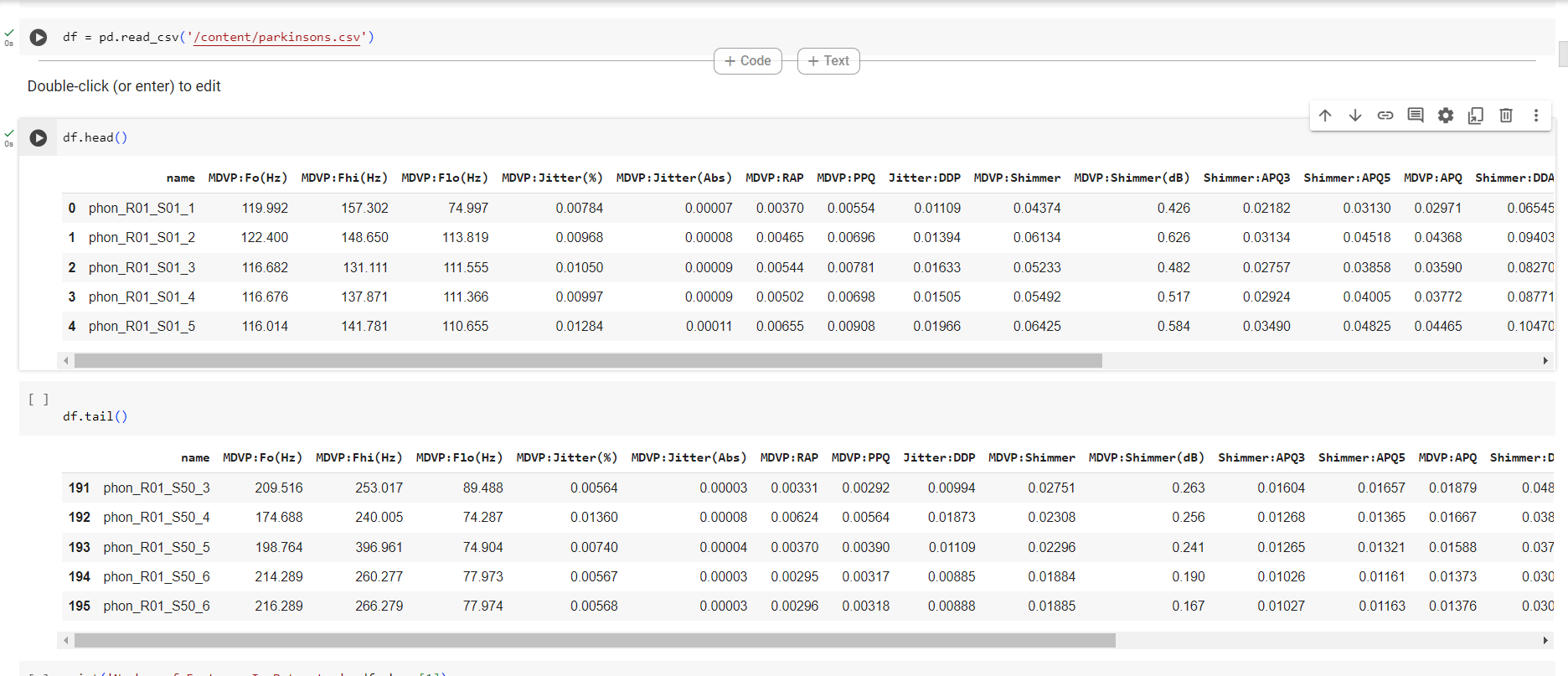


Figure 6.1: Importing the dataset • Data is sorted to obtain accurate results.

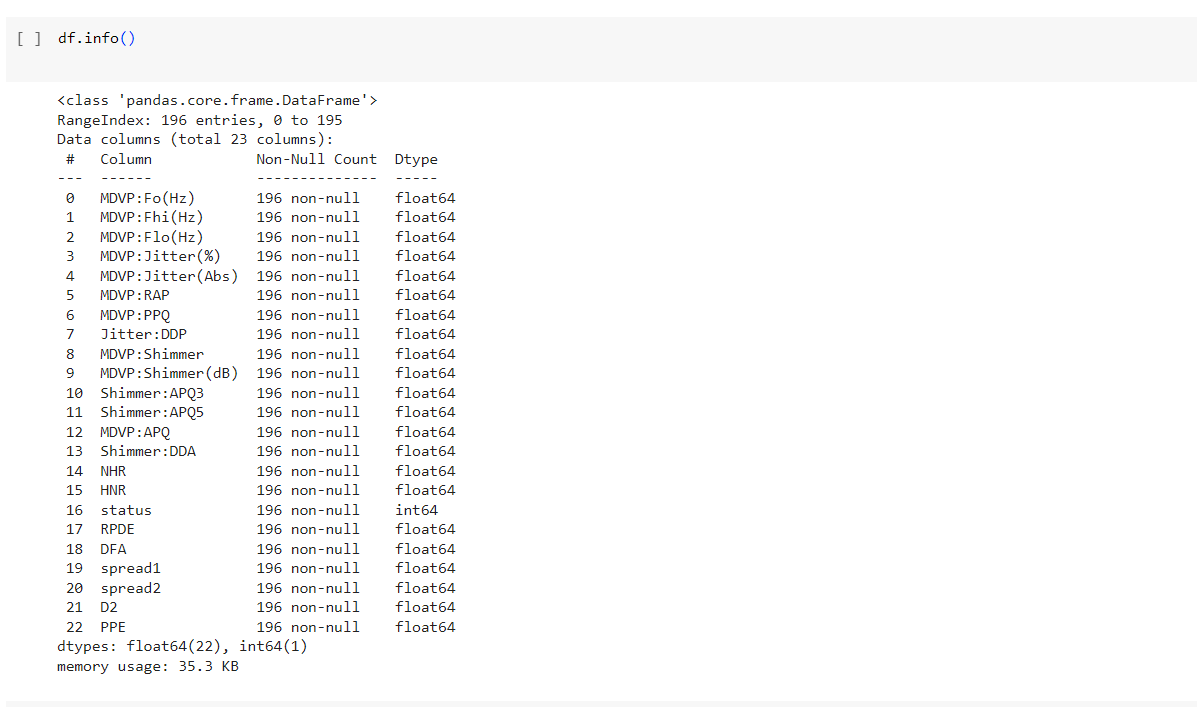


Figure 6.2: Sorting the dataset

* Countplot of Balanced Data

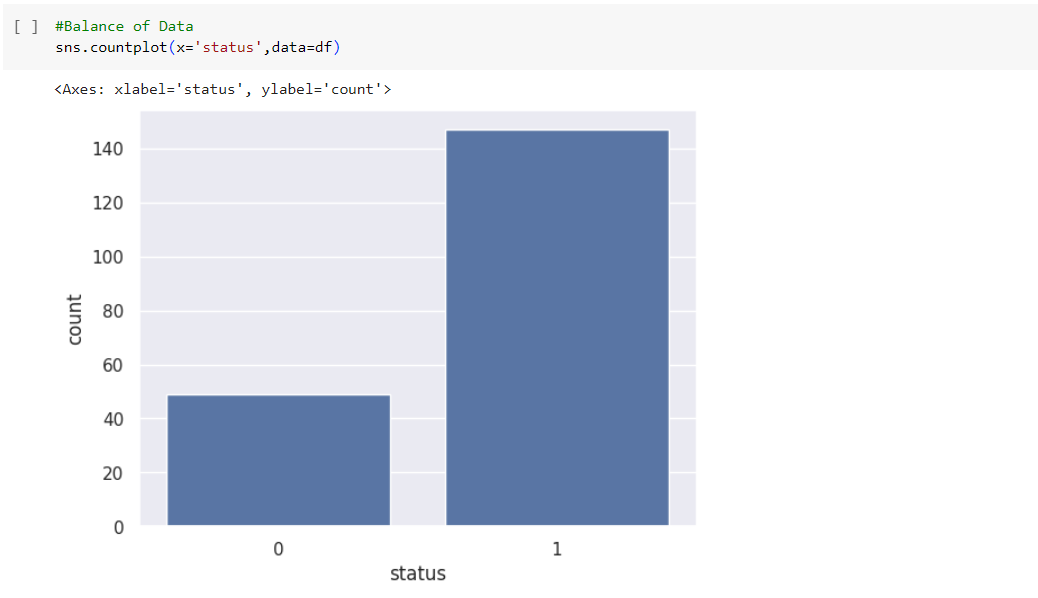


Figure 6.3: Balanced Data

* After analyzing the prediction model, we can clearly determine that the how many peoples effected by Parkinsion’s disease.



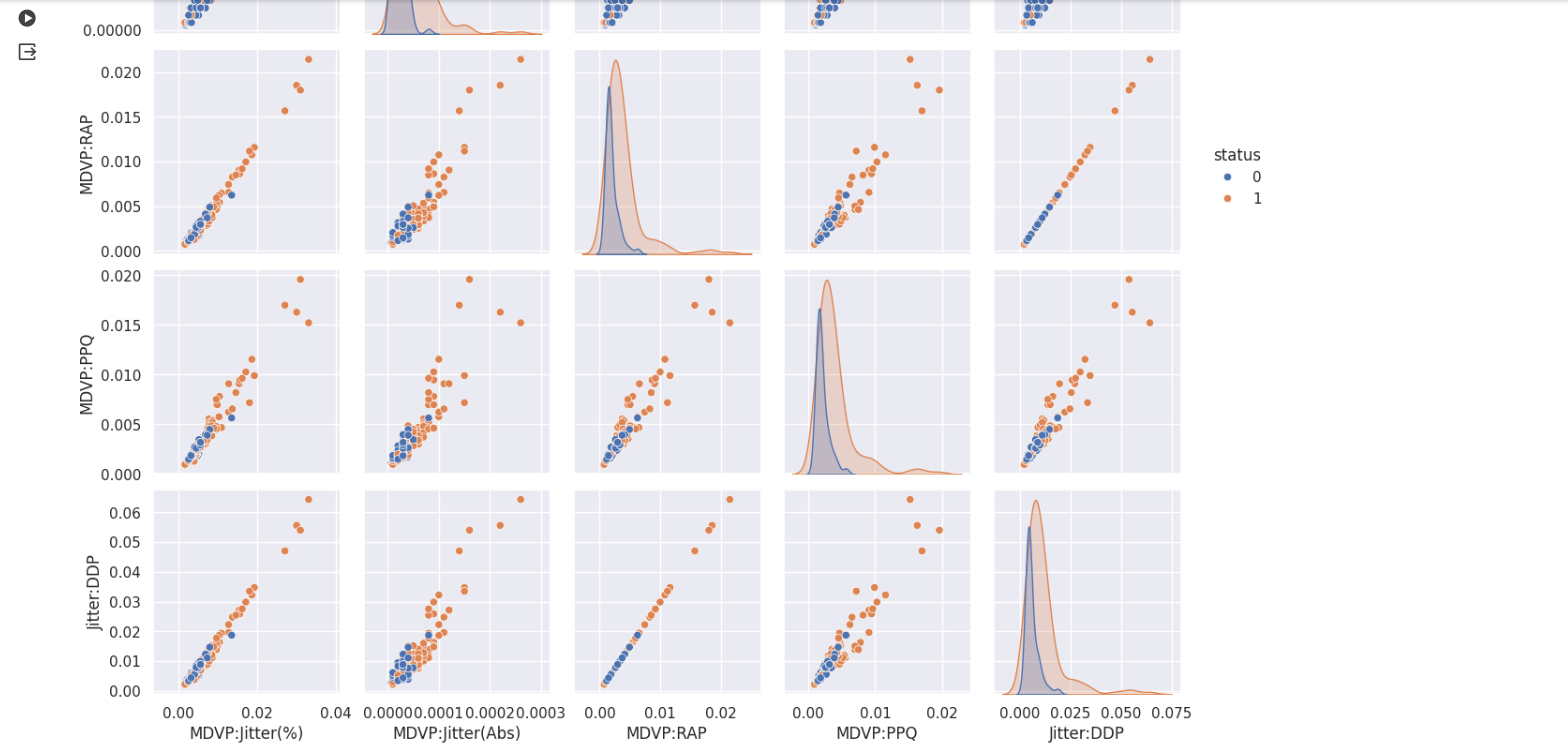


Figure 6.4: SNS Pairplot

* We have used three algorithm for prediction model
* Logistic regression
* KNN
* Naive bayes

