HEALTH CARE – DRUG PERSISTENCY

20TH AUGUST 2022

Background

One of the challenge for all Pharmaceutical companies is to understand the persistency of drug as per the physician prescription.

To solve this problem, ABC pharma company is seeking to automate this process of identification.

Problem Statement

- To understand the persistency of a drug as per the prescription given by the physician is an important question faced by pharmaceutical companies.
- The problem here is to build a classification model to understand the persistency (persistent or not) of a drug for the given dataset.

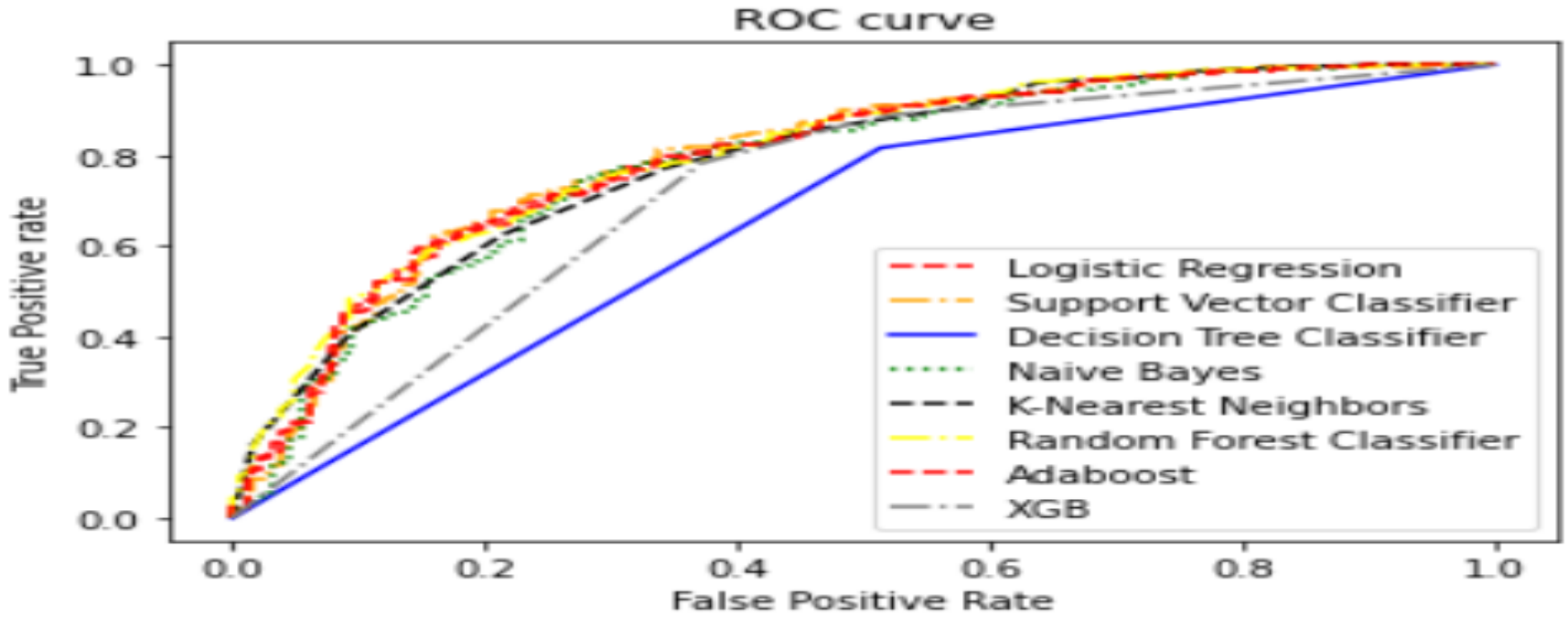
Models

- Linear
  - Logistic Regression – Linear
  - Support Vector Classifier
  - Naïve Bayes
  - K-Nearest Neighbors
  - Decision Tree
- Ensemble Bagging
  - Random Forest
- Boosting
  - Adaboost
  - XGBoost

