TITANIC SURVIVAL PREDICTION

By

ELAKYA K 21ITR026

KAVYA S 21ITR052

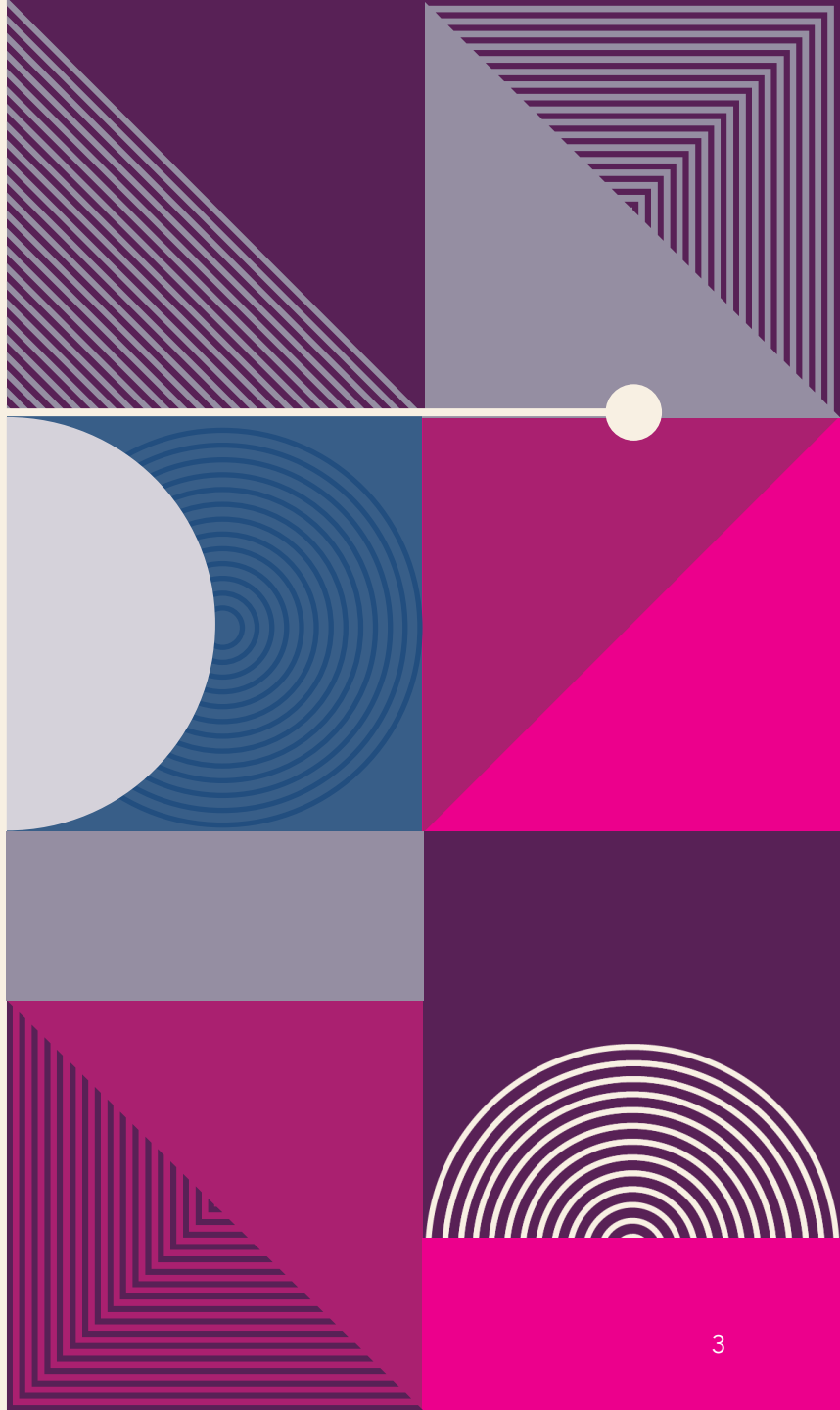
MAMTHA B 21ITR058

OBJECTIVE

To utilize supervised learning algorithms to predict the survival outcomes of Titanic passengers based on available demographic and ticket information. Through the exploration of machine learning techniques, we seek to develop accurate predictive models and evaluate their performance.

PROBLEM STATEMENT:

The project focuses on predicting the survival probability of passengers aboard the Titanic using machine learning algorithms. By analyzing demographic and ticket information, the goal is to develop models capable of accurately predicting whether a passenger survived or not. This endeavor aims to uncover patterns and factors influencing survival rates, thereby contributing to historical data analysis and predictive modeling.



DATASET DESCRIPTION

The dataset comprises information on passengers aboard the Titanic, including attributes such as age, gender, ticket class, fare, and survival status. It contains both numerical and categorical features, offering insights into the demographics and ticketing details of passengers. This dataset serves as the basis for developing predictive models to infer survival outcomes and understand the factors contributing to passenger survival rates.

