SMOKE DETECTION

1 MOSTAFA TAREK

2 MOHAMED EZZ AL-REGAL

3 ABD EL-RAHMAN AHMED

Abstract

This research extensively explores various machine learning models for Smoke Detection, evaluating their strengths and weaknesses. It also examines how different feature selection techniques impact model performance. By comparing these methods on benchmark datasets, the study aims to guide the selection of optimal models and features for effective smoke detection. Overall, it provides valuable insights for practitioners and researchers in this field.

PYTHON

Introduction

Our objective is to deeply explore machine learning models for Smoke Detection and their performance with various feature selection techniques. Motivated by the urgency of timely smoke detection for safety, it covers established algorithms like SVM, XGB, Logistic Regression, KNN, NN, Decision Trees, and Random Forests. It emphasizes how feature quality impacts system effectiveness and uses techniques like variance thresholding, correlation analysis, feed-forward, and k-best selection. The structure details the experimental setup, datasets, and evaluation metrics, aiming to aid practitioners in system development. Overall, it contributes to both machine learning knowledge and critical improvements in smoke detection, offering insights applicable to diverse classification tasks.

Data Exploration

01

reading the CSV file

02

Rename the column's name

03

Check for Nulls.

04

Check for Duplicates.

Feature Selection

01
Variance Thresholding

02
Correlation Analysis

03
Feed-Forward Selection

04
k-Best Selection

05
Mutual Information Classification

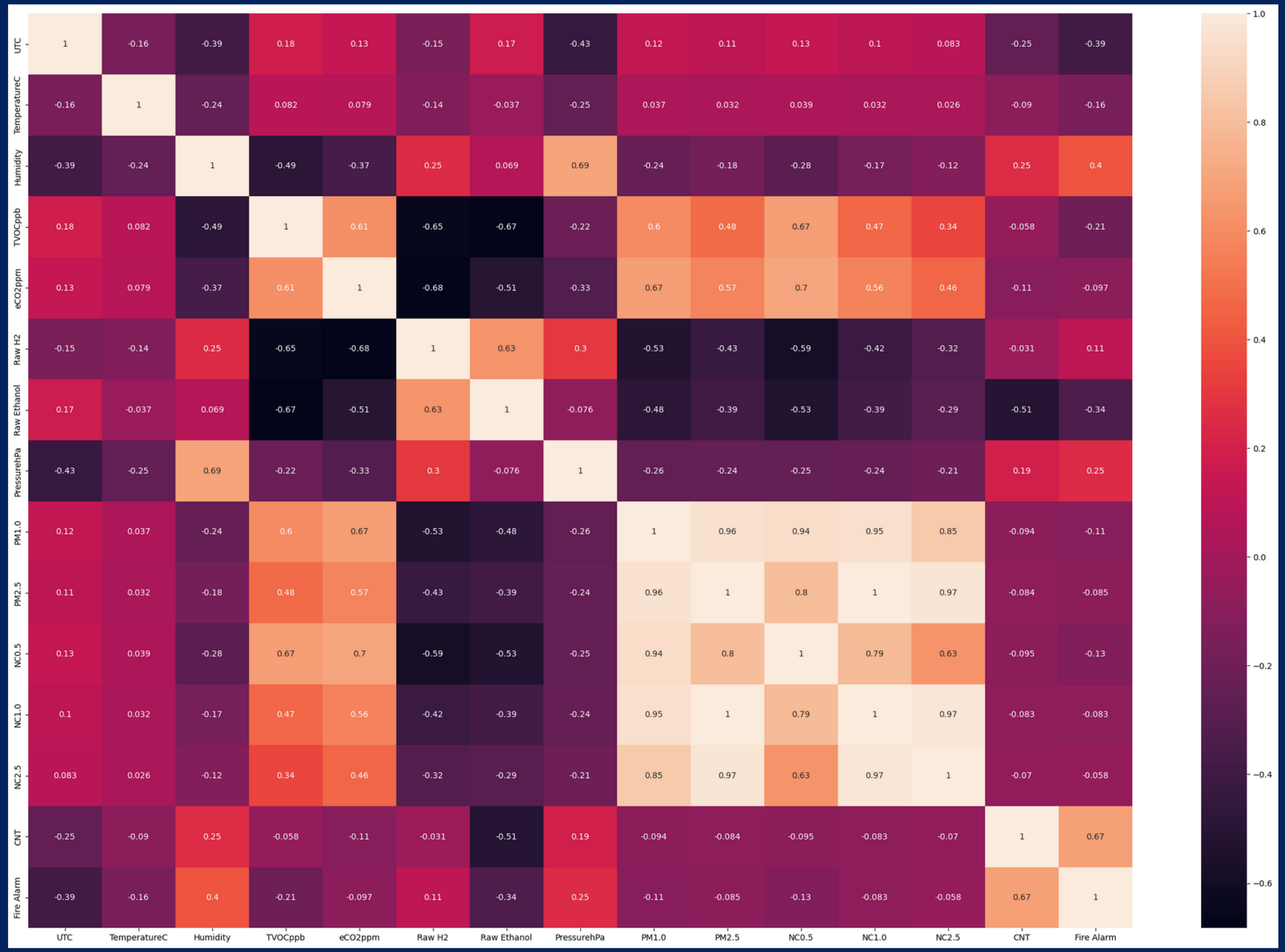
06
Mean Absolute Difference

07
Model Fit Selection for XGBoost

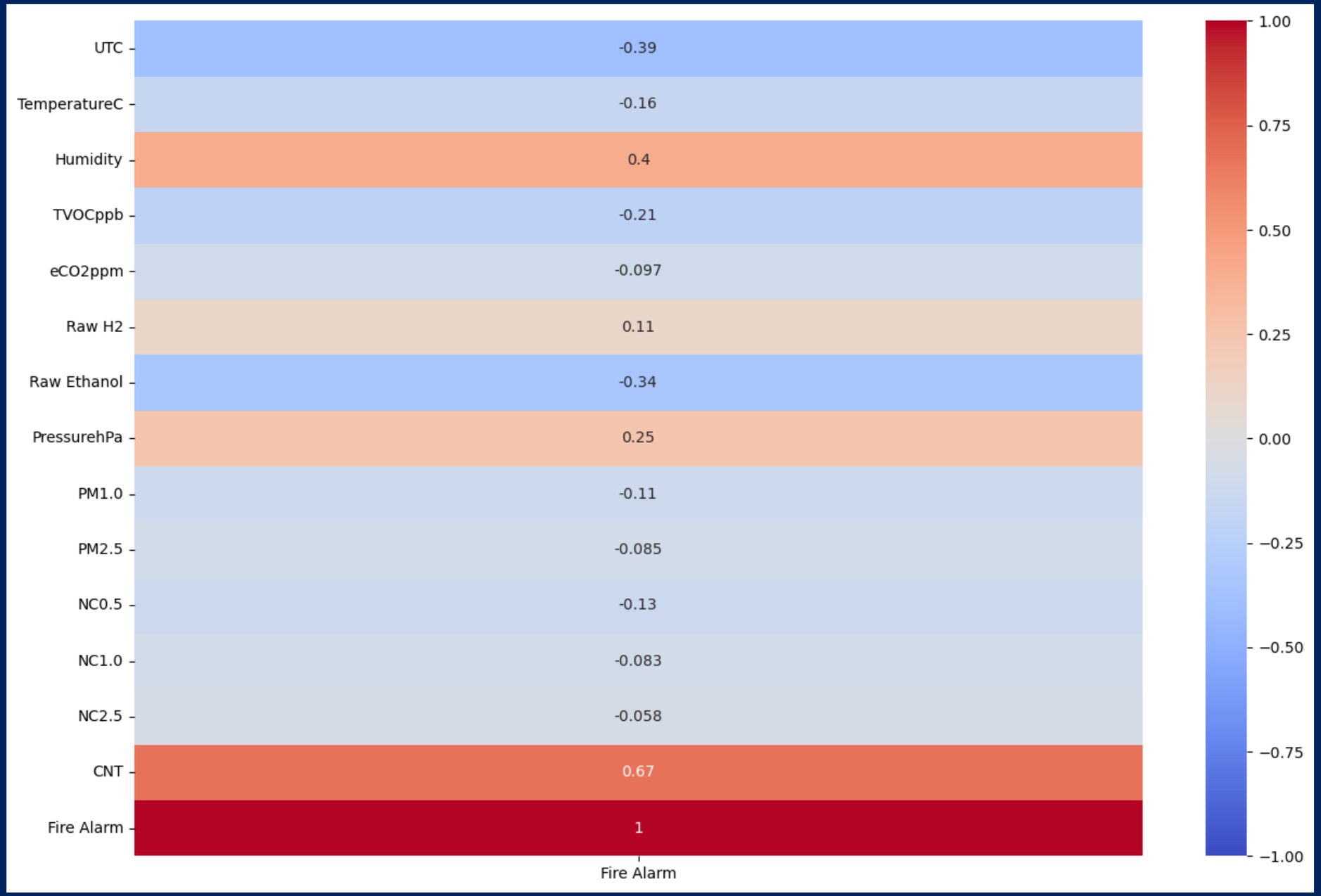
08
Model Fit Selection for Random Forest

09
Model Fit Selection for Logistic Regression

Correlation Map



Correlation Map for target feature



Model Used

- 1. Random Forest
- 2. Logistic Regression: loss dice coef loss
- 3. SVM (Support Vector Machine)
- 4. Naive Bayes
- 5. K-Nearest Neighbors (KNN):
- 6. Decision Tree:
- 7. AdaBoost:
- 8. Neural Network (Sigmoid):
- 9. XGBoost:

Metrics

01

Accuracy

02

Precision

03

Recall

04

F1 Score

road map of building model

01

Load data

02

Data Preprocessing

03

Data Visualization

04

Feature Extraction

05

Split data into training and testing sets

06

build model Architecture

07

Train Modle

08

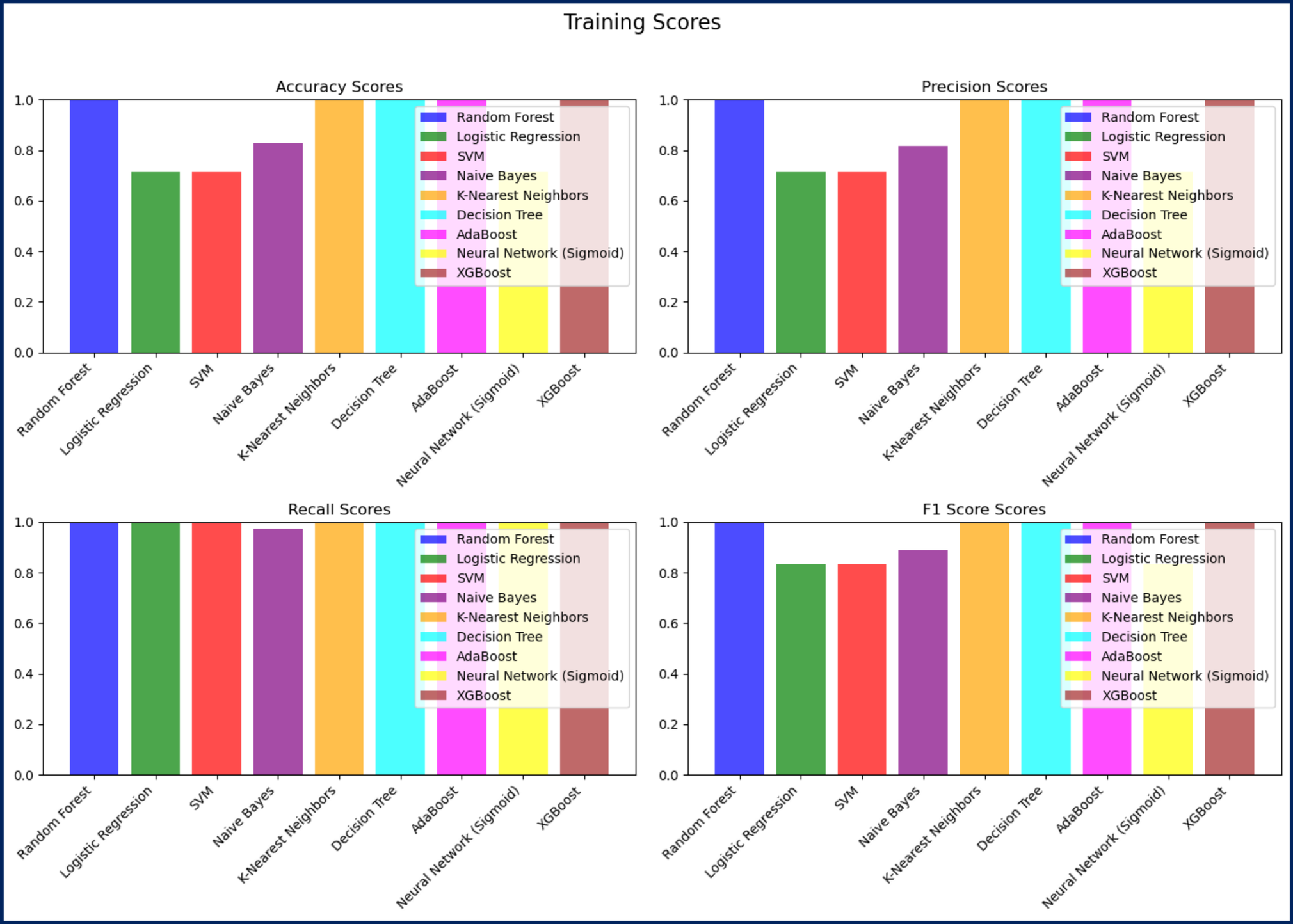
Evaluate the model with test data

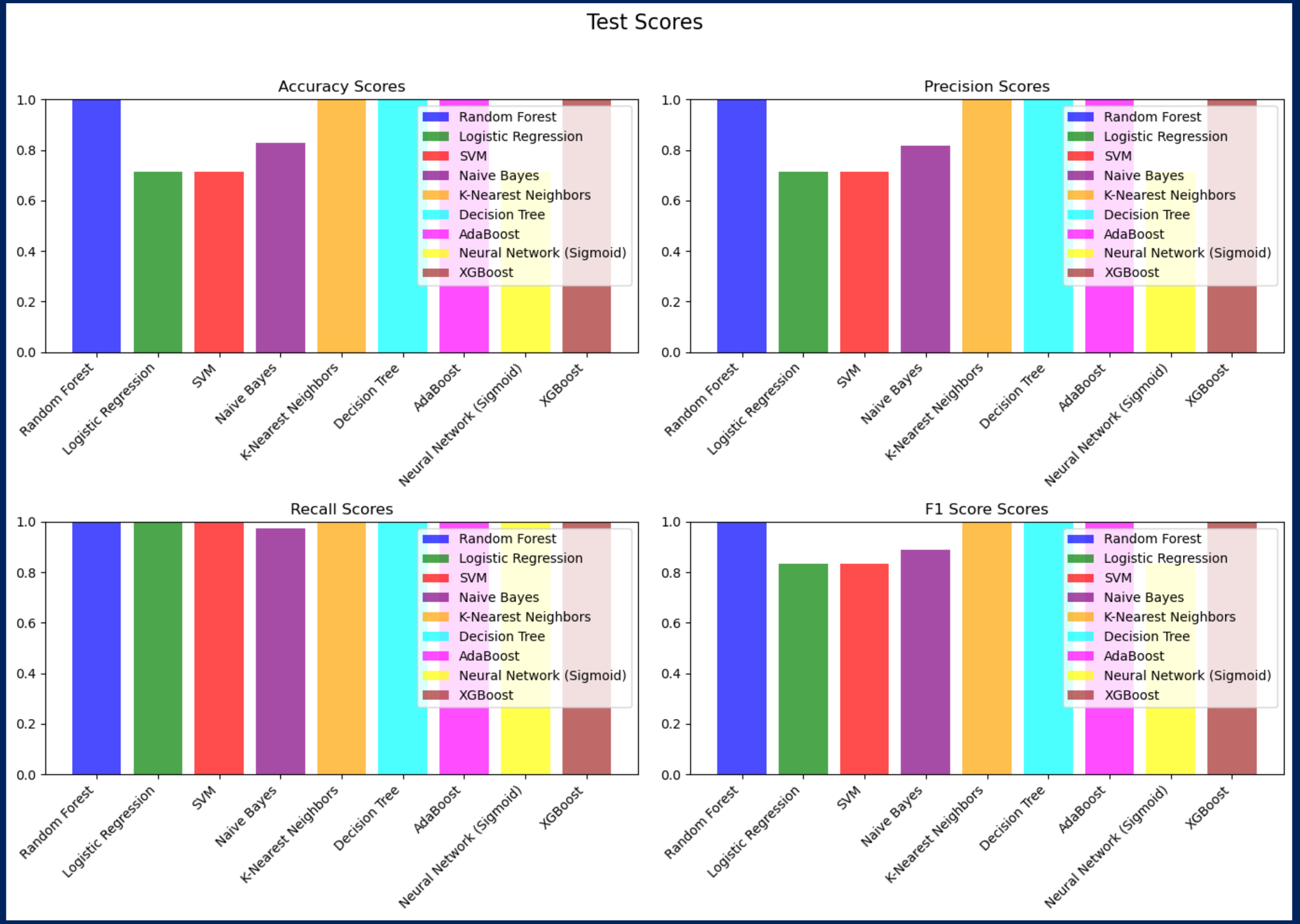
Train with all features				
	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.715	0.715	1	0.8338
Random Forest	1	1	1	1
SVM	0.715	0.715	1	0.8338
Naive Bayes	0.8263	0.8171	0.9754	0.8892
KNN	0.9997	0.9997	0.9999	0.9998
Decision Tree	1	1	1	1
AdaBoost	1	1	1	1
Neural Network Sigmoid	0.715	0.715	1	0.8338
XGBoost	1	1	1	1

Test with all features				
	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.7131	0.7131	1	0.8325
Random Forest	1	1	1	1
SVM	0.7131	0.7131	1	0.8325
Naive Bayes	0.8268	0.8184	0.973	0.8891
KNN	0.9998	0.9999	0.9999	0.9999
Decision Tree	0.9998	0.9998	1	0.9999
AdaBoost	0.9999	0.9999	1	0.9999
Neural Network Sigmoid	0.7131	0.7131	1	0.8325

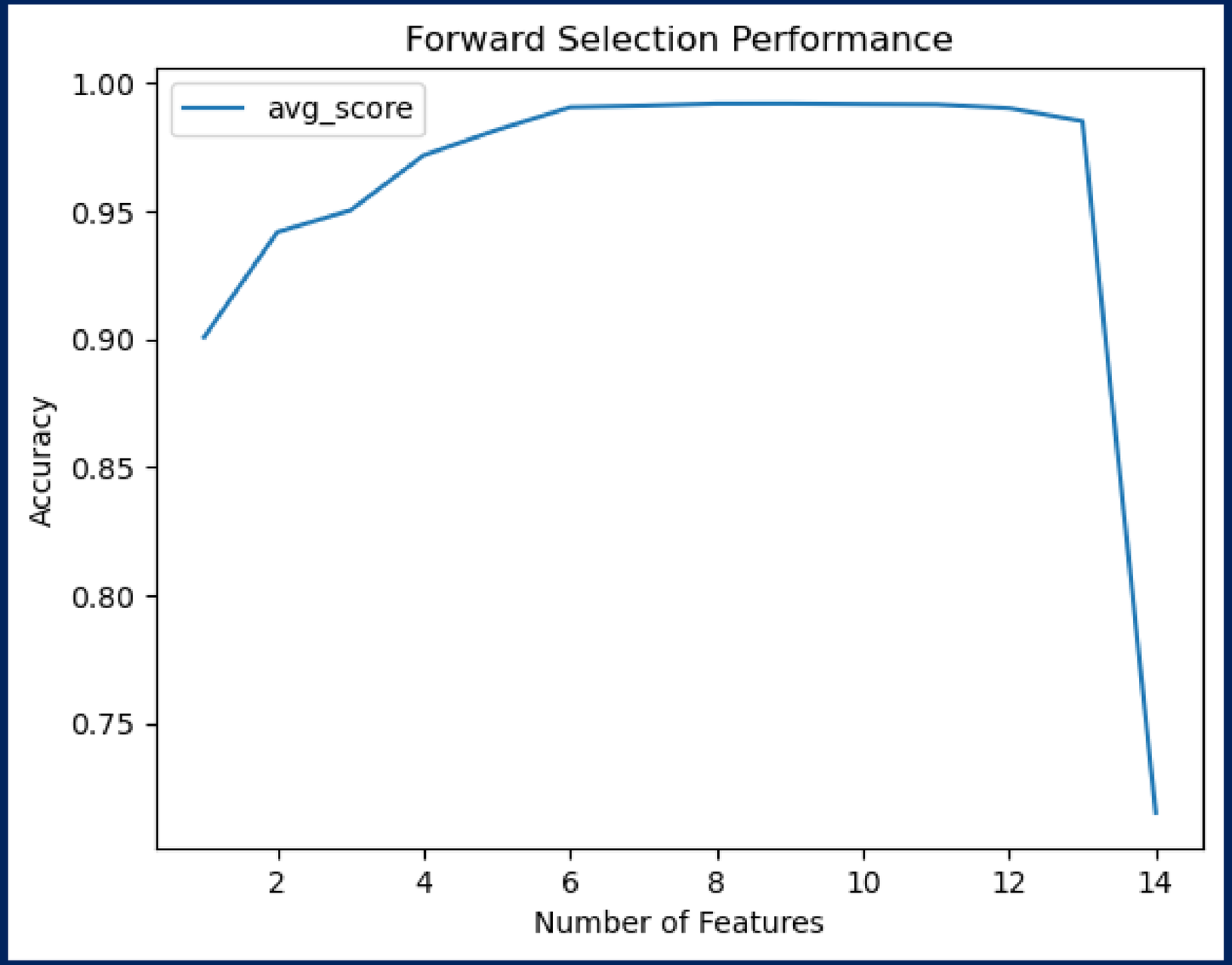
Train with 5 Features using K Selection				
	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.9817	0.9972	0.9771	0.9871
Random Forest	1	1	1	1
SVM	0.9813	0.9946	0.9791	0.9868
Naive Bayes	0.8799	0.8708	0.977	0.9208
KNN	0.9997	0.9996	1	0.9998
Decision Tree	0.9999	0.9999	1	0.9999
AdaBoost	0.9956	0.9967	0.9972	0.9969
Neural Network Sigmoid	0.9865	0.9925	0.9886	0.9906
XGBoost	1	1	1	1

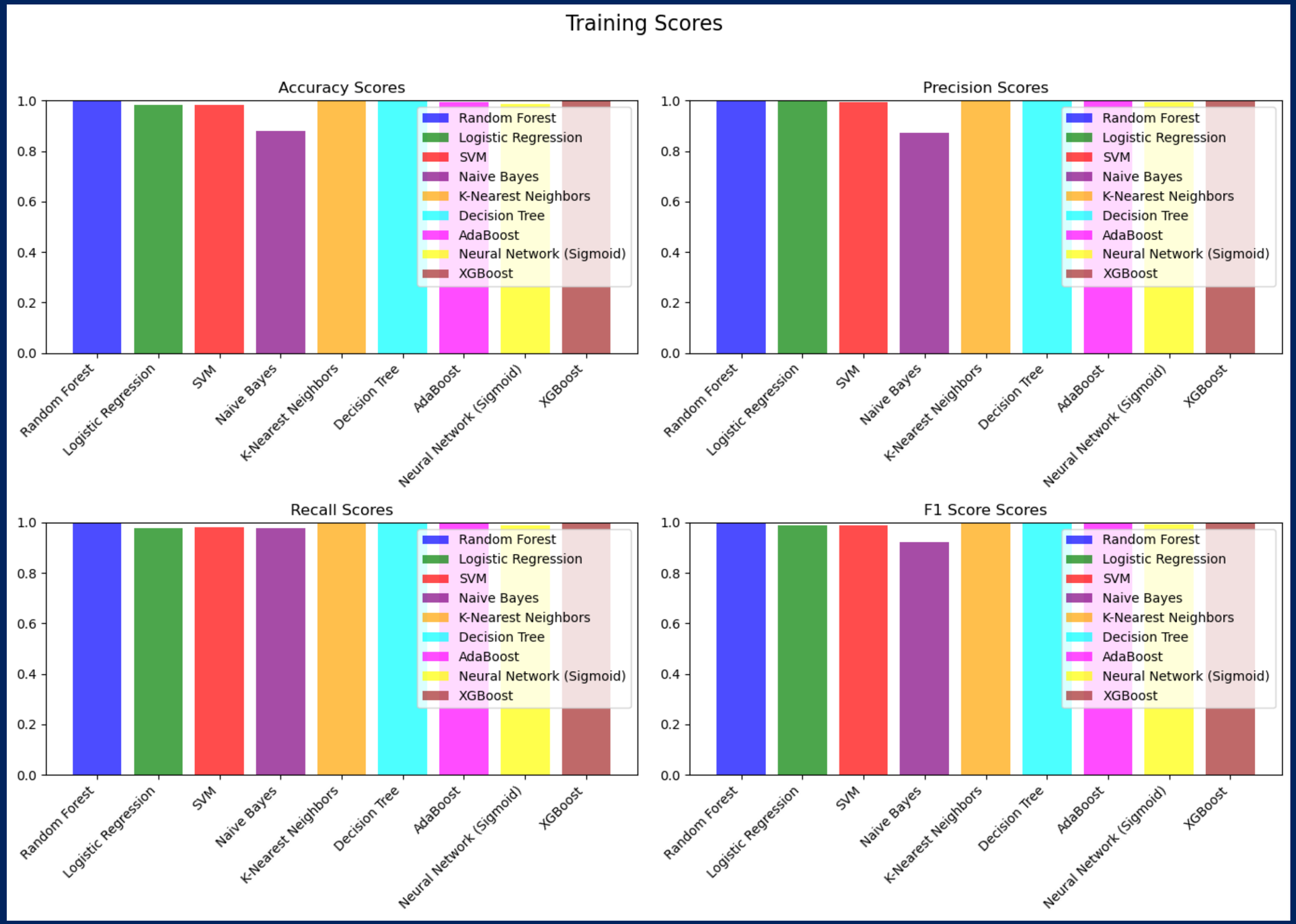
Test with 5 Features using K Selection				
	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.981	0.9973	0.976	0.9865
Random Forest	0.9999	1	0.9999	0.9999
SVM	0.9796	0.9948	0.9765	0.9855
Naive Bayes	0.8814	0.8731	0.9755	0.9215
KNN	0.9996	0.9994	1	0.9997
Decision Tree	1	1	1	1
AdaBoost	0.9954	0.9962	0.9973	0.9968
Neural Network Sigmoid	0.987	0.9929	0.9888	0.9909
XGBoost	1	1	1	1