Tools Used for Comparison

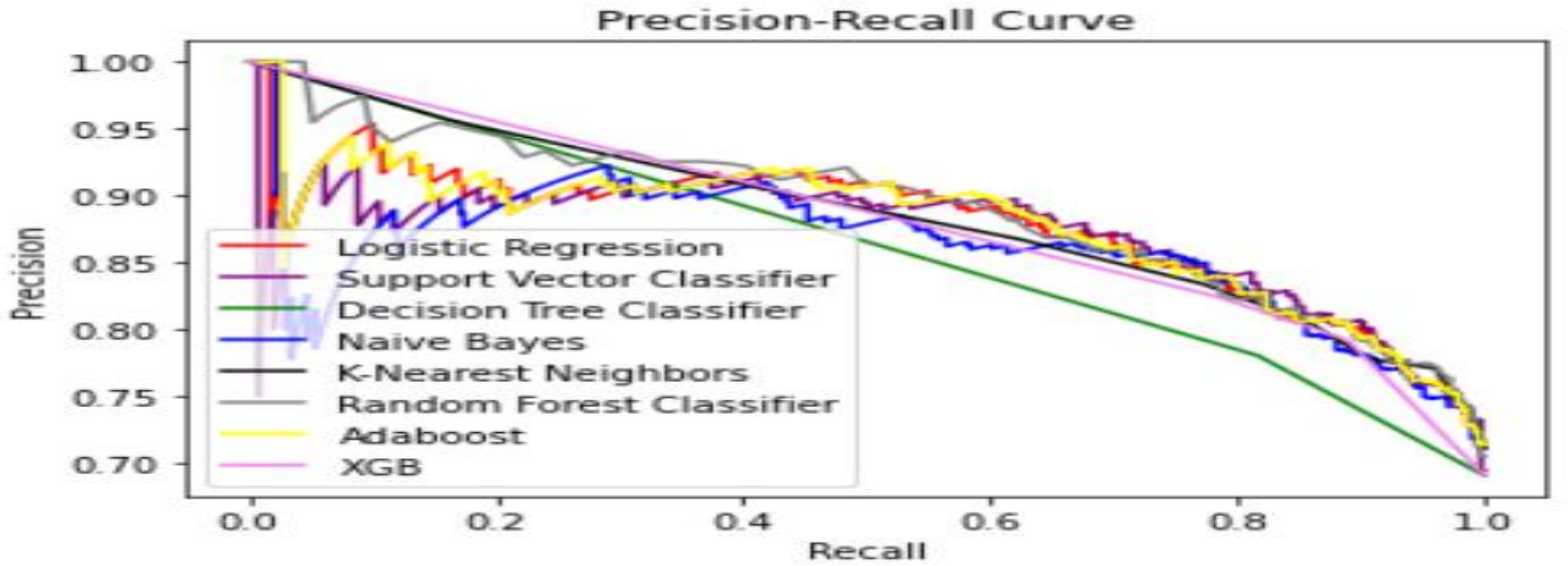
- Accuracy Score
- Confusion Matrix
- Classification Report
- Specificity, Sensitivity
- Receiver Operating Characteristics (ROC) Curve
- Precision Recall Curve
- AUC Value
- F1 Score

ROC CURVE



Random Forest Classifier and Adaboost has similar Curve.

