TELECOM CHURN CASE STUDY

SUBMITTED BY
DINESH KUMAR (DS C53)

PROBLEM STATEMENT

- Within the dynamic landscape of the telecommunications industry, customers enjoy the privilege of selecting the numerous service providers and retaining the flexibility to switch between operators as they see fit.
- This environment of robust competition has resulted in the telecommunications sector witnessing an annual churn rate ranging from 15% to 20%.
- In this case study, we will predict the churn for pre-paid customers who are predominant in the Indian and South Asian market.

BUSINESS OBJECTIVE

The business objective is to predict the churn in the last (i.e. the ninth) month using the data (features) from the first three months. To do this task well, understanding the typical customer behavior during churn will be helpful.

EDA AND MODEL DEVELOPMENT

- Our journey begins with a comprehensive exploration of the dataset, addressing missing data through imputation, and eliminating columns that do not contribute meaningful insights.
- We then identify a subset of high-value customers, those who've recharged an amount equal to or exceeding the 70th percentile of the average recharge amount in the first two months.
- Additionally, we engineer six new features to enhance our dataset, providing richer information for model training.
- After data preparation, we proceed with model development, employing various machine learning algorithms. We refine these models through hyperparameter tuning to optimize their performance.

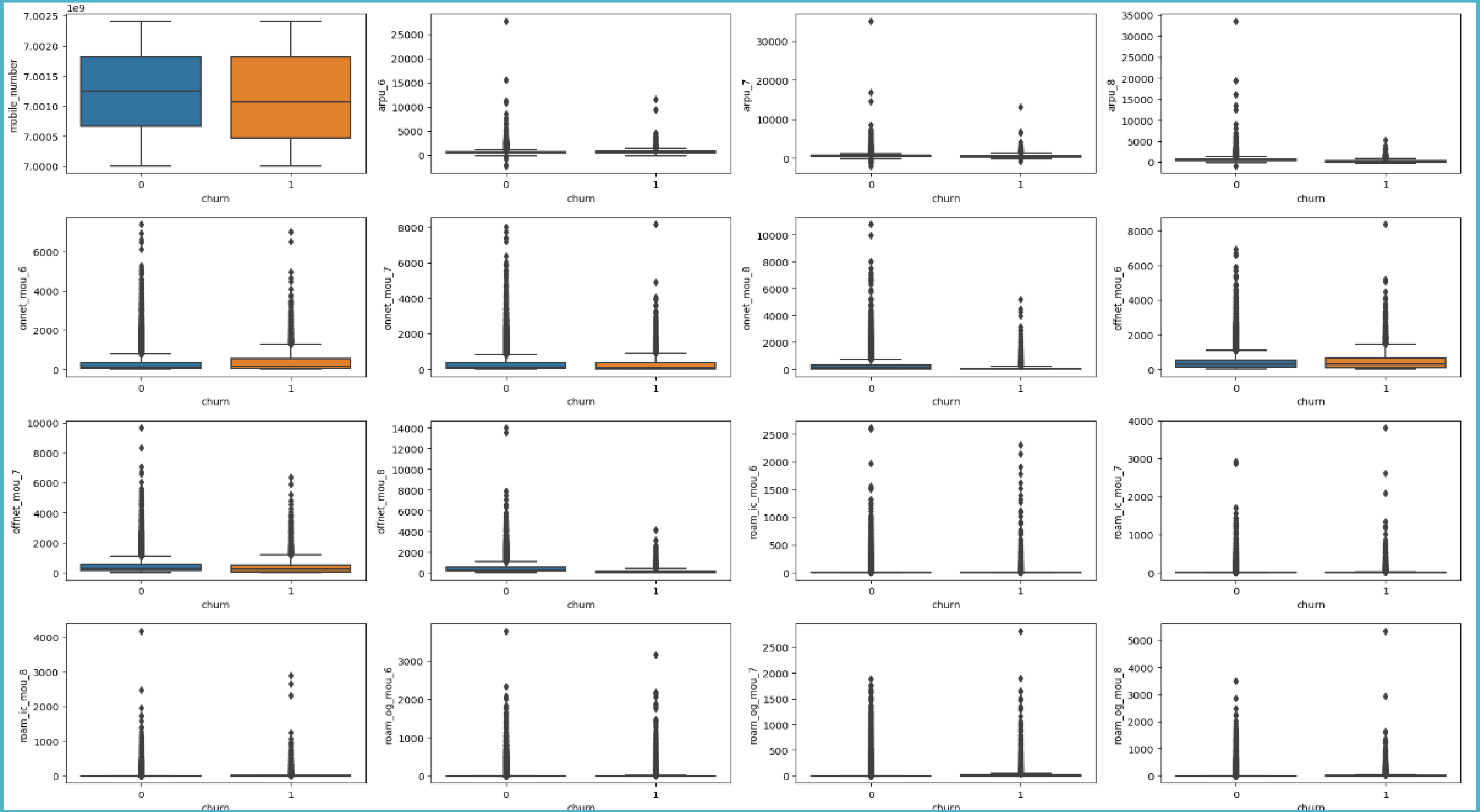
SOURCING AND UNDERSTANDING DATA

- Importing dataset
- Converting dataset into a DataFrame
- Understanding the data dictionary
- Inspecting Data for EDA
- Performing Data Cleaning

EXPLORATORY DATA ANALYSIS (EDA)

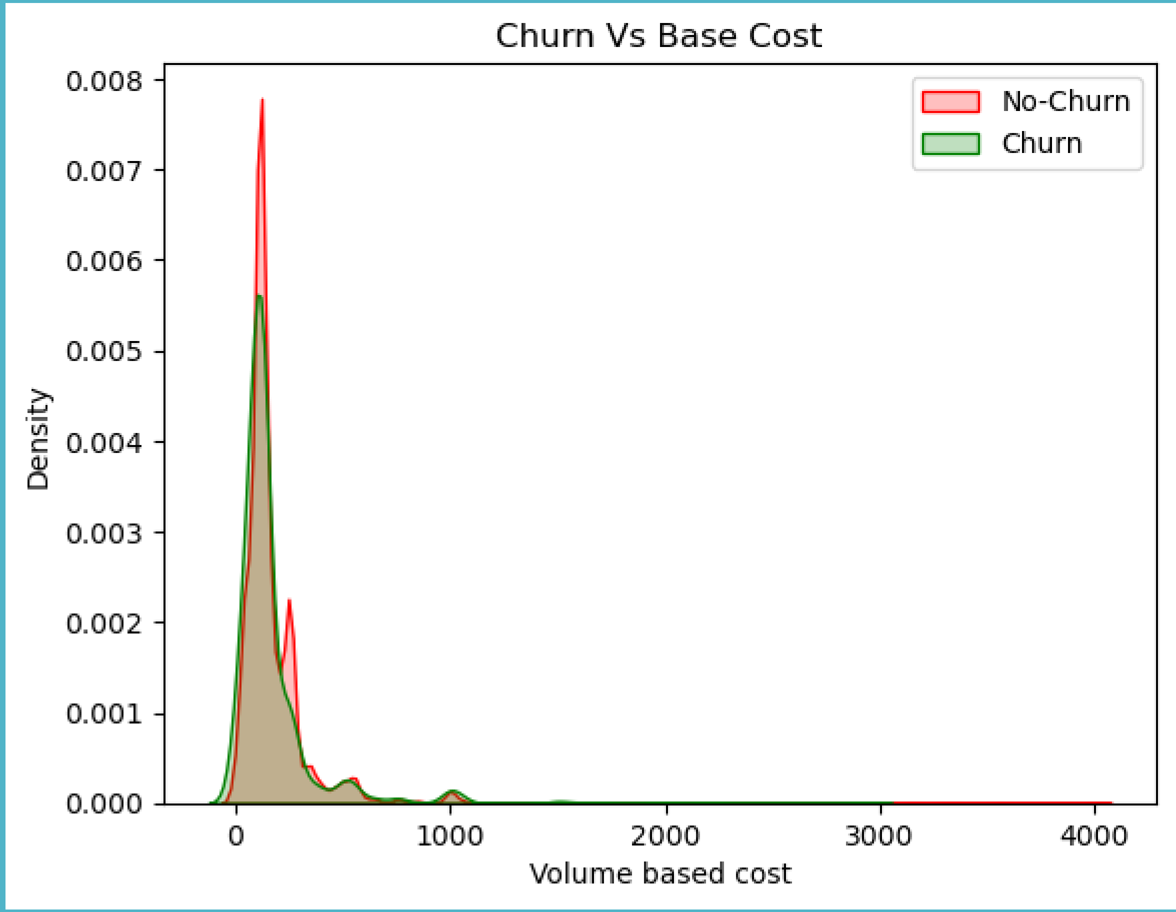
EXPLORATORY DATA ANALYSIS (EDA)

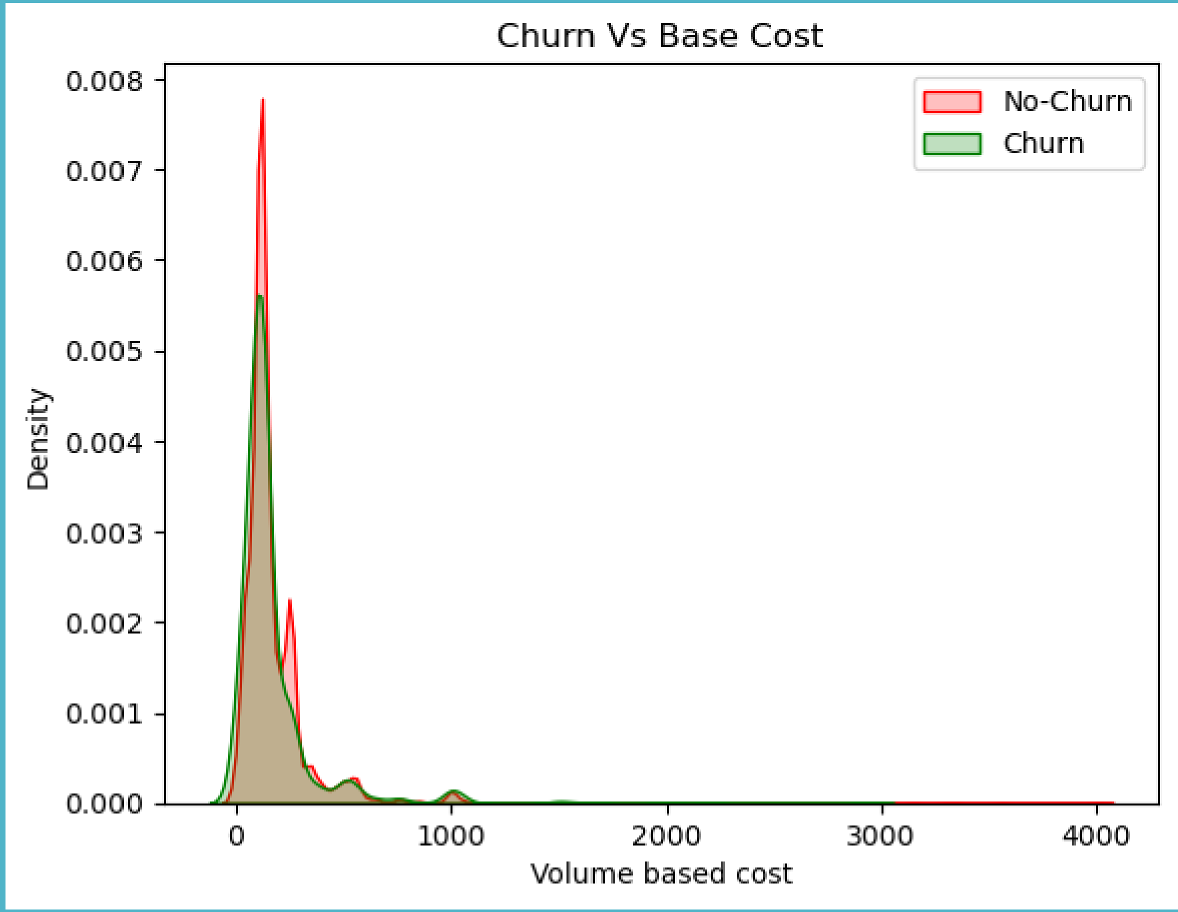
OUTLIER ANALYSIS USING BOX PLOT



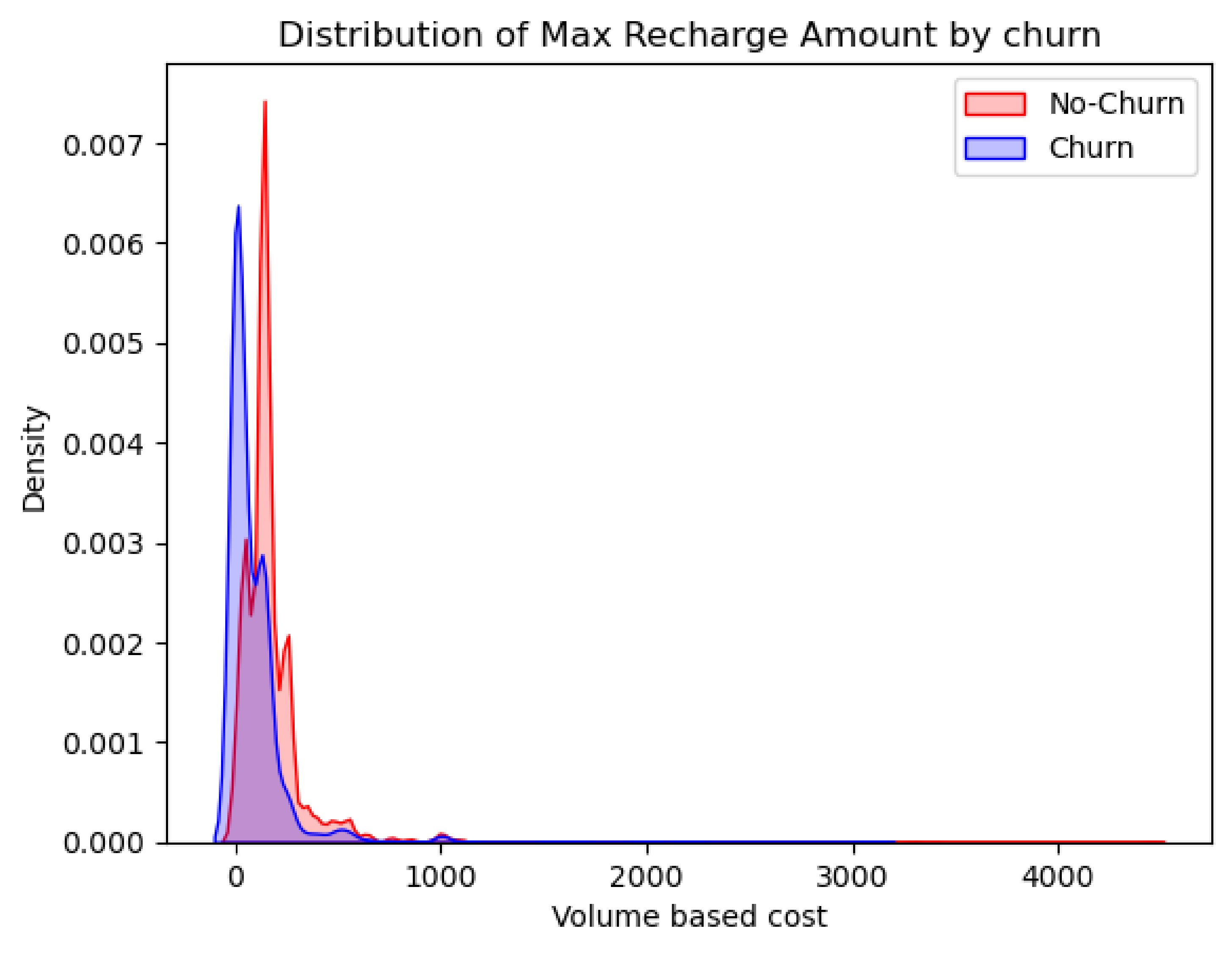
EXPