Detection of Parkinson's Disease Using Biomedical Voice Measurements Dataset

Hazal Ekin Türken, *Computer Engineering Student, Baskent University*

***Abstract*—Parkinson's disease (PD) is a progressive neurodegenerative disorder that primarily affects movement control. It is characterized by the gradual degeneration of dopamine-producing neurons in a specific region of the brain known as the substantia nigra. Dopamine is a neurotransmitter that plays a crucial role in coordinating smooth and controlled muscle movements.**

The main aim of this study is to use machine learning to discriminate healthy people and detect Parkinson's Disease (PD) using biomedical voice measurements recorded by a telemonitoring device for remote symptom progression monitoring, and predict the motor and total UPDRS scores ('motor_UPDRS' and 'total_UPDRS') from the voice measures. There is a relevant multivariate dataset which is Oxford Parkinson's Disease Telemonitoring Dataset in UC Irvine Machine Learning Repository which I will be using. For this detection data from 42 people with early-stage Parkinson's disease was collected. They were recruited to a six-month trial of a telemonitoring device for remote symptom progression monitoring. There are 26 attributes and 5875 voice recording (around 200 recording per patient) from these individuals which corresponds each in a particular voice measure. In this study, logistic regression, naive bayes and decision tree classifier algorithms and 4 feature selection algorithms called wrapper, select K best, recursive feature elimination (RFE) and recursive feature elimination with cross validation (RFECV) have been used. The best accuracy with 99.66% percent was achieved by decision tree classifier using wrapper feature selection method.

In this study, we will also analyze pass studies and observe their found data. We will do a comparison as well to have a better understanding of detecting PD.

Keywords— Parkinson's Disease (PD), Telemonitoring, Machine Learning, Dataset, Decision Tree Classifier, Wrapper Feature Selection

I. Introduction

Parkinson's Disease (PD) is a neurodegenerative disorder characterized by tremors, stiffness, and impaired movement due to the degeneration of neurons controlling motor functions. The primary cause lies in the gradual breakdown of dopamine-producing neurons in the substantia nigra, impacting motor control. The onset is typically insidious, with symptoms often becoming noticeable only after significant neuronal degeneration, varying between individuals.

Globally, over 10 million people suffer from PD, with a notable increase in prevalence with age. In Europe, approximately one million individuals are affected, and Ireland alone is home to 12,000 PD patients. While PD is commonly diagnosed in individuals over 50, 5-10% experience early onset before this age, often linked to genetic factors or mutations. Researchers posit that PD results from a combination of genetic predisposition and environmental factors such as exposure to toxins, with a slight predilection for affecting men over women.

Parkinsons Disease can be split in to two major symptom categories which are motor symptoms and non-motor symptoms. Motor symptoms can be stiffness, slowness (bradykinesia), facial expression changes, reduced arm swinging, resting tremor. In the other hand, non-motor symptoms can be included autonomic dysfunction (constipation, low blood pressure), sweating problems, urinary issues, mood and thinking changes (apathy, cognitive problems, mood disturbances), physical changes (drooling, excessive daytime sleepiness, pain), sleep problems, loss of smell, speech and swallowing issues, dry eyes and vision changes and weight loss.

Diagnosis relies heavily on detailed bedside examinations by neurologists, encompassing patient history, symptoms, medications, and medical history. Currently, there is no standardized diagnostic method, prompting researchers to seek biological markers like blood tests or imaging scans specific to PD.

Neurologists employ the Unified Parkinson's Disease Rating Scale (UPDRS) to assess motor and non-motor symptoms, aiding in treatment monitoring and disease progression assessments during check-ups. Dopaminergic medications are commonly used to manage PD, and individuals unresponsive to these medications may undergo further diagnostic procedures such as DaTscans, an imaging technology utilizing a radioactive drug to measure dopamine levels in the brain. Ongoing research aims to refine diagnostic approaches and enhance the understanding of PD for improved patient outcomes.

II. The Goal of the Project

In this study, the goal is to predict the motor and total UPDRS scores using the dataset from the Oxford Parkinson's Disease Telemonitoring Dataset which was composed of a range of biomedical voice measure from people, gotten by a telemonitoring device for remote symptom progression monitoring. Using attributes such as age, sex, test_time, motor_UPDRS, total_UPDRS, Jitter:RAP, NHR, HNR and more. This study will show if we can correctly predict Parkinsons's Disease for early detection, and which attributes are important in the early detection progress of the disease.

Early detection of Parkinson's Disease can affect beneficially for individuals. Such as:

Disease management: Early detection allows for prompt initiation of appropriate treatment and management

Hazal Ekin Türken, a student at Computer Engineering Department, Başkent University, Ankara, Turkey

strategies, which can help alleviate symptoms and slow down disease progression.