Precision-Recall CURVE



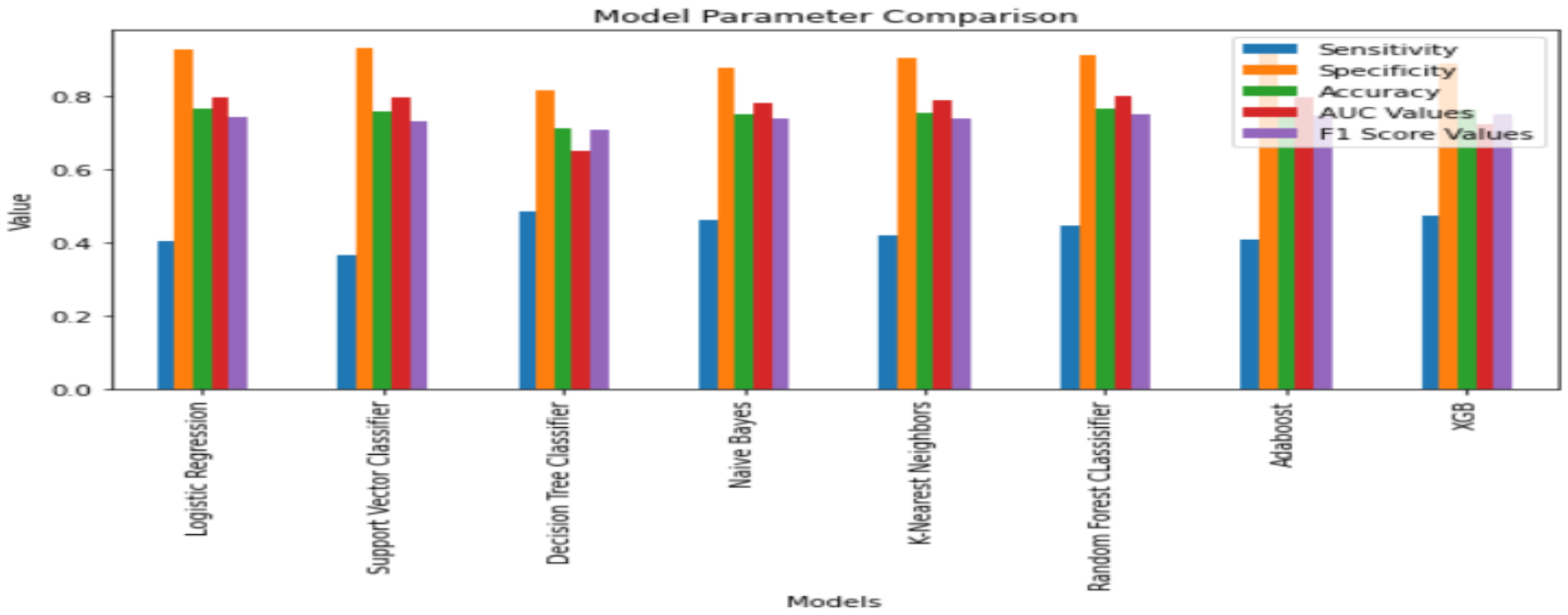
Comparison Matrix

1 to 8 of 8 entries Filter

Models	Sensitivity	Specificity	Accuracy	AUC Values	F1 Score Values
Logistic Regression	0.40437158469945356	0.9290953545232273	0.7668918918918919	0.7965315911125362	0.7446727939489631
Support Vector Classifier	0.366120218579235	0.9339853300733496	0.7584459459459459	0.7982951888519246	0.731491659766629
Decision Tree Classifier	0.48633879781420764	0.8166259168704156	0.714527027027027	0.6514823573423116	0.7099516198358813
Naive Bayes	0.4644808743169399	0.8801955990220048	0.7516891891891891	0.7804855238018892	0.7395148319266621
K-Nearest Neighbors	0.4207650273224044	0.9070904645476773	0.7567567567567568	0.7882279850895828	0.738338562781425
Random Forest CLassisifier	0.44808743169398907	0.9119804400977995	0.768581081081081	0.8031050008684382	0.7521116723854484
Adaboost	0.4098360655737705	0.9290953545232273	0.768581081081081	0.7972263417371437	0.7469214302459266
XGB	0.47540983606557374	0.8899755501222494	0.7618243243243243	0.7228546234318008	0.7495324808615949

The best model from the above values is Random Forest Classifier.

Comparison Matrix



The best model from the above values is Random Forest Classifier.

Proposed Model

The proposed model is Random Forest Classifier.

- Testing Accuracy: 0.769
- Sensitivity: 0.448
- Specificity:0.912
- AUC Value:0.803
- F1 Score:0.752

Prediction System

- Inputs a series of 6 values (Assumption: Input happens after PCA)
- Apply Standard Scaler
- Input to RFC model to predict the output
- Convert the numerical output to Categorical

Summary

- From various Classification models, Random Forest Classifier was chosen based on the various parameter values.
- It has the highest accuracy of 76.9
- This could be the efficient model for the automation of the prediction of persistent or non-persistent drugs.

Thank You