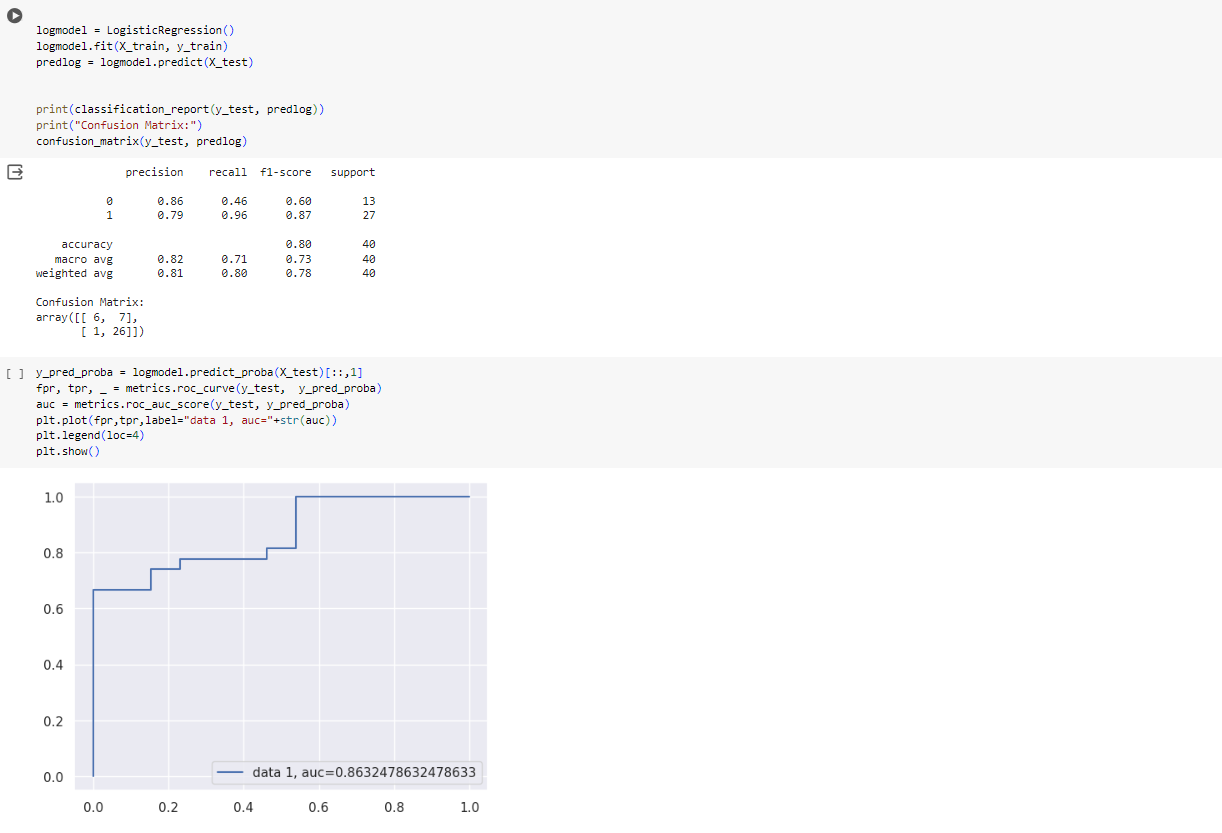


Figure 6.5: Logistic Regression Algorithm

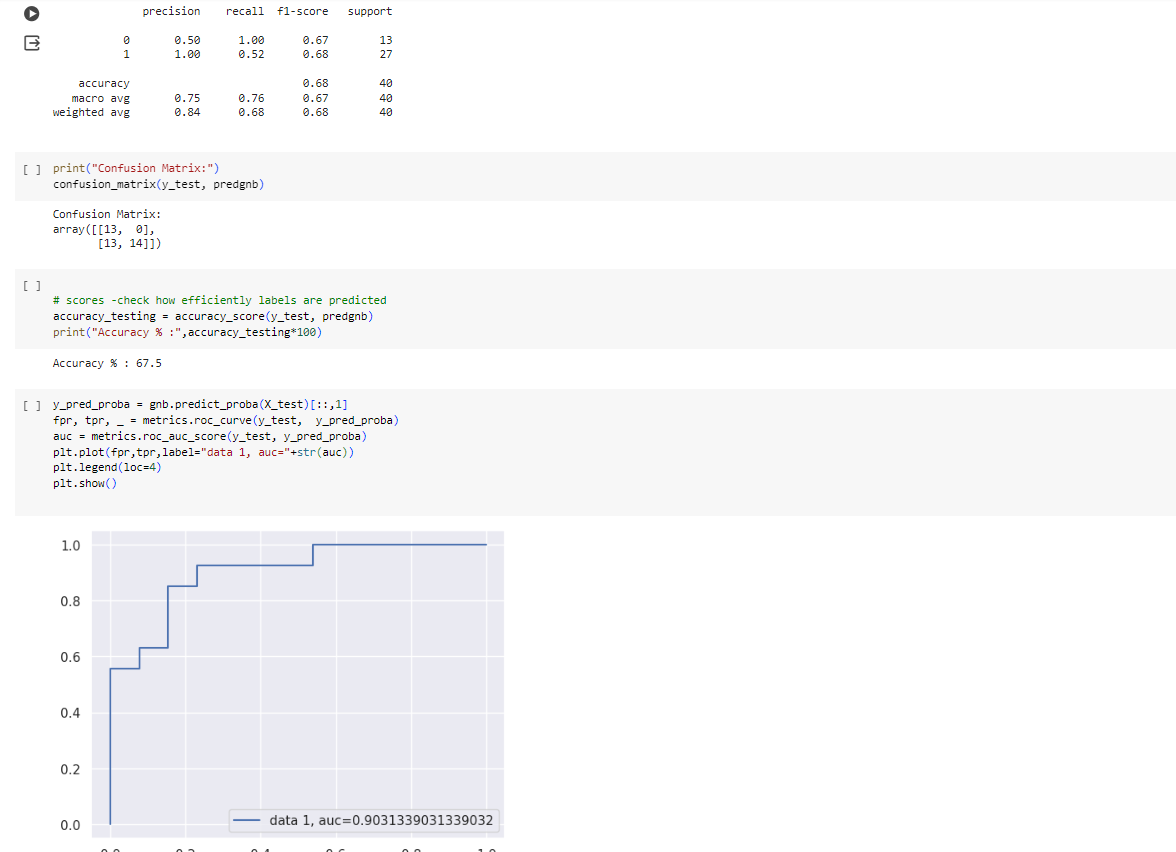


Figure 6.6: Naïve bayes Algorithm

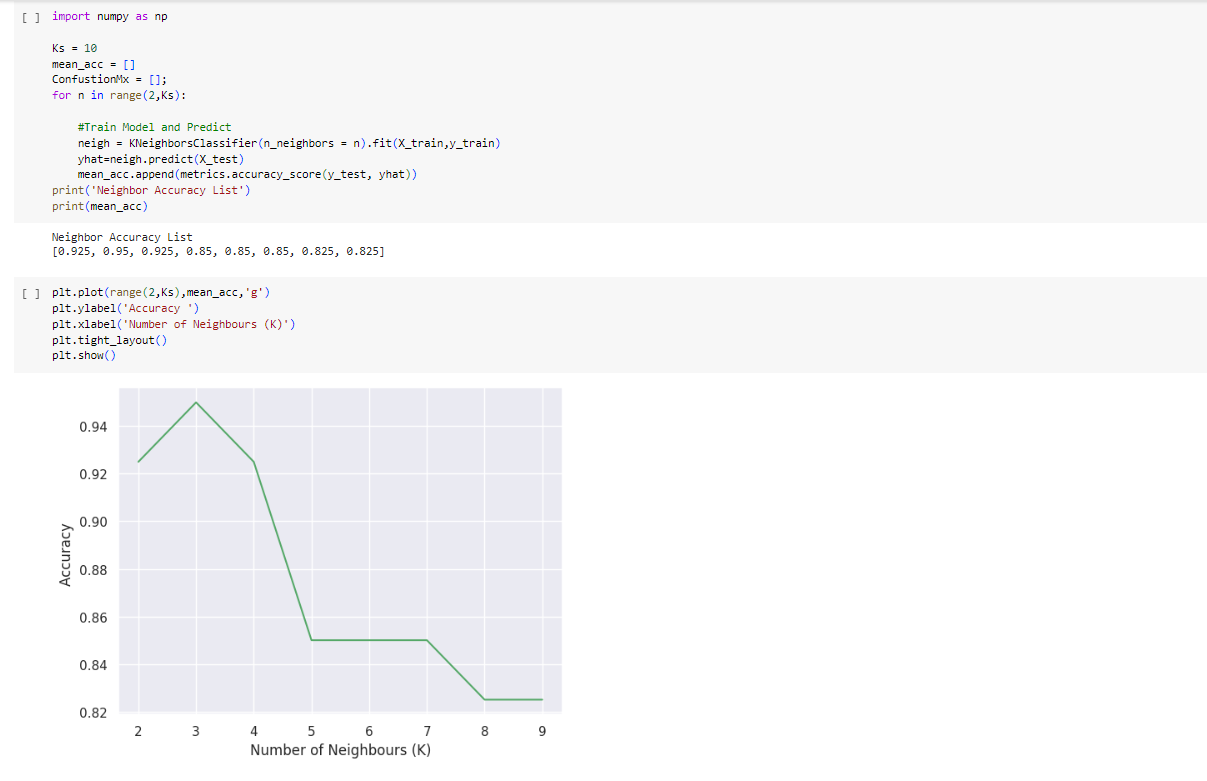


Figure 6.7: KNN Algorithm

* We have used Hybrid approach to get more clarity how are prediction model will be best so by Hybriding KNN and Naïve bayes it gives accuracy of 0.97 compare to other two algorithm .



Figure 6.8: Hybrid Accuracy of KNN + Naive Bayes



Figure 6.9: AUC score of comparison

* This shows the comparison between the three models and hybrid approach too.