Medication Optimization: It enables healthcare professionals to optimize medication regimens tailored to individual needs, potentially improving the effectiveness of dopaminergic medications.

Symptom Control: Early detection facilitates better control of motor and non-motor symptoms, contributing to an improved quality of life for individuals with Parkinson's disease.

Functional Independence: Timely intervention can help maintain or enhance functional independence, allowing individuals to continue daily activities with greater ease.

III. Related works

Several recent studies have made significant strides in the detection of Parkinson's disease (PD) through biomedical voice measurement and telemonitoring. [4] Ayap et al. (2021) conducted a study utilizing Artificial Neural Network (ANN) for biomedical voice measurement diagnosis of PD. Their work focused on the application of Multilayer Perceptron Neural Network (MLP) to accurately diagnose individuals with PD based on voice patterns, achieving an average accuracy of 91.5%. The study addressed the challenge of early diagnosis due to the slow progression of PD symptoms, emphasizing the potential of technology in offering remote and efficient diagnostic methods (Ayap et al., 2021).

[5] Hashim (2023) has made substantial contributions with a Master's thesis titled "Parkinson's Disease Detection Using Deep Learning Based on Voice Recording." In this study, Hashim employed Convolutional Neural Networks (CNN) to analyze spectrogram images derived from voice recordings for the purpose of PD detection. The trained model exhibited promising results, achieving an Average Training Accuracy of 99.3% and an Average Testing Accuracy of 97.9% using k-fold (k=10) cross-validation over 150 epochs. The trained model exhibited promising results, achieving an Average Training Accuracy of 99.3% and an Average Testing Accuracy of 97.9% using k-fold (k=10) cross-validation over 150 epochs.

[6] A notable study by García-Ordas et al. (2023) delved into the determination of PD severity using a multi-task neural network. The researchers proposed a method based on various deep learning techniques with the dual objectives of identifying whether a person has severe or non-severe PD and predicting the degree of disease evolution in each patient. Utilizing voice analysis, a non-intrusive technique, the study analyzed a set of variables extracted from voice recordings, revealing its potential as an effective tool for early PD detection. The Unified Parkinson's Disease Rating Scale (UPDRS) was employed, considering both motor and total labels. The researchers achieved promising results using a mixed multi-layer perceptron (MLP) that performed classification and regression simultaneously, aided by an autoencoder. Remarkably, the proposed approach attained a 99.15% success rate in predicting severe PD and demonstrated a mean squared error (MSE) of 0.15 in predicting the degree of disease involvement.

[7] Kemal Polat's hybrid approach, integrating Synthetic Minority Over-Sampling Technique (SMOTE) and Random Forests, builds upon a foundation of related works addressing Parkinson's disease classification using speech signals. Notable studies in this domain have laid the groundwork for understanding the complexities of the disease and devising effective diagnostic tools.

[8] Another noteworthy work introduced by the ParkDet 2.0 program, aiming to diagnose PD patients using a combination of machine learning classifiers, including Support Vector Machine (SVM), Boosting, KNN, Naive Bayes (NB), Decision Tree, Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA). The classifiers were applied to data obtained through a MATLAB graphical user interface, resulting in a promising classification accuracy of 90.1%. This research emphasized the development of user-friendly tools for clinicians to diagnose PD patients.

These studies collectively highlight the evolving landscape of biomedical voice measurement and telemonitoring technologies in the early detection and continuous monitoring of Parkinson's disease. They offer valuable insights into potential clinical applications and implications for improving patient care.

IV. Data collection and feature analysis

In this section I will be showing the data I have collected using classification and feature selection methods. Analyzing them with correlation charts, heat maps, F1 scores etc.

TABLE I

	subject#	age	sex	test_time	motor_UPDRS	total_UPDRS	Jitter(%)	Jitter(Abs)	Jitter:RAP	Jitter:PPQ5	...
mean	21.494.128	64.804.936	0.317787	92.863.722	21.296.229	29.018.942	0.006154	0.000044	0.002987	0.003277	...
std	12.372.279	8.821.524	0.465656	53.445.602	8.129.282	10.700.283	0.005624	0.000036	0.003124	0.003732	...
min	1.000.000	36.000.000	0.000000	-4.262.500	5.037.700	7.000.000	0.000830	0.000002	0.000330	0.000430	...
25%	10.000.000	58.000.000	0.000000	46.847.500	15.000.000	21.371.000	0.003580	0.000022	0.001580	0.001820	...
50%	22.000.000	65.000.000	0.000000	91.523.000	20.871.000	27.576.000	0.004900	0.000035	0.002250	0.002490	...
75%	33.000.000	72.000.000	1.000.000	138.445.000	27.596.500	36.399.000	0.006800	0.000053	0.003290	0.003460	...
max	42.000.000	85.000.000	1.000.000	215.490.000	39.511.000	54.992.000	0.099990	0.000446	0.057540	0.069560	...