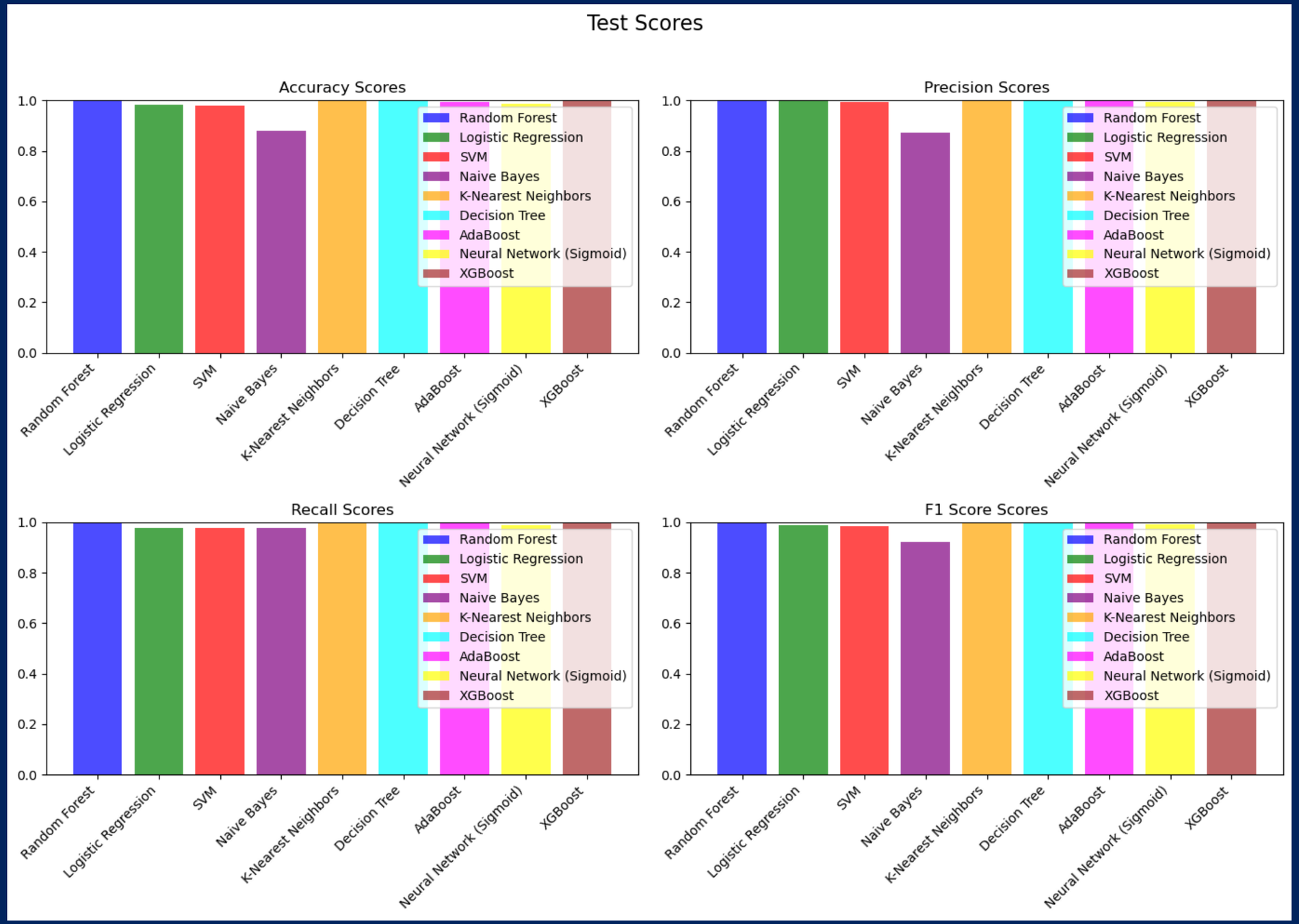

Training Scores
Accuracy Scores
Precision Scores
Recall Scores
F1 Score Scores
1.0
0.8
0.6
0.4
0.2
0.0
Random Forest
Logistic Regression
SVM
Naive Bayes
K-Nearest Neighbors
Decision Tree
AdaBoost
Neural Network (Sigmoid)
XGBoost


Test Scores
Accuracy Scores
Precision Scores
Recall Scores
F1 Score Scores
1.0
0.8
0.6
0.4
0.2
0.0
Random Forest
Logistic Regression
SVM
Naive Bayes
K-Nearest Neighbors
Decision Tree
AdaBoost
Neural Network (Sigmoid)
XGBoost

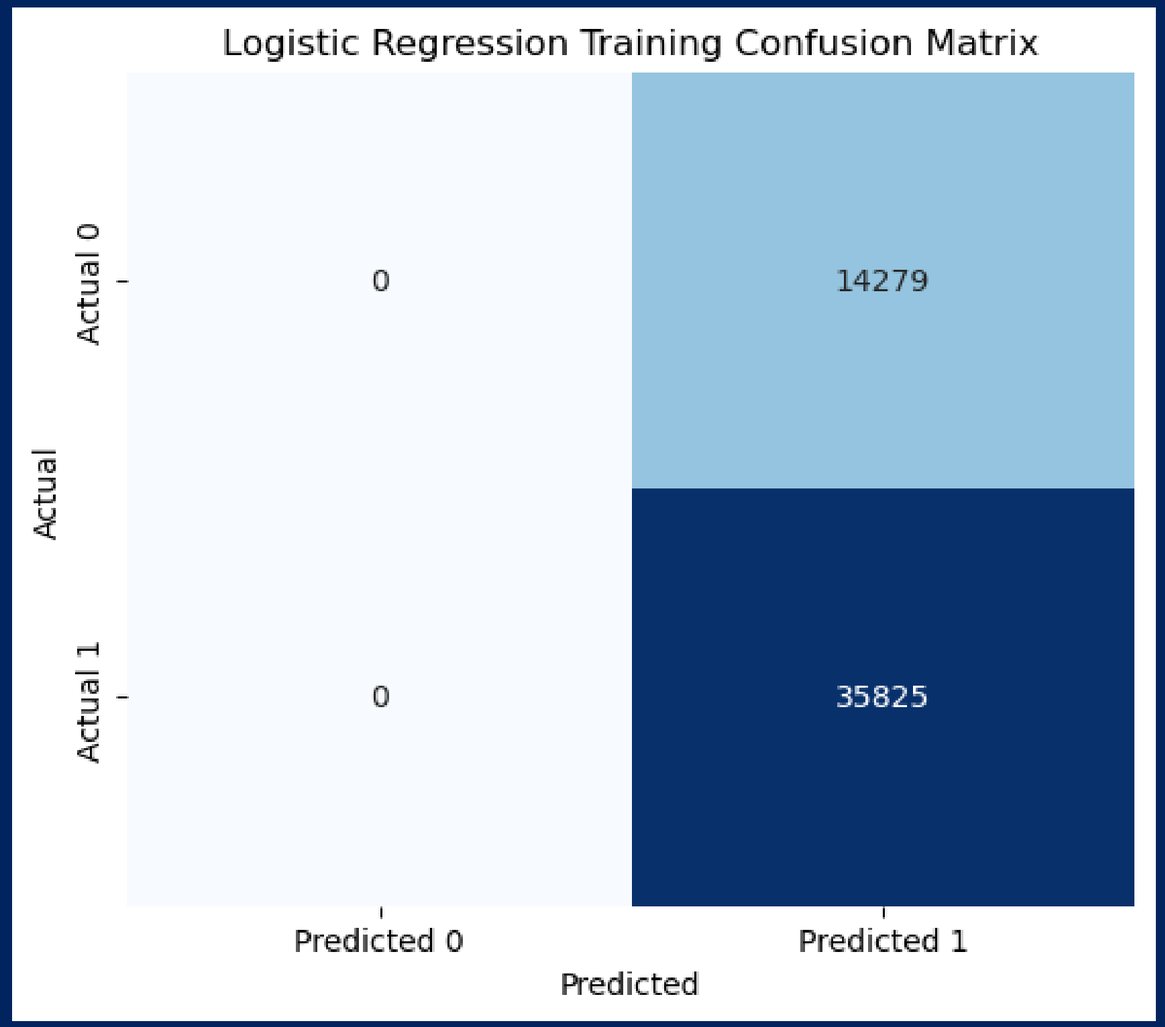
Why do we stop at K = 5?

