Group\_09

Group Members:

Vamsidhar Mothe (1002083347)

Sreeja Sree Kukkala (1002091624)

Tejaswi Kundur (1002120036)

Sadvik Mente (1002120393)

Arpith Rameshbhai Gabani (1001881217)

Instructor Name: Ramakrishna Koganti

Course Number: CSE-5301-009

Date: 30 April 2023

**Customer churn prediction model**

# 1. Introduction

Businesses use churn prediction to anticipate the likelihood of a client abandoning services or ceasing purchases from the company. Customer data such as purchase history, demographic information, and behaviour patterns are analysed by churn prediction algorithms to identify customers who are likely to churn. The fundamental purpose of churn prediction is to assist organizations in retaining customers by detecting possible churners early on and taking proactive actions to keep them. Businesses may keep their customers happy and loyal by recognizing the reasons for customer churn and making improvements to their products or services, pricing strategy, or customer service.

# 2. Problem definition

A churn prediction model's issue definition is to create a machine learning model that can accurately forecast which consumers are at risk of churning or leaving a firm. The idea is to identify consumers who are likely to leave soon so that firms can take proactive steps to keep them. The problem of churn prediction models necessitates a thorough understanding of machine learning algorithms, feature engineering, and data analysis methodologies. The ultimate goal is to create a model that accurately predicts customer turnover, which may then be used to improve customer retention, reduce churn, and generate revenue for the company.

# 3. Predictive Model Selection

To test the proposed churn prediction model, measures such as accuracy, precision, recall, f-measure, and receiver operating characteristics (ROC) area are used. For model selection, we use many machine learning techniques such as decision trees, classifiers utilizing Random Forest, and Linear Regression (LR) to discover which algorithm provides the best accuracy. Random forest is a powerful ensemble technique that can handle large datasets and reduce overfitting. Decision tree is a simple and easy-to-understand algorithm, logistic regression is appropriate for modelling the connection between variables, and decision tree is a simple and easy-to-understand algorithm. The algorithm chosen is determined by the churn prediction problem's specific needs, the quantity and complexity of the dataset, and the required level of accuracy and interpretability.

# 4. Analysis

Machine learning techniques utilized in churn prediction include decision trees, logistic regression, and random forest. We trained these models on a churn prediction dataset and compared their performance using accuracy, precision, recall, and F1-score criteria. The decision tree attained an accuracy of 72%, a precision of 82%, a recall of 81%, and an F1-score of 81%. Logistic regression obtained 79% accuracy, 84% precision, 90% recall, and an F1-score of 87%. Random forest achieved 78% accuracy, with 82% precision, 89% recall, and an F1-score of 86%. It's critical to remember that when assessing a machine learning algorithm's performance, accuracy is not always the only or even the most significant statistic to consider. Other measures like precision, recall, and F1-score may be more pertinent depending on the challenge and business requirements. Additionally, it's possible that the characteristics of the dataset will affect the algorithms' performance. To obtain a more thorough understanding of the performance of various algorithms, it is advised to carry out several evaluations and compare their results using a variety of measures.

# 5. Conclusion

Churn prediction is a critical challenge for CRM (Customer Relation Management) in the current competitive telecom market to retain valuable customers by recognizing comparable groups of consumers and offering competitive offers/services to the relevant groups. To keep customers and address CRM issues, researchers in this field have been examining the primary causes of churn. In this study, a customer churn model for data analytics is offered and verified using common assessment indicators. The observed findings demonstrate that our suggested churn model outperformed other machine learning approaches.

# 6. Recommendations

The ability to predict customer attrition is a crucial tool for telecom firms to lower churn rates and increase customer retention. Companies can create prediction models that identify at-risk consumers and take proactive efforts to retain them by studying customer behaviour data. Improved customer service, tailored offers, and individualized marketing efforts are a few examples of this. Companies may lower their turnover rates and boost customer happiness, which can enhance sales and profitability, by taking action before a customer leaves.

# Appendix

# Python Code:

Python Code is attached in a submitted zip file.