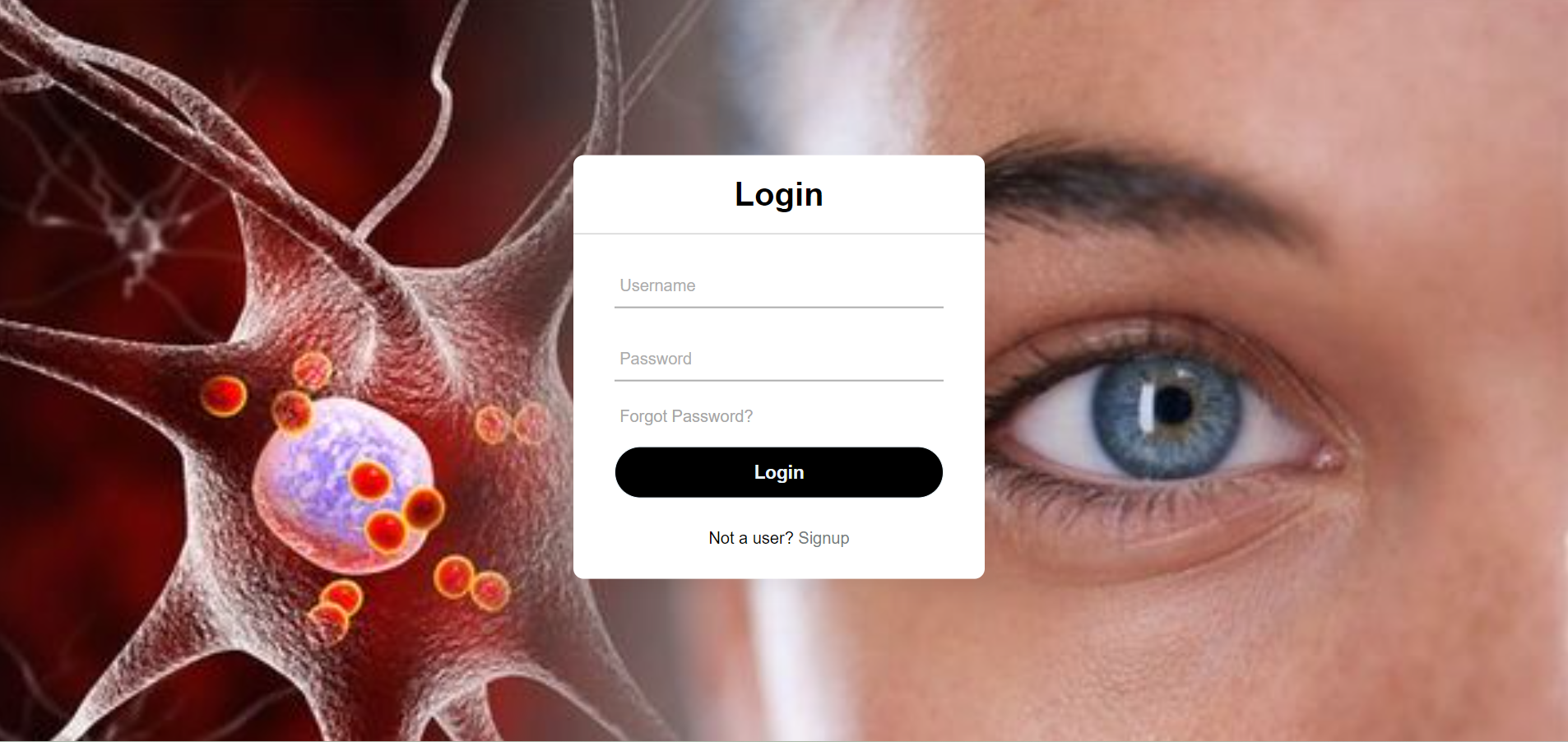


Figure 6.10: Login page



Figure 6.11:Diet Page

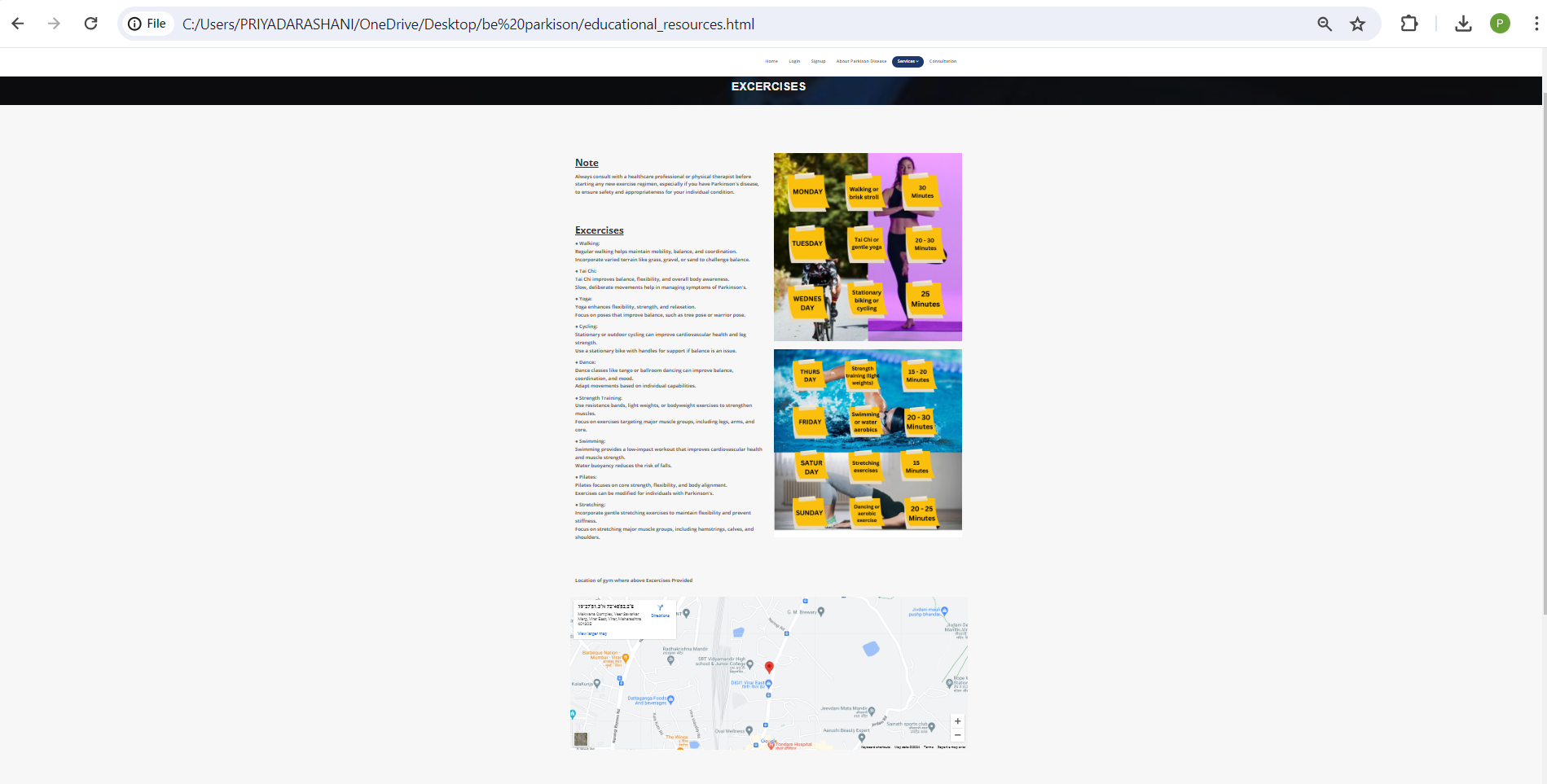
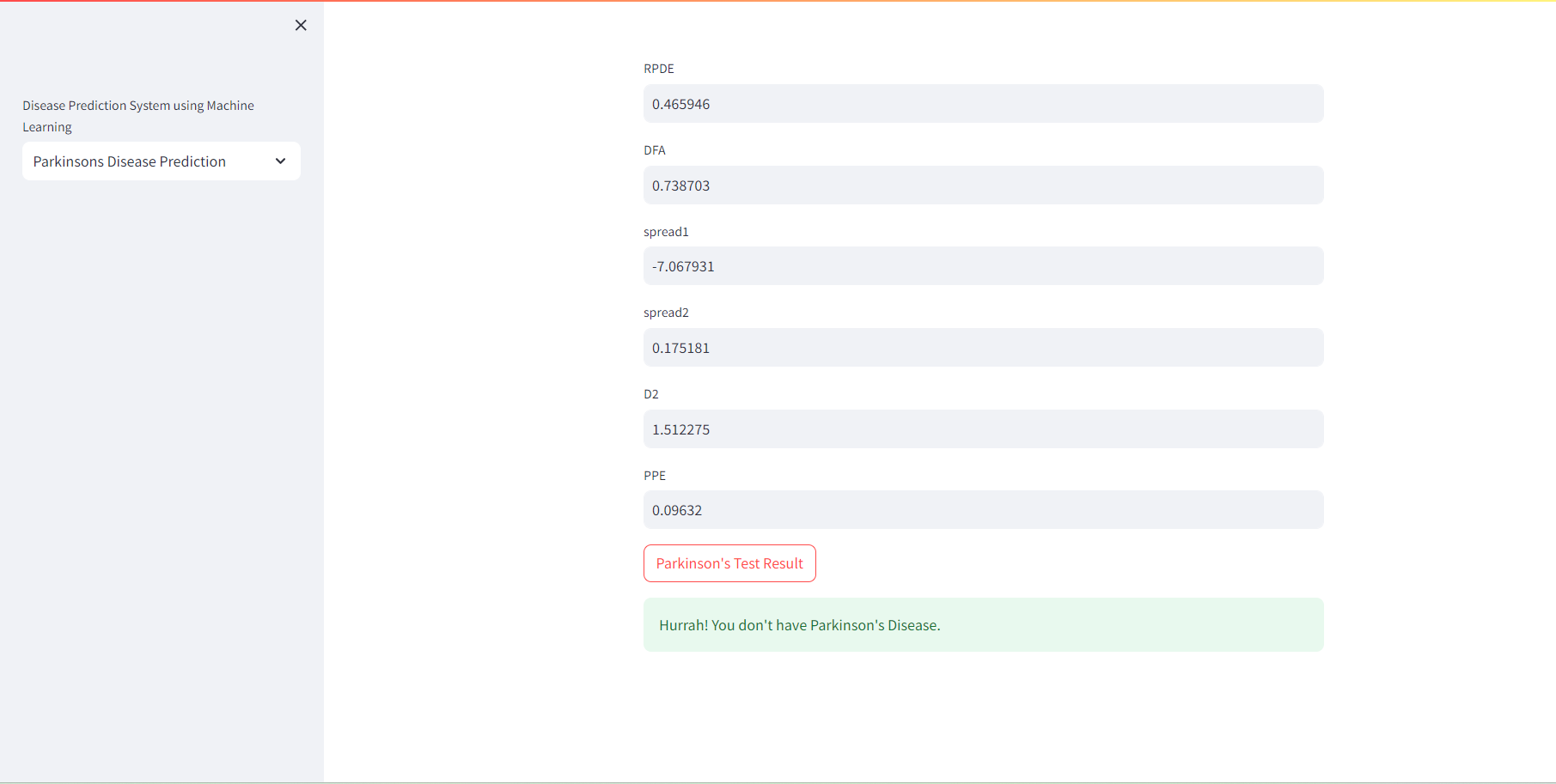


Figure 6.12:Exercise page

* In this prediction page it shows that if person has prakinsion’s disease or not so here we have shown the two pages where it have shown that the person has parkinsion’s disease according to readings.



* Here it shows that person don’t have parkinsion’s disease .



## Chapter 7

### Conclusion

* ML has been helping us to solve many problems in our day-to-day life. It has helped to analyze data provided and provide appropriate solutions to problems that occur. Due to which, the study uses the k-means algorithm.
* This study aimed to determine the reason behind the major cause of the increase in the number of road accidents happening around. For the past few years, it was noticed that the rate of road accidents had been increasing at an alarming rate due to various factors like drunk driving, problems related to climate, human error, etc. Considering this, the study of road accidents can play an important role to prevent road accidents that would have happened in the near future.
* From the data collected, we found a few major reasons for the cause of road accidents and the relationship between them. The data, when represented meaningfully, can help people to

avoid places or take caution while going through such areas as per the severity of a location in terms of accidents.

* Moreover, the study would help road transport systems to improve their work efficiency and help in controlling the number of casualties and loss of personal property taking place due to road accidents.

## References

1. ChangqinQuan , Kang Ren, And Zhiwei Luo, “A Deep Learning Based Method for Parkinson’s Disease Detection Using Dynamic Features of Speech,” Kobe University, Kobe 657-8501, Japan: Academic, January 20, 2021.
2. M. Nithya ,V.Lalitha , R K paveethra , M S Kumari, “Early Detection of Parkinson's Disease using Machine Learning & Image Processing,” Computer Science and Engineering Sri Sairam Engineering College 4Chennai ,Jan. 25 – 27, 2022.
3. Sandhiya.S,Dr.Ashok.S,Mr.G.Vishnu Vardhan Rao,Dr.Prabhu V , K.Mohanraj , Dr.R.Azhagumurugan, “Parkinson's Disease Prediction Using Machine Learning Algorithm,” Electrical & Elctronics Engineering Sri Sairam Engineering College

Chennai, India,

1. H. Gunduz, "Deep learning-based Parkinson’s disease classification using vocal feature sets", IEEE Access, vol. 7, pp. 115540-115551, 2019.
2. Sushila Palwea , Aditi Govindu , "Early detection of Parkinson's disease using machine learning",School of SCET, MIT WPU, Pune 411038, India,2023.
3. Mirza Muntasir Nishat1, Tasnimul Hasan1, Sarker Md. Nasrullah2, Fahim Faisal1, Md. Asfi-Ar-Raihan Asif1 and Md. Ashraful Hoque “Detection of Parkinson’s Disease by Employing Boosting Algorithms”Department of Electrical and Electronic Engineering, Islamic University of Technology, Dhaka, Bangladesh,2021
4. JieMei1,ChristianDesrosiers2 andJohannesFrasnelli1,3 “Machine Learning for the Diagnosis of Parkinson’s Disease: A Review of Literature”6 May 2021.
5. Imran Ahmed1, Sultan Aljahdali2, Muhammad Shakeel Khan1 and sSanaa Kaddoura3,"Classification of Parkinson Disease Based on Patient’s Voice Signal Using Machine Learning"1 Jan 2022
6. Md Abu Sayed1 , Duc Minh Cao2, Maliha Tayaba3, MD Tanvir Islam4, Md Eyasin Ul Islam Pavel5, Md Tuhin Mia6, Eftekhar Hossain Ayon7, Nur Nobe8 and Bishnu Padh Ghosh9 "Parkinson's Disease Detection through Vocal Biomarkers and Advanced Machine Learning Algorithms"2 Dec 2023.
7. Jawad Rasheed, Alaa Ali Hameed, Naim Ajlouni,Akhtar Jamil," Application of Adaptive Back-Propagation Neural Networks for Parkinson’s Disease Prediction National University of Computer and Emerging Sciences, Punjab, India, October 2020
8. Wu Wang , Junho Lee , FouziHarrou , (Member, Ieee), And Ying Sun, “Early Detection of Parkinson’s Disease Using Deep Learning and Machine Learning,” King Abdullah University of Science and Technology, August 21, 2020

## 

## Appendix

### Technologies Used

1. **Python**

Python is a popular programming language. It was created by Guido van Rossum, and released in 1991 .

It is used for:

* + web development (server side),
  + software development,
  + mathematics,
  + system scripting.

**What can Python do?**

* + Python can be used on a server to create web applications.
  + Python can be used alongside software to create workflows.
  + Python can connect to database systems. It can also read and modify files.
  + Python can be used to handle big data and perform complex mathematics.
  + Python can be used for rapid prototyping, or for production-ready software development.

1. **NumPy**

NumPy is a library for the Python programming language, adding support for large, multi- dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. The ancestor of NumPy, Numeric, was originally created by Jim Hugunin with contributions from several other developers. In 2005, Travis Oliphant created NumPy by incorporating features of the competing Numarray into Numeric, with extensive modifications. NumPy is open-source software and has many contributors.

## Publications

1. Jayesh Patil, Mandar Prabhu, Dhaval Walavalkar, and Vivian Brian Lobo. “Road accident analysis using machine learning.”  *In 2020 IEEE Pune Section International Conference (PuneCon)* , pp. 108-112 . IEEE, 2020, DOI: 10.1109/PuneCon50868.2020.9362403 .
2. Jayesh Patil, Vaibhav Patil, Dhaval Walavalkar, and Vivian Brian Lobo. “Road Accident

Analysis and Hotspot Prediction using Clustering.”  *In 2021 6 th  International Conference on Communication and Electronics Systems (ICCES)* , pp. 763-768. IEEE, 2021 , DOI:

10.1109 /ICCES51350.2021.9489074.

## Acknowledgments

We owe our deepest gratitude and regards towards the ones who offered their valuable guidance in the hour of need. We thank our guide  **Mrs. Aditi Raut** ( Assistant Professor, Department of Computer Engineering, St. John College of Engineering and Management) for his guidance and precious insights. His useful comments and feedback during the discussions we had and the encouragement to question every technical detail that we came across while partially completing this B.E. project helped us to a great extent.

We also take the opportunity to thank  **Ms. Deepika Nadar**  and (B.E. Project Coordinators) and  **Dr. Nilesh T. Deotale**  (Head of Department) who have always rendered their support and assistance in the best possible way.

We would also like to thank  **Dr. G. V. Mulgund**  (Principal, St. John College of Engineering and Management) and the  **members of Aldel Education Trust**  who have given us the background to conduct this B.E project.

26