ALGORITHMS USED

SUPERVISED LEARNING ALGORITHMS

1. SUPPORT VECTOR MACHINE(SVM)
2. NAÏVE BAYES
3. DECISION TREE
4. LOGISTIC REGRESSION
5. KNN CLASSIFIERS
6. GRADIENT BOOST
7. RANDOM FOREST
8. LINEAR REGRESSION
9. KNN REGRESSOR

VISUALIZATION

1. COUNTER PLOT
2. DISTRIBUTION PLOT
3. PIE CHART
4. SCATTER PLOT
5. VIOLIN PLOT
6. BOX PLOT
7. SWARM PLOT
8. BAR PLOT
9. HISTOGRAM

DIMENSIONALITY REDUCTION

- PRINCIPAL COMPONENT ANALYSIS

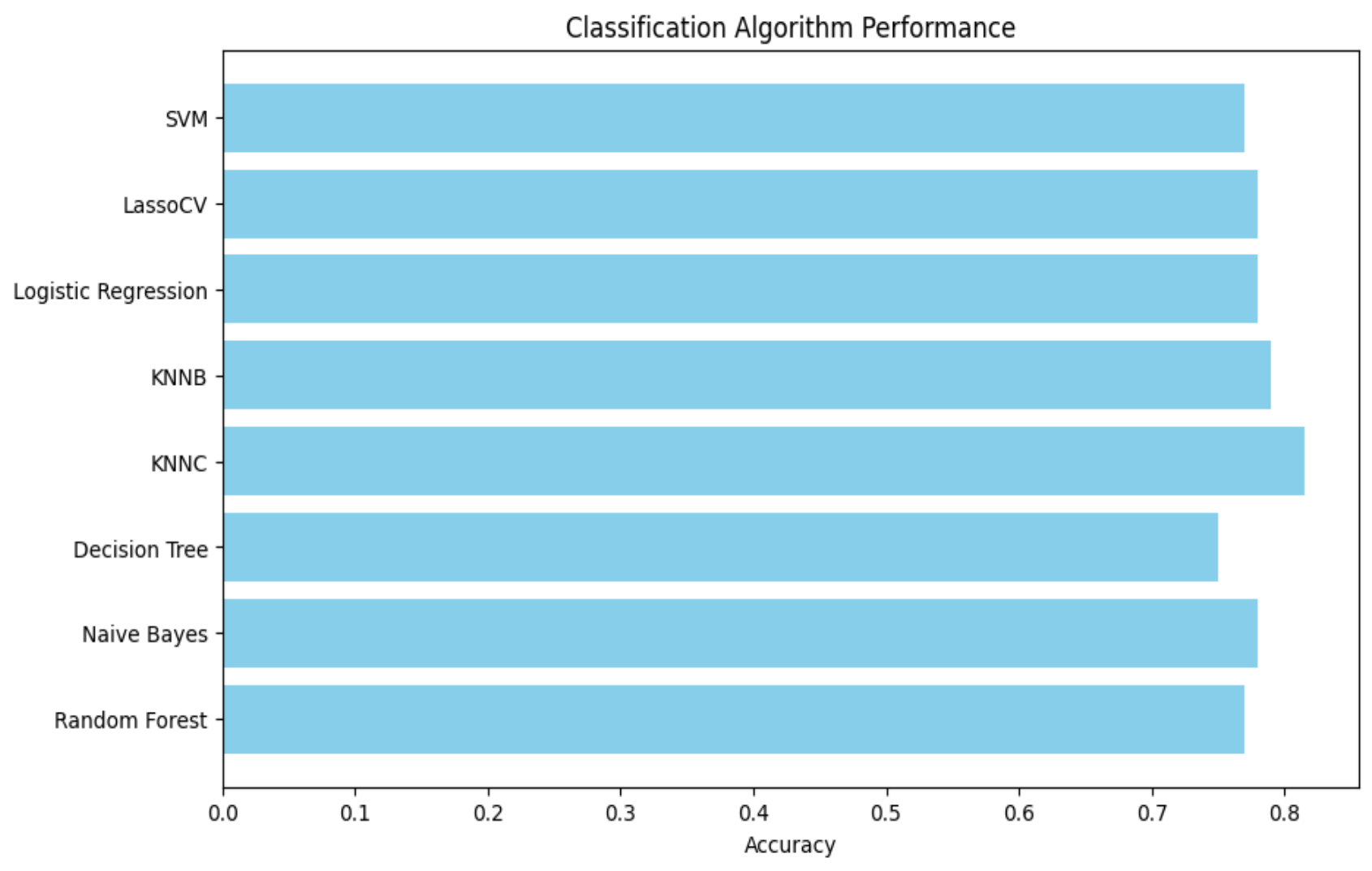
CROSS VALIDATION

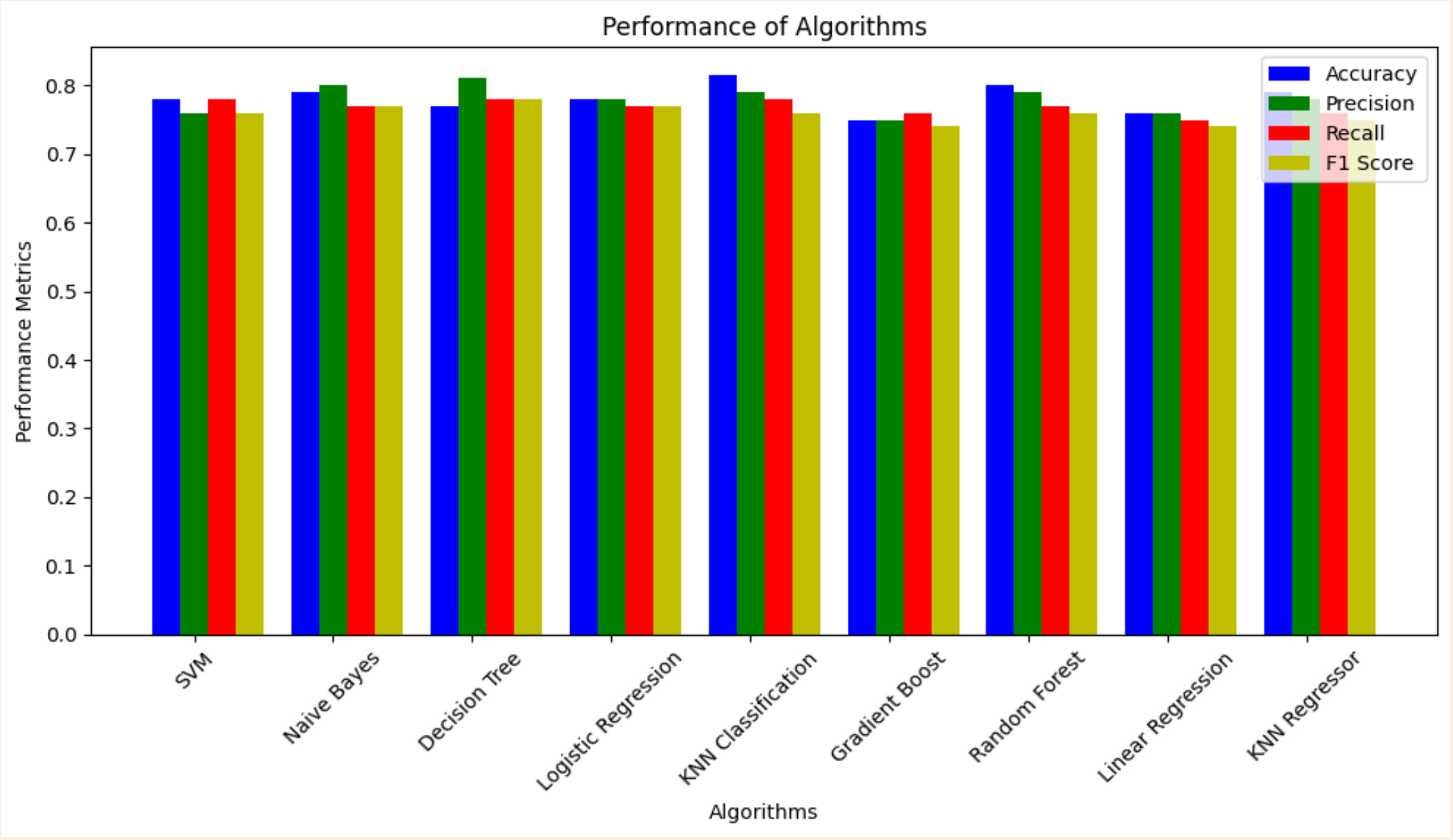
- LASSOCV

RESULT

Algorithms / Performance metrics	Accuracy	MSE
SVM	0.775	-
Naive Bayes	0.782	-
Decision Tree	0.796	-
Logistic Regression	0.787	-
KNN Classifier	0.805	-
Gradient Boost	-	0.12
Random Forest	0.776	-
Linear Regression	-	0.21
KNN Regressor	-	0.21

GRAPH AND INTERPOLATION




Performance of Algorithms
Accuracy
Precision
Recall
F1 Score
Performance Metrics
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
SVM
Naive Bayes
Decision Tree
Logistic Regression
KNN Classification
Gradient Boost
Random Forest
Linear Regression
KNN Regressor
Algorithms

COLAB LINK

https://colab.research.google.com/drive/1VEH7Ea9saopxpieUMVg3pswpBGbpdYhX?usp=sharing

CONCLUSION

In conclusion, after evaluating multiple supervised learning algorithms on the Titanic dataset, we found that KNN Classifier achieved the highest accuracy of 0.815(82%) among all algorithms tested. While accuracy is an important metric, it's essential to consider other factors such as interpretability, computational efficiency, and scalability when selecting the best algorithm for a specific task. Overall, KNN Classifier showed promising performance and may be a suitable choice for further exploration or deployment in real-world applications.

THANK YOU!!