# Additional pictures:

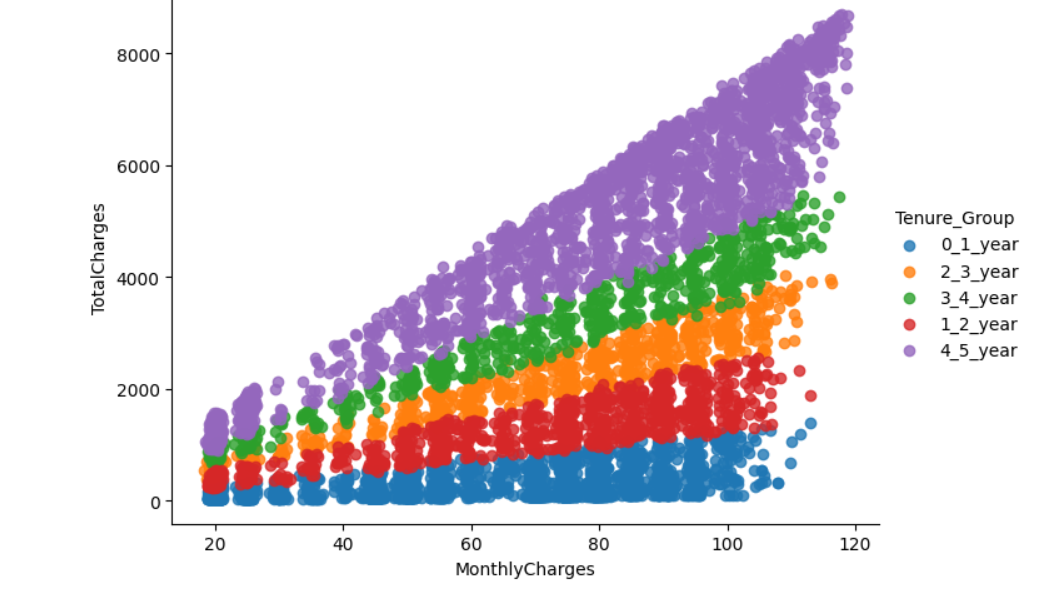


Fig. 1. Scatter plot for Total charges vs Monthly charges of five years.

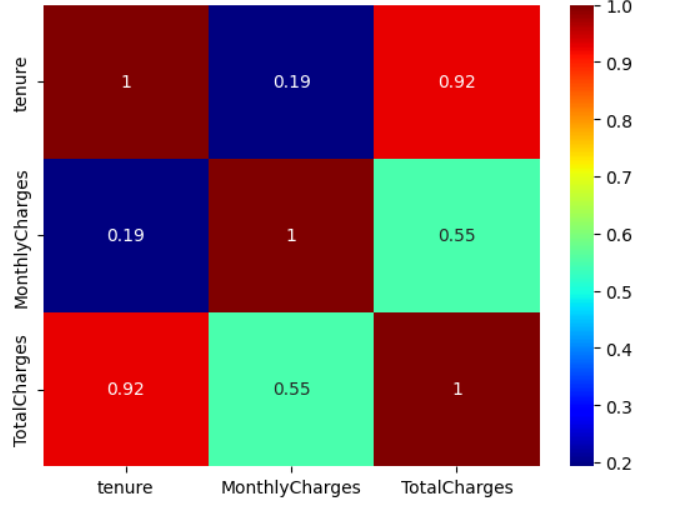


Fig. 2. Heat Map.