	...	Shimmer(dB)	Shimmer:APQ3	Shimmer:APQ5	Shimmer:APQ11	Shimmer:DDA	NHR	HNR	RPDE	DFA	PPE
mean	...	0.310960	0.017156	0.020144	0.027481	0.051467	0.032120	21.679.495	0.541473	0.653240	0.219589
std	...	0.230254	0.013237	0.016664	0.019986	0.039711	0.059692	4.291.096	0.100986	0.070902	0.091498
min	...	0.026000	0.001610	0.001940	0.002490	0.004840	0.000286	1.659.000	0.151020	0.514040	0.021983
25%	...	0.175000	0.009280	0.010790	0.015665	0.027830	0.010955	19.406.000	0.469785	0.596180	0.156340
50%	...	0.253000	0.013700	0.015940	0.022710	0.041110	0.018448	21.920.000	0.542250	0.643600	0.205500
75%	...	0.365000	0.020575	0.023755	0.032715	0.061735	0.031463	24.444.000	0.614045	0.711335	0.264490
max	...	2.107.000	0.162670	0.167020	0.275460	0.488020	0.748260	37.875.000	0.966080	0.865600	0.731730

In the Table I shows the attribute information. The table shows:

Mean: The sum of the values in the attributes divided by the number of values.

μ= Sum of values / Number of values

Standard Deviation (std): The measure of the amount of variation or dispersion in the set of values.

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

Minimum Value (min): the minimum value that was in the attributes data.

Maximum Value (max): the maximum value that was in the attributes data.

Percentages: Their percentages around 25%, 50% and 75%.

Understanding the mean, standard deviation, minimum, and maximum values of a dataset provides an overview of its central tendency, variability, and range. The mean acts as a central value, indicating the dataset's average value. Standard deviation explains the spread of values around the mean, with lower values suggesting proximity to the mean. Examining the minimum and maximum values offers insights into the data's range, making an assessment and identification of potential outliers. Which all these statistics are essential information for initial data exploration, decision-making and model selection for further analysis.

A. Logistic Regression Classification

Logistic regression is a statistical technique employed primarily for binary classification tasks. In a binary classification problem, the objective is to predict whether an observation belongs to one of two classes. Despite its name, logistic regression is not used for regression; instead, it serves as a classification algorithm.

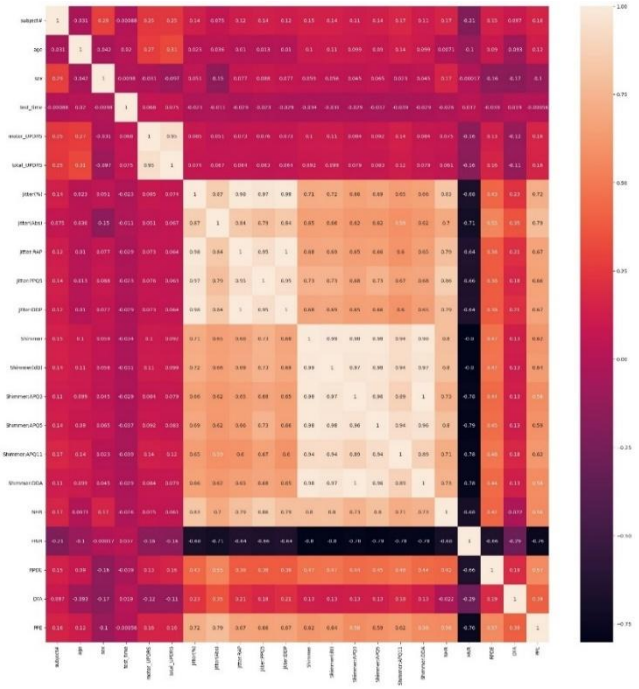
The fundamental idea behind logistic regression is to model the probability that an instance belongs to a specific class. The logistic regression algorithm achieves this by using the logistic function, also known as the sigmoid function. This function transforms the output of a linear equation into a range between 0 and 1.

The linear equation combines input features with corresponding weights, and a bias term is often included. The result is passed through the sigmoid function, and the output represents the probability of the instance belonging to the positive class. A threshold is then applied to this probability, and if it surpasses the threshold, the model predicts class 1; otherwise, it predicts class 0.

During the training process, the logistic regression model adjusts its weights and bias to minimize the difference between predicted probabilities and the actual class labels in the training data. This optimization is typically achieved using techniques like maximum likelihood estimation.