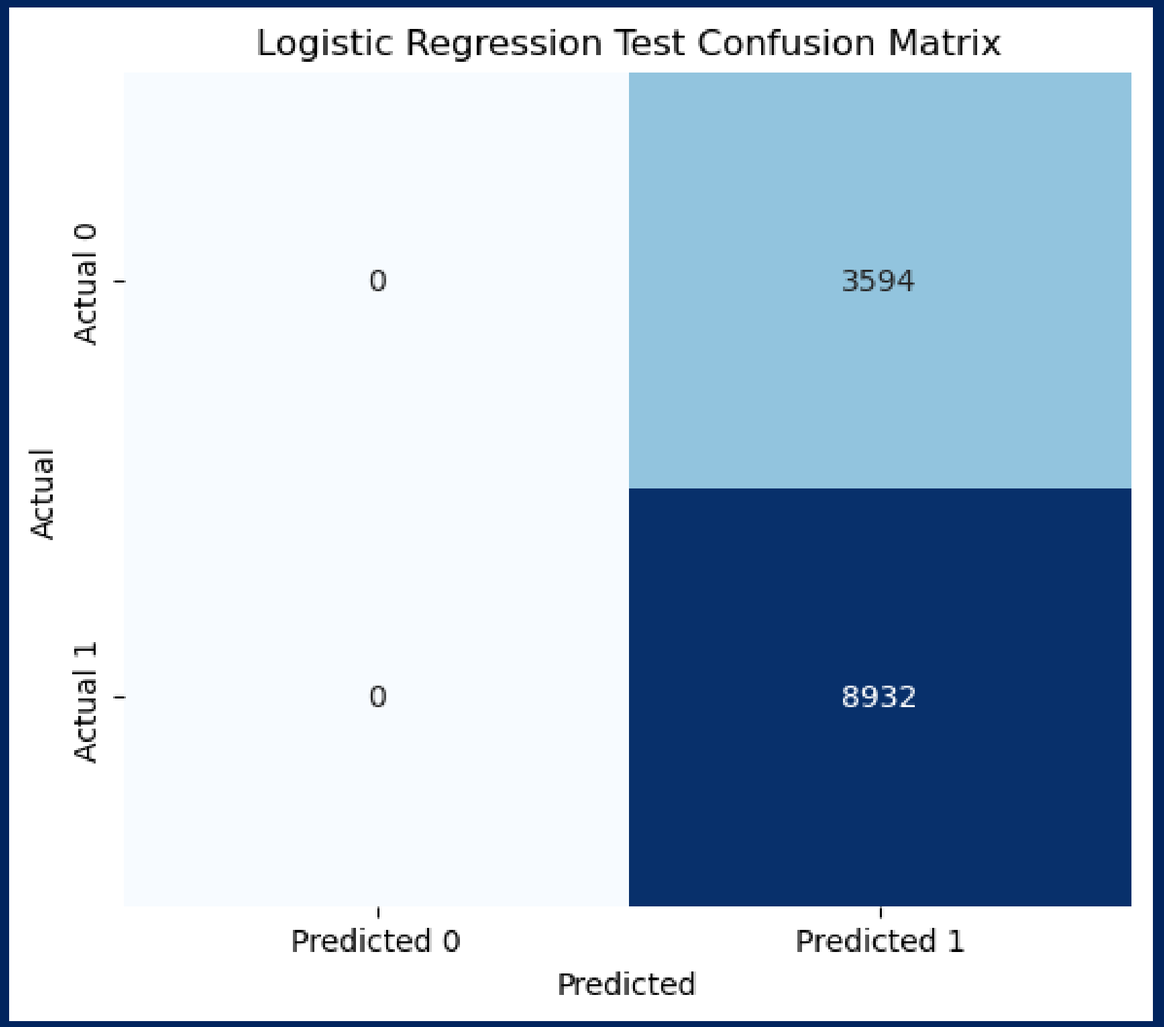



Training Scores
Accuracy Scores
Precision Scores
Recall Scores
F1 Score Scores
1.0
0.8
0.6
0.4
0.2
0.0
Random Forest
Logistic Regression
SVM
Naive Bayes
K-Nearest Neighbors
Decision Tree
AdaBoost
Neural Network (Sigmoid)
XGBoost


Test Scores
Accuracy Scores
Precision Scores
Recall Scores
F1 Score Scores
Random Forest
Logistic Regression
SVM
Naive Bayes
K-Nearest Neighbors
Decision Tree
AdaBoost
Neural Network (Sigmoid)
XGBoost
0.0
0.2
0.4
0.6
0.8
1.0

Confusion matrix for all features





Confusion matrix for best 5 features