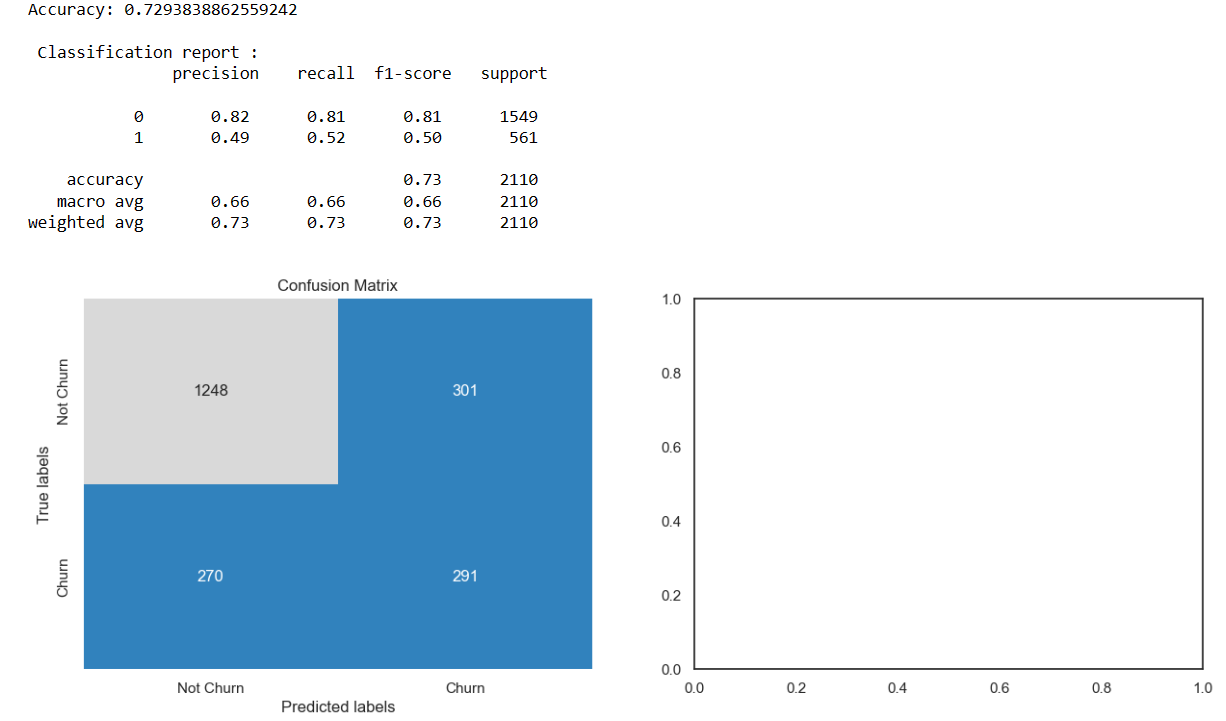


Fig. 3. Classification report of Decision tree classifier.

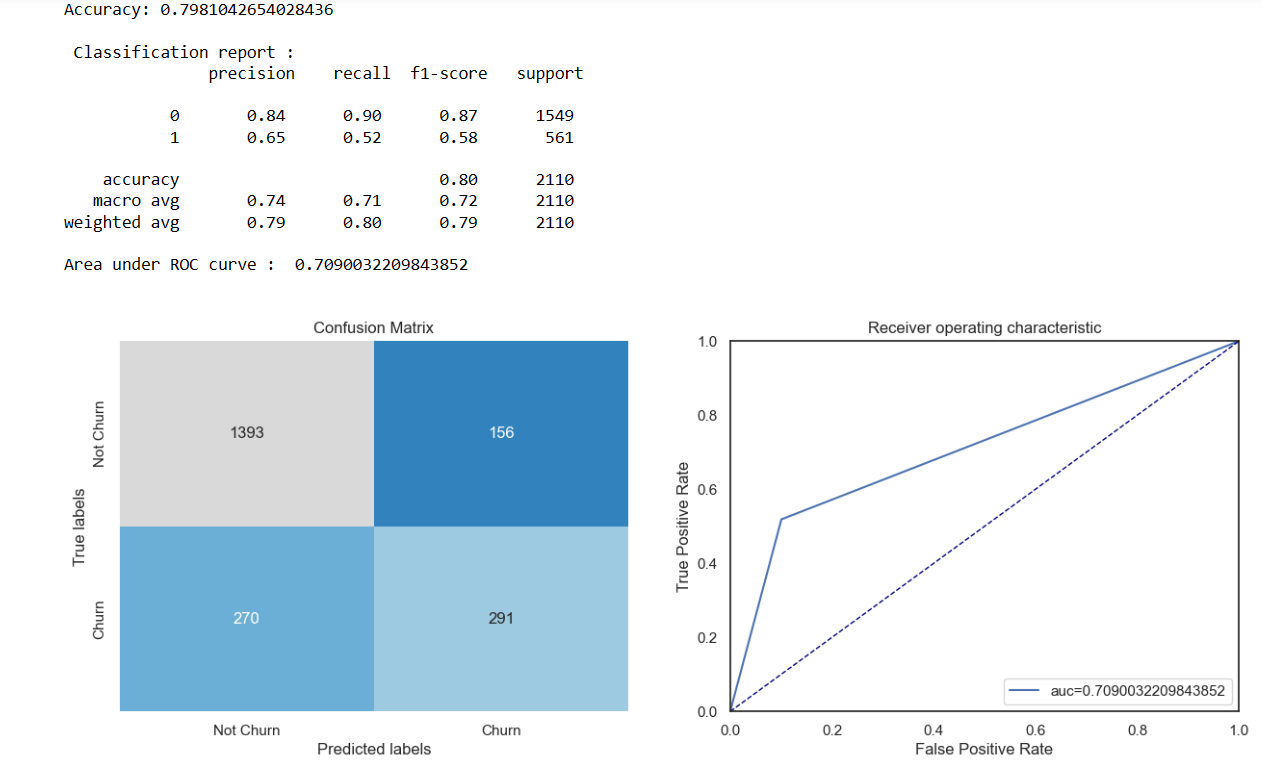


Fig. 4. Classification report of Logistic regression.

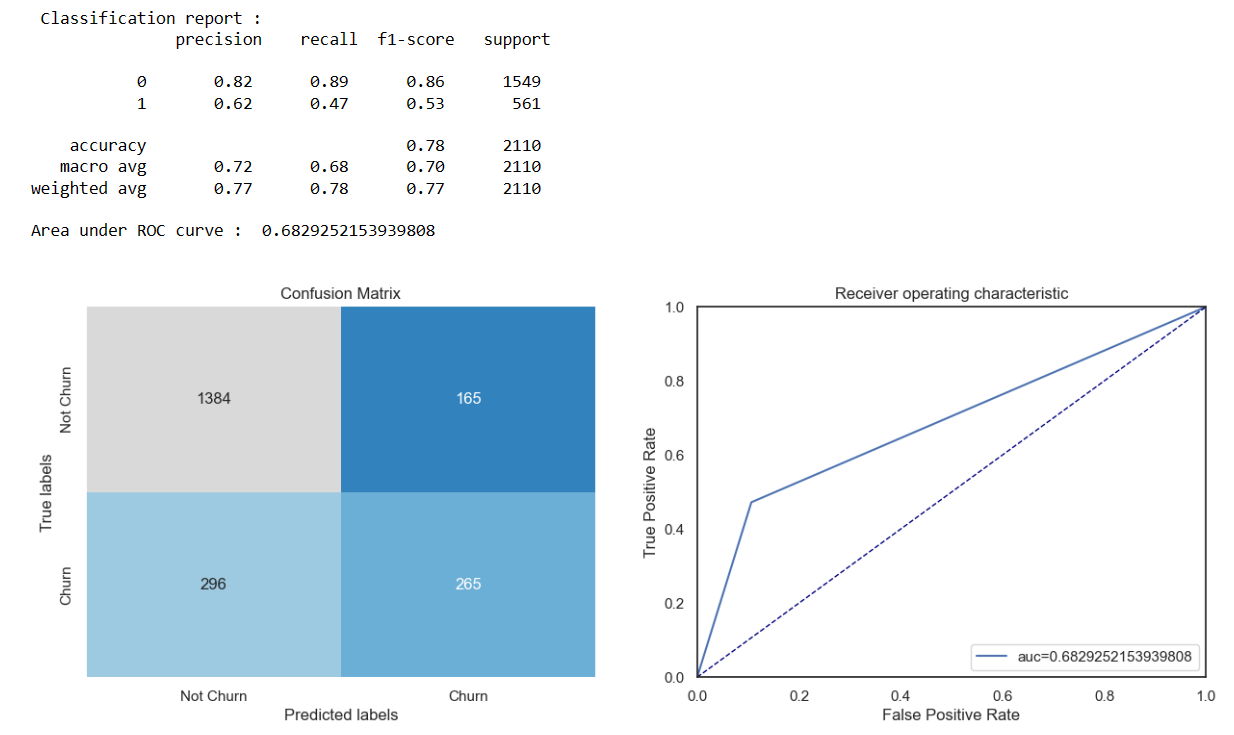


Fig. 5. Classification report of Random Forest.