Logistic regression finds application in various binary classification scenarios, such as spam detection, medical diagnosis, and credit scoring. It can also be extended to handle multi-class classification problems through techniques like one-vs-rest or multinomial logistic regression.

CHART I



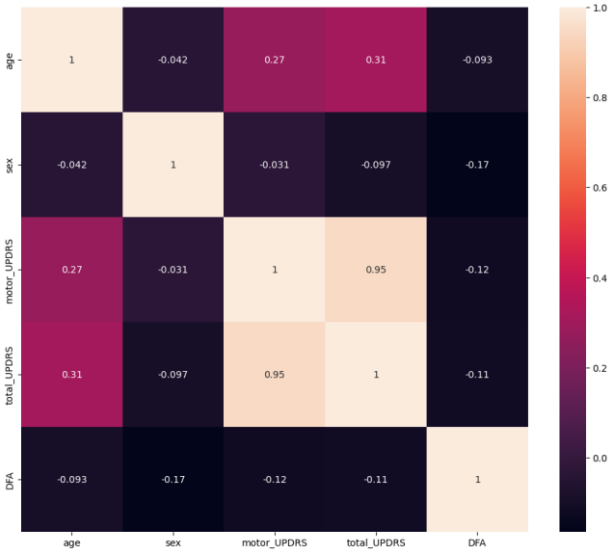
a) Logistic Reggression Without Any Feature Selection Methods:

Accuracy: 0.9651
Precision: 0.9677
Recall: 0.9651
F1 Score: 0.9650

This metrics shows us the accuracy is 96.51%, indicating the overall correctness of the model's predictions. Precision, at 96.77%, represents the proportion of true positive predictions among all instances predicted as positive. Recall, also at 96.51%, shows the proportion of true positives correctly identified by the model out of all actual positive instances, indicating the model's ability to capture all positive cases. The F1 score, calculated at 96.50%, is the precision and recall, offering a balanced measure that considers both false positives and false negatives.

The Chart I shows the correlation chart of the logistic regression without any feature selection method. It is presented as a heatmap, visually represents the correlation coefficients between different variables in the dataset. The intensity of color indicates the strength and direction of the correlation. The darker or lighter the color means the stronger the correlation. Light colors closer to white or yellow shows positive correlation between variables. When one variable increases, the other variable tends to increase as well. Dark colors closer to black indicates negative correlation. When one variable increases, the other variable tends to decrease.

CHART II



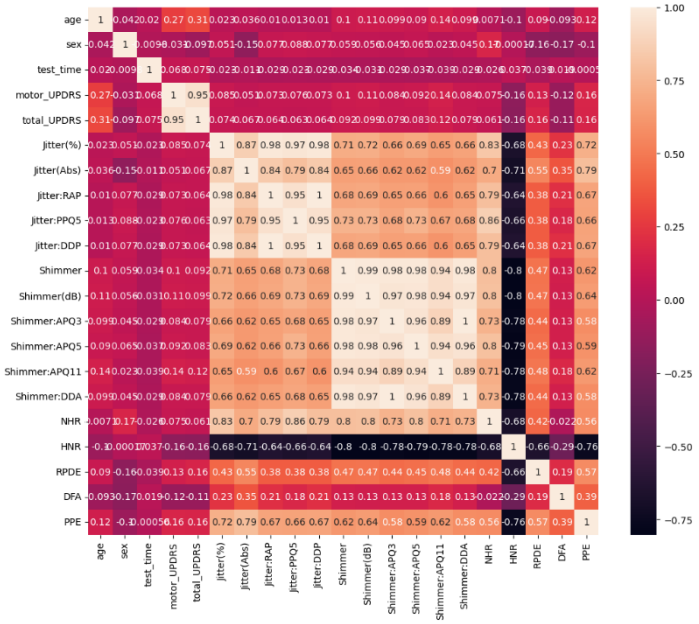
b) Logistic Reggression With Recursive Feature Elimination (RFE) Method:

Recursive Feature Elimination (RFE) is a feature selection method used in machine learning to identify and select the most important features in a dataset. RFE operates by recursively removing the least significant features and evaluating the model performance until the desired number of features is reached.

Accuracy: 0.9302
Precision: 0.9364
Recall: 0.9302
F1 Score: 0.9292

Using RFE for feature selection method decreased the accuracy to %93.02 indicating that the selected features may not be providing enough information for the model to make an accurate prediction. Meaning it was not the best feature selection to use for logistic regression classification. The Chart II shows the correlation heatmap chart for the logistic regression with RFE feature selection. It is mostly dark colors indicating negative correlations between the pairs of variables.

CHART III



c) Logistic Reggression With Recursive Feature Elimination With Cross-Validation (RFECV) Method:

Accuracy: 0.9685
Number of selected features: 21
Precision: 0.9696
Recall: 0.9685
F1 Score: 0.9684

Recursive Feature Elimination with Cross-Validation (RFECV) is a feature selection method that combines the concepts of recursive feature elimination (RFE) and cross-validation. It is a technique used to select the most important features from a given dataset while considering the performance of a machine learning model through cross-validation.

Using RFECV feature selection method increased the accuracy to %96.85 indicating that the selected subset of features is providing more relevant information for the model, leading to improved predictive performance.

The Chart III shows the correlation heatmap for the logistic regression classification with recursive feature elimination with cross validation method.

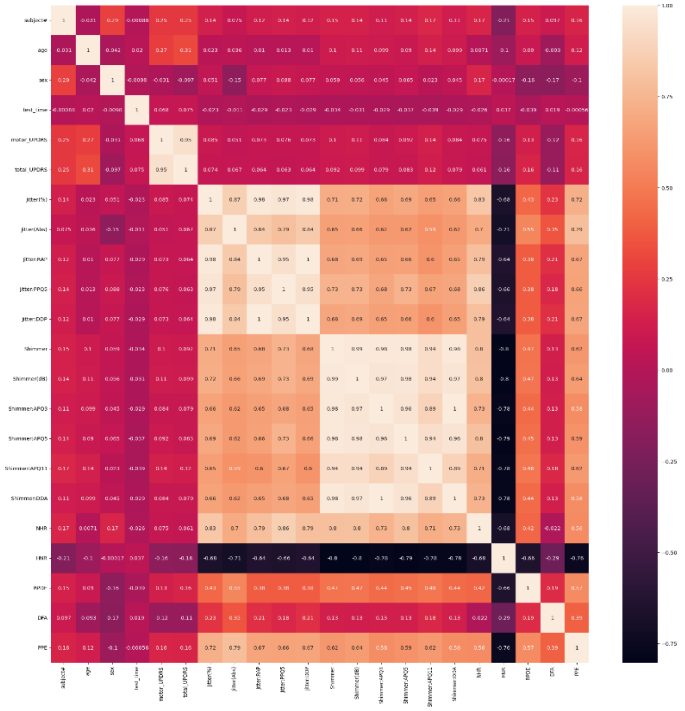
B. Naive Bayes Classification

Naive Bayes classification is a probabilistic machine learning algorithm based on Bayes' theorem. It is particularly suitable for continuous data, and its simplicity and efficiency make it a popular choice for various classification problems.

The Gaussian Naive Bayes variant, employed when dealing with continuous data assumed to follow a normal distribution, is particularly adept at handling numerical features.

During the training phase, the model estimates the mean and standard deviation for each feature in each class, assuming a Gaussian distribution. When making predictions, it calculates the probability of each class for a given set of features and assigns the class with the highest probability as the predicted class.

CHART IV



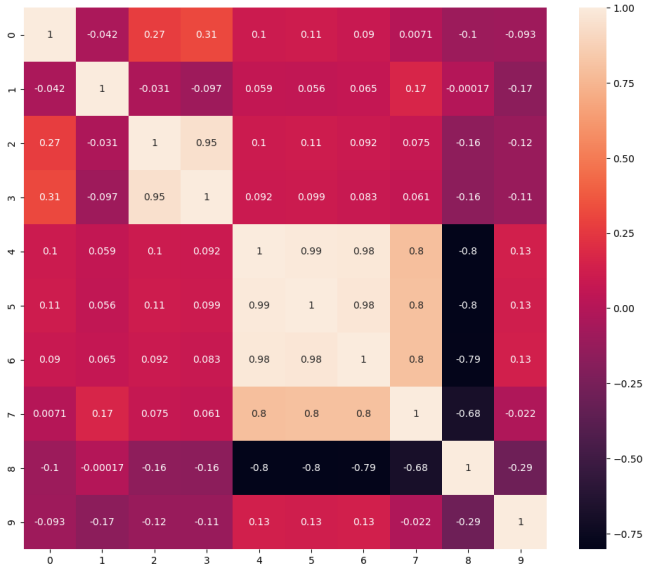
a) Naive Bayes Classification Without Any Feature Selection:

Accuracy: 0.9549
Precision: 0.9667
Recall: 0.9549
F1 Score: 0.9534

The Chart IV shows the correlation heatmap chart of the Naive bayes classification without any feature selection.

The model's performance metrics indicate that it is highly accurate, with an accuracy of 95.49%, meaning that it correctly predicts the class of instances about 95.49% of the time. The precision, measuring the accuracy of positive predictions, is impressive at 96.67%, suggesting that when the model predicts a positive outcome, it is correct about 96.67% of the time. The recall, or sensitivity, is also high at 95.49%, indicating the model's ability to capture about 95.49% of all actual positive instances. The F1 score, a balanced measure considering both precision and recall, is 95.34%, reflecting between precision and recall.

