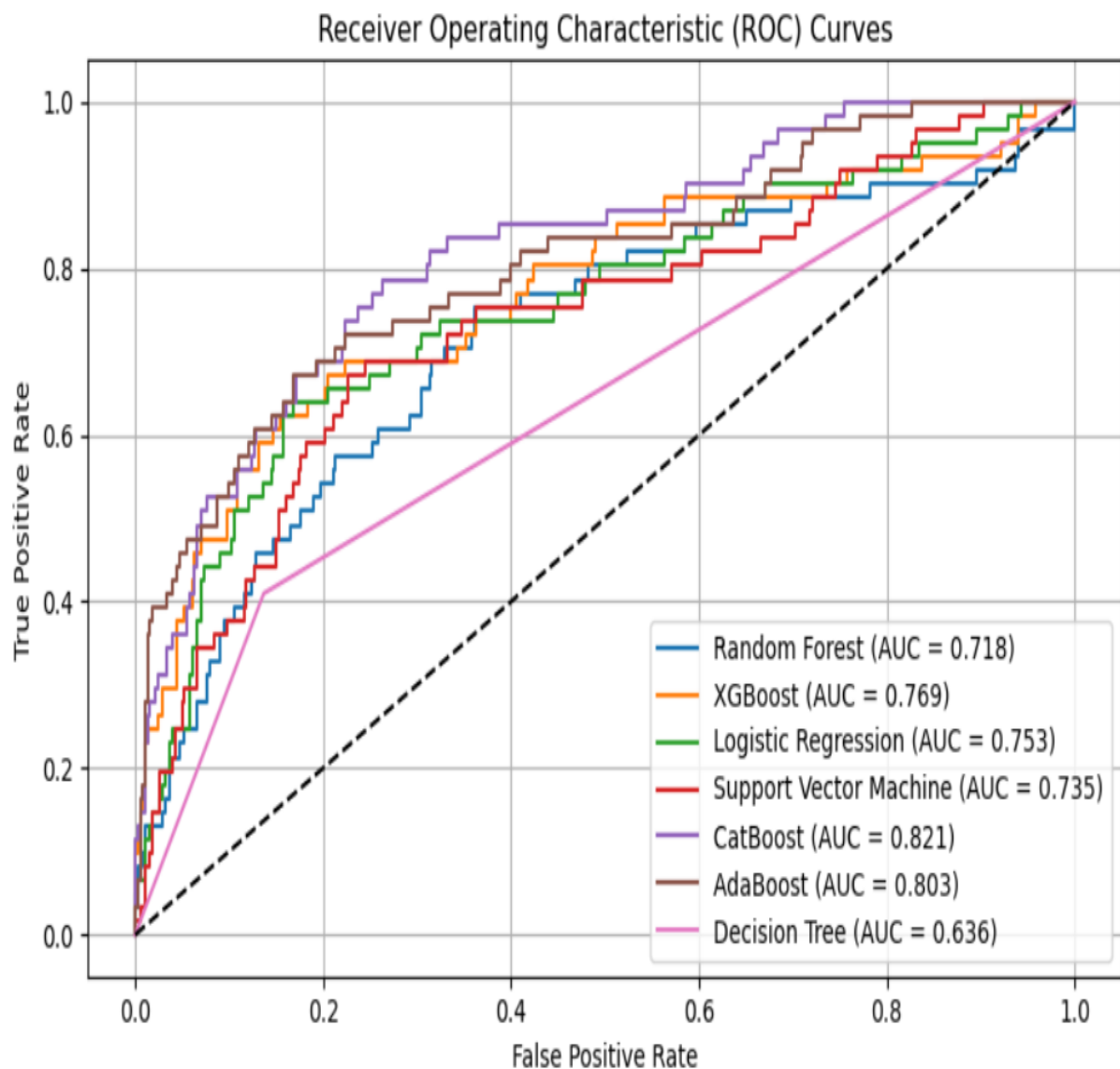


## INTRODUCTION

Employee attrition refers to an employee's voluntary or involuntary resignation from a workforce. Organizations spend many resources in hiring talented employees and training them. Every employee is critical to a company's success. Goal was to predict employee attrition and identify the factors contributing to an employee leaving a workforce. Trained various classification models on the dataset and assessed their performance using different metrics such as accuracy, precision, recall and F1 Score. Also analyzed the dataset to identify key factors contributing to an employee leaving a workforce. This project will assist organizations in gaining fresh insights into what drives attrition and thus enhance retention rate.

## RESULTS AND ANALYSIS



---

RANDOM FOREST	Accuracy_score: 0.794 Precision:0.330 Recall0.475
XGBOOST	Accuracy_score: 0.864 Precision:0.511 Recall0.393
LOGISTIC REGRESSION	Accuracy_score: 0.723 Precision:0.287 Recall0.672
SUPPORT VECTOR MACHINE	Accuracy_score: 0.719 Precision:0.286 Recall0.689
CATBOOST	Accuracy_score: 0.884 Precision:0.692 Recall0.295
ADABOOST	Accuracy_score: 0.862 Precision:0.500 Recall0.492
DECISION TREE	Accuracy_score: 0.800 Precision:0.325 Recall0.410

1. Since the data set is highly imbalanced with the negative class having 84% of the total instances, accuracy is not a good measure for model comparison as any naive classifier would also yield 84% accuracy.
2. Precision defines how correctly is the minor class being detected and recall states how much is the minor class being detected.
3. In this problem statement our aim is to decrease False Negative.
4. The higher the precision and lower recall values in baseline condition conclude that the minority class is not detected much but whenever detected the degree of trust is high.
5. on contrary, the higher recall and lower precision value conclude that even though the minority class is being detected but degree of trust in that detection is very low.
6. The models specialized for imbalanced class scenarios showed promising outcomes in all metrics.
7. Finally it can be conclude that CatBoost,XGBoost,AdaBoost,RandomForest and LogisticRegression are the best models among the above mentioned.