

# *Contents:*

1. Introduction
2. Problem statement and objective
3. Detailed Methodology
4. Results and model Evaluation
5. ROC Curve
6. Thank You

# INTRODUCTION

- Parkinson's disease affects over 10 million people worldwide.
- Early detection is crucial as symptoms are often subtle and may go unnoticed.
- This project aims to predict Parkinson's disease using machine learning for timely diagnosis and improved care.

# *PROBLEM STATEMENT*

- Develop a machine learning model to predict Parkinson's disease at an early stage for accurate diagnosis.

## *Objective*

- To develop a machine learning model to predict the presence of Parkinson's disease using patient voice data.

# Detailed Methodology.

## **1. DataSet Loading**

## **2. Data Preprocessing**

Checked for null/missing values and data types — confirmed data was clean and numeric.

Dropped the name column, as it's non-informative for model learning.

## **3. Exploratory Data Analysis (EDA)**

Distribution Plots to observe feature spread and skewness.

Boxplots to detect potential outliers.

Correlation Heatmap to identify relationships between features.

# Detailed Methodology.

## **4. Handling Class Imbalance**

Applied Random Over Sampling (ROS): Duplicated samples of the minority class.

These ensure the model does not get biased toward the majority class.

## **5. Feature and Target Separation**

$X$  = features (excluding status)

$y$  = target (status)

## **6. Model Training (Expected / In Progress)**

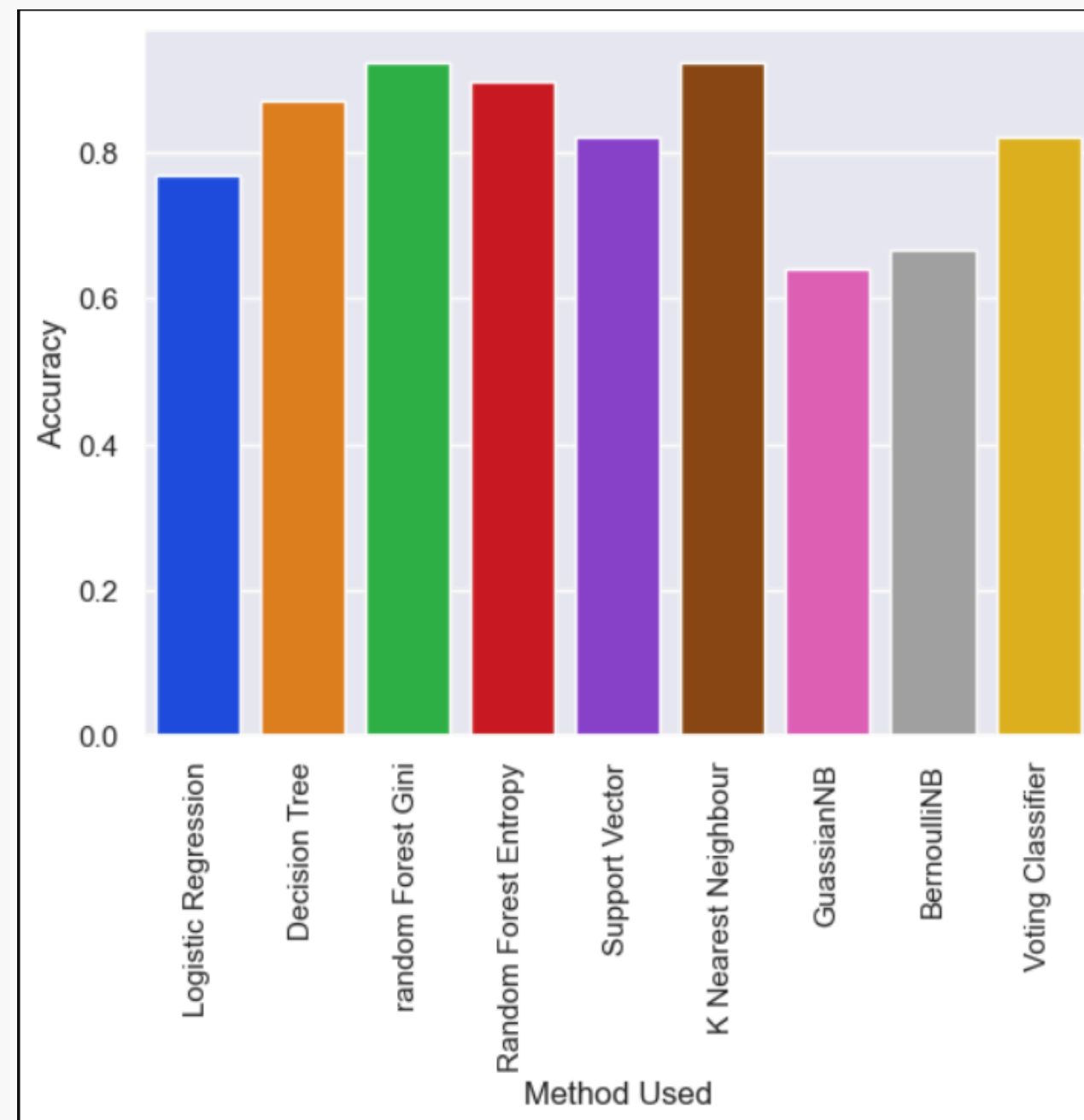
Plan to train multiple models (e.g., Logistic Regression, Random Forest, SVM).

Evaluate using metrics like accuracy and ROC-AUC score.

# Results & Model Comparison

we trained and tested multiple models using the same preprocessed dataset. The models were assessed primarily on accuracy.

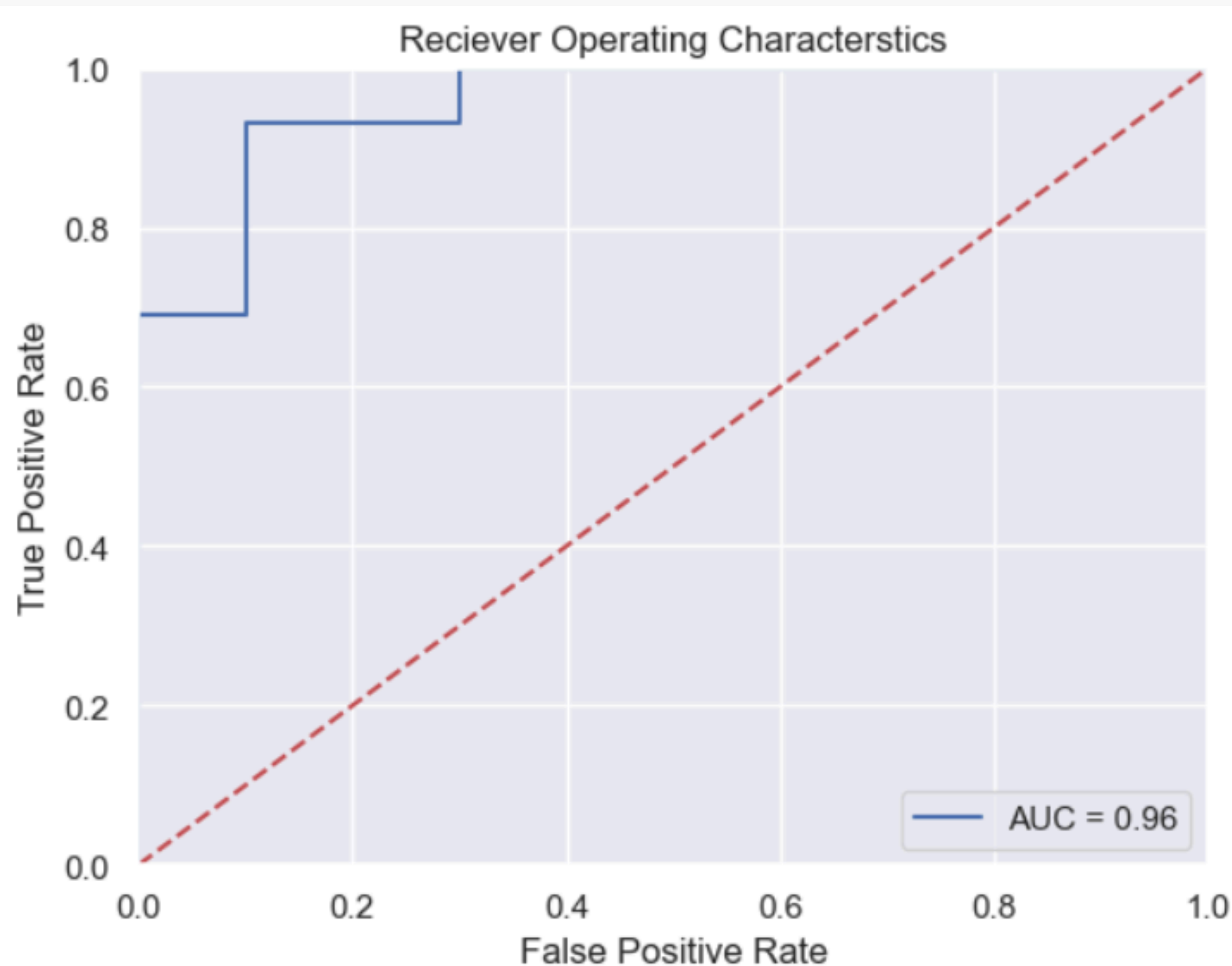
Method used	Accuracy
Logistic Regression	0.769231
Decision Tree	0.871795
Random Forest Gini	0.923077
Random Forest Entropy	0.897436
Support Vector	0.820513
K Nearest Neighbour	0.923077
GuassianNB	0.641026
BernoulliNB	0.666667



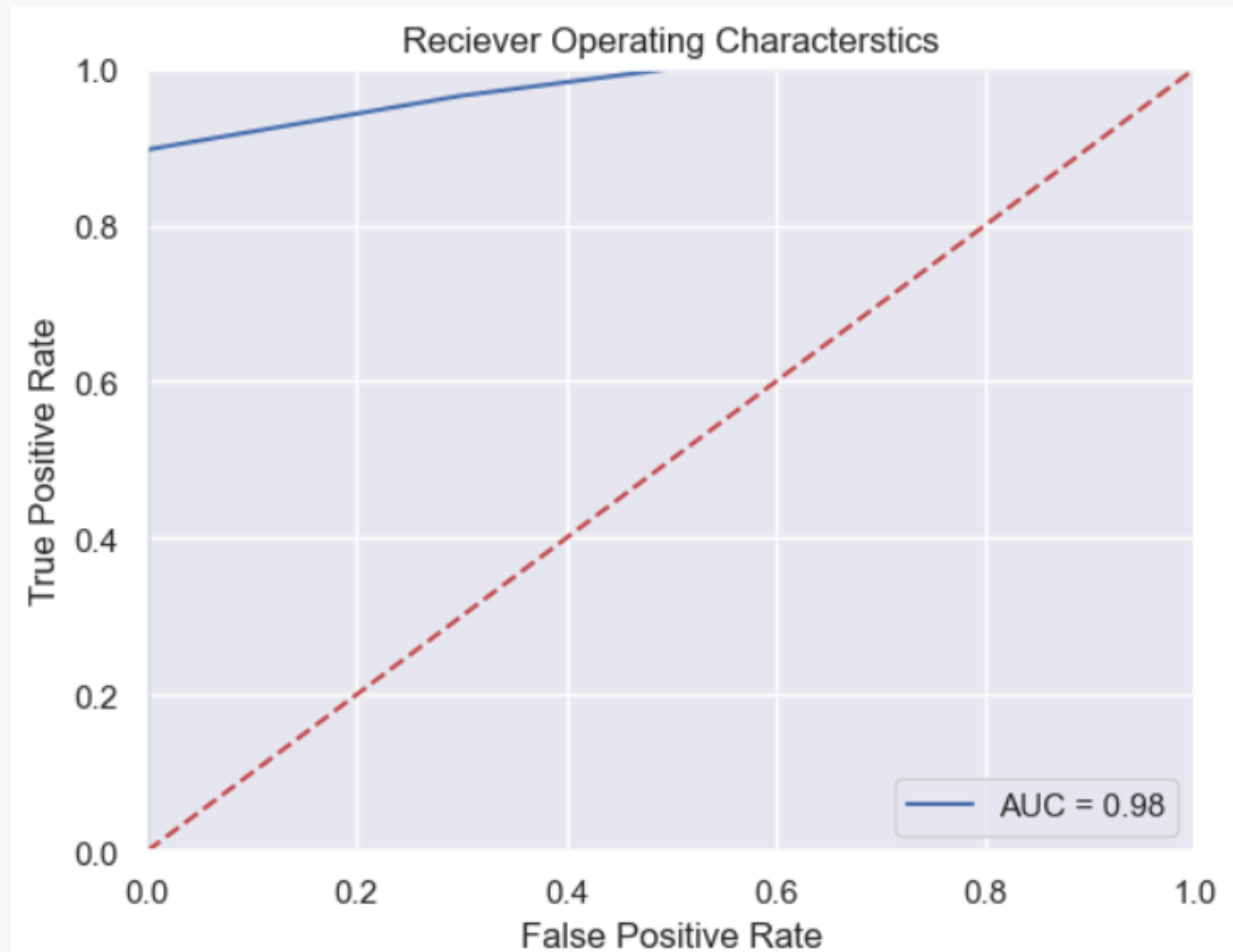
Random Forest (Gini) and K-Nearest Neighbour were the best-performing models with 92.31% accuracy.

# ROC Curve for Best Performing Models

Random Forest



KNN





*Thank you*