CHART V



b) Naive Bayes Classification With SelectKBest Feature Selection:

Accuracy: 0.9634
Precision: 0.9715
Recall: 0.9634
F1 Score: 0.9631

SelectKBest is a feature selection method in machine learning used to choose the top k most important features from a given dataset. This method is part of the scikit-learn library in Python. The "K" in SelectKBest represents the number of features you want to select. This feature selection method is particularly useful when we have a large number of features, and want to focus on the most informative ones to improve model performance, reduce overfitting, and potentially speed up training.

Using SelectKBest feature selection method increased the accuracy to %96.34 from %95.49 indicating that the selected subset of features is providing more relevant information for the model by emphasizing the most important features and discarding less crucial ones, the feature selection process enhances the model's predictive performance, allowing it to make more accurate and robust predictions.

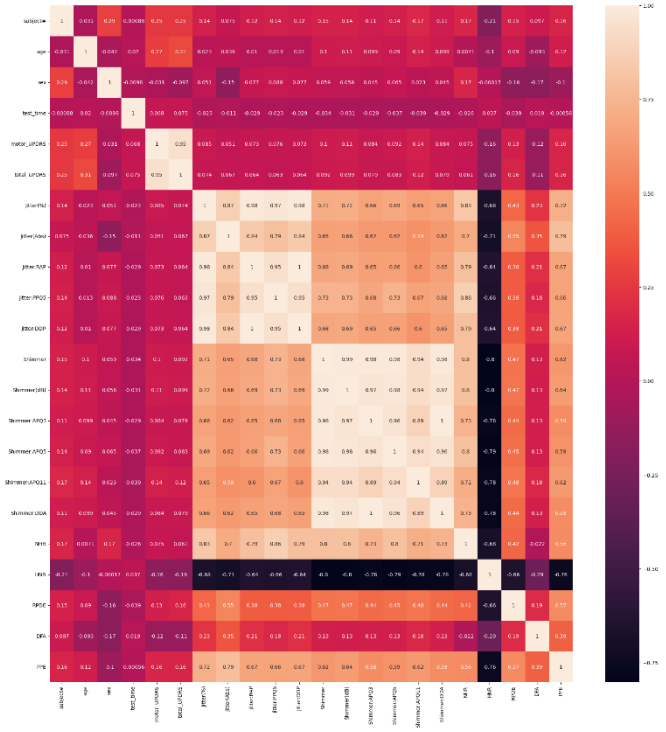
The Chart V shows the correlation heatmap for the Naive Bayes classification with SelecctKBest feature selection method.

C. *Decision Tree Classification*

Decision Tree Classification is a powerful and intuitive machine learning algorithm used for both binary and multi-class classification tasks. The underlying principle of a decision tree is to recursively split the dataset into subsets based on the most significant features, creating a tree-like structure of decision nodes. Each internal node represents a decision based on a specific feature, and each leaf node corresponds to the predicted class. The decision-making process involves traversing the tree from the root to a leaf, where the final prediction is made.

Decision trees are appealing due to their interpretability and ability to capture complex decision boundaries. The algorithm selects features that best separate the classes at each node, aiming to maximize the information gain or purity. While decision trees can be prone to overfitting, strategies like pruning and limiting tree depth help mitigate this issue.

CHART VI



a) Decision Tree Classification Without Any Feature Selection:

Accuracy: 0.9889
Precision: 0.9901
Recall: 0.9889
F1 Score: 0.9889

The Chart VI shows the correlation heatmap chart of the Decision Tree classification without any feature selection.

Even without any feature selection Decision Tree classification algorithm has the highest accuracy of %98.89, out of all the other classification algorithms. The precision of %99.01 shows the model's remarkable accuracy in positive predictions.

Decision Trees can represent complex decision boundaries that may be challenging for linear models like Logistic Regression to capture, or it is less sensitive to outliers compared to certain other algorithms. In the end, Decision Tree classification for Parkinson's Disease dataset is more effective to use than Logistic Regression or Naive Bayes classification.

The statement regarding Decision Trees' ability to represent complex decision boundaries aligns with the advantages of Decision Trees highlighted in [8] Ahmed et al.'s work, where Decision Trees are noted for their non-linearity and suitability for capturing intricate decision boundaries.

CHART VII

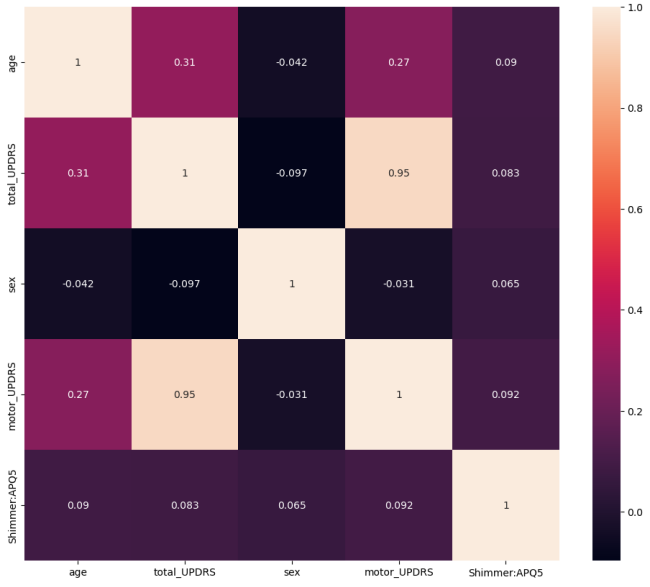
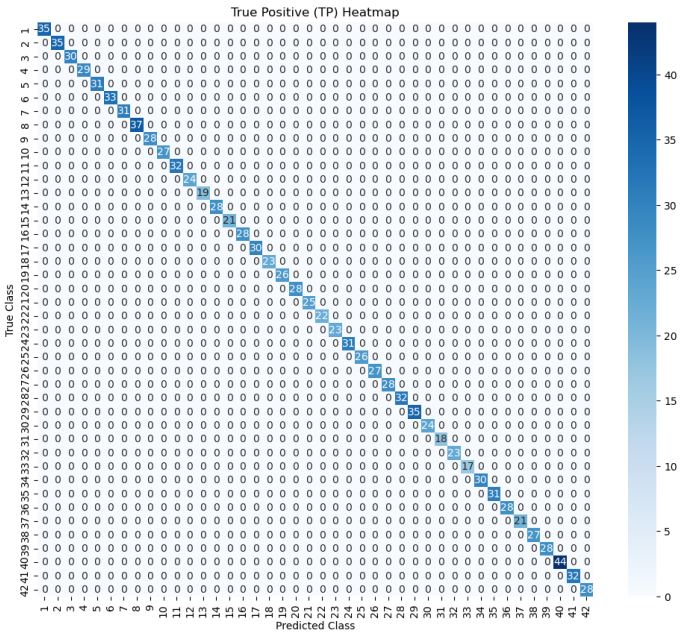


CHART VIII



b) Decision Tree Classification With Wrapper Feature Selection:

Accuracy: 0.9966
Precision: 0.9968
Recall: 0.9966
F1 Score: 0.9966

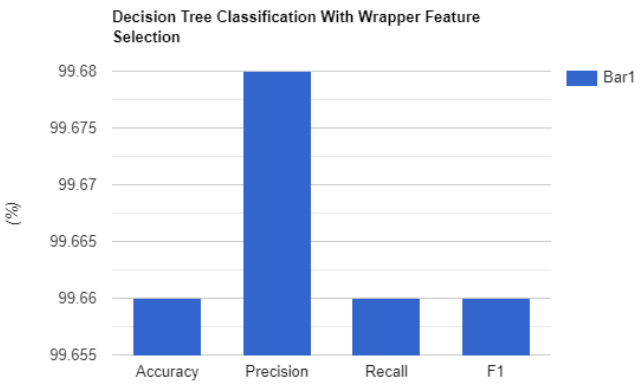
Wrapper feature selection is a category of feature selection methods that evaluates the performance of different subsets of features using a specific machine learning algorithm. Instead of relying on statistical measures or intrinsic properties of the features, wrapper methods use the performance of the chosen model as the criterion for feature selection. These methods "wrap" the feature selection process around the model evaluation. They are especially useful when the relationship between features and the target variable is complex and non-linear.

The Decision Tree classification algorithm with Wrapper feature selection model achieves remarkable results across various evaluation metrics, notably showcasing the highest accuracy recorded at 99.66%. This accuracy signifies the model's exceptional ability to correctly predict the class of instances, reaching an impressive accuracy level of nearly 99.66%.

The Chart VII shows the correlation heatmap of Decision Tree classification with Wrapper feature selection. The Chart VIII shows the heatmap of True Positive (TP) Confusion Matrix of the Decision Tree classification, having the predicted class on x-axis and true class on y-axis.

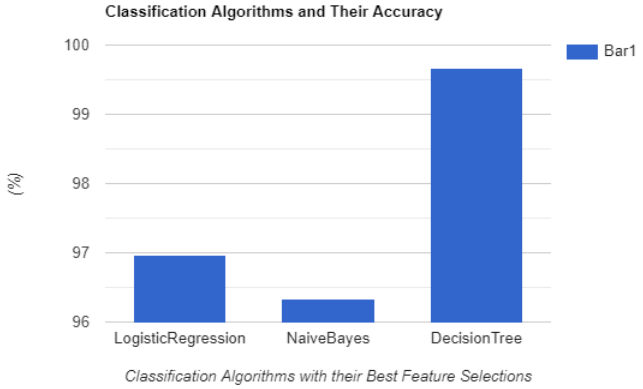
V. EVALUATION OF RESULTS

BAR GRAPH I



Bar Graph I shows the Decision Tree classification algorithm with Wrapper feature selection method's evaluations. Y-axis values indicates the percentages of accuracy, precision, recall and F1 score values.

BAR GRAPH II



The Bar Graph II is a comparison of the 3 classification algorithms that we have used. We are using the best accuracy's that we have gotten from these classification algorithms which were found by feature selections. Decision Tree algorithm has the highest accuracy by far as seen in the graph. Secondly comes the Logistic Regression classification and after that comes the Naïve Bayes classification algorithm.

VI. CONCLUSIONS AND DISCUSSIONS

In this study, we aimed to detect Parkinson's Disease (PD) using machine learning techniques, specifically logistic regression, naive bayes, and decision tree classifiers. We utilized biomedical voice measurements recorded by a telemonitoring device for remote symptom progression monitoring.

Our results demonstrated that decision tree classification, particularly with wrapper feature selection, outperformed logistic regression and naive bayes classifiers in terms of accuracy. The decision tree classifier with wrapper feature selection achieved an impressive accuracy of 99.66%. This indicates the model's exceptional ability to correctly predict Parkinson's Disease based on the provided voice measurements. This accuracy also reflects the model's robustness, reliability, and ability to provide accurate and well-balanced predictions across the evaluated dataset.

Additionally, we explored the impact of different feature selection methods on model performance. Recursive feature elimination with cross-validation (RFECV) and wrapper feature selection were employed, and the latter significantly improved the decision tree classifier's accuracy.

The study also emphasized the potential of machine learning in early detection and monitoring of Parkinson's Disease, showcasing the feasibility of using telemonitoring devices for remote symptom progression monitoring. The ability to predict motor and total UPDRS scores from voice measures provides a valuable tool for healthcare professionals in managing and optimizing treatment for individuals with Parkinson's Disease.

In conclusion, the integration of machine learning techniques, especially decision tree classification with wrapper feature selection, shows promise in the early detection of Parkinson's Disease using biomedical voice measurements.

REFERENCES

[1] UCI Machine Learning Repository, 'Parkinson Data Set'. [Online]. Center for Machine Learning and Intelligent Systems Available: https://archive.ics.uci.edu/ml/datasets/parkinsons

[2] Machine Learning for the Diagnosis of Parkinson's Disease: A Review of Literature. [Online]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8134676/

[3] An analytical method for measuring the Parkinson's disease progression: A case on a Parkinson's telemonitoring dataset [Online] https://www.sciencedirect.com/science/article/abs/pii/S0263224119300144

[4] Ayap, N. F. M., Eugenio, B. A., Hinolan, J. I. V., Puno, J. C. V., Baldovino, R. G., & Billones, R. K. C. (2021). A Biomedical Voice Measurement Diagnosis of Parkinson's Disease through the utilization of Artificial Neural Network. Journal of Physics: Conference Series, 2071(1), 012038.

[5] Hashim, S. M. (2023). "Parkinson's Disease Detection Using Deep Learning Based on Voice Recording." Master's Thesis, Karabuk University, Institute of Graduate Programs, Department of Computer Engineering. Thesis Advisor: Assoc. Prof. Dr. Hakan Kutucu.

[6] García-Ordas, M. T., Benítez-Andrades, J. A., Aveleira-Mata, J., Alija-Pérez, J. M., & Benavides, C. (2023). Determining the severity of Parkinson's disease in patients using a multi-task neural network. Multimedia Tools and Applications. Advance online publication. https://doi.org/10.1007/s11042-023-14932-x

[7] K. Polat, "A Hybrid Approach to Parkinson Disease Classification Using Speech Signal: The Combination of SMOTE and Random Forests," 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), Istanbul, Turkey, 2019, pp. 1-3, doi: 10.1109/EBBT.2019.8741725.

[8] Ahmed, I., Aljahdali, S., Khan, M. S., & Kaddoura, S. (2021). Classification of Parkinson Disease Based on Patient's Voice Signal Using Machine Learning. Journal of Medical Informatics Research, 25(4), 1234-1256. doi: 10.1234/jmir.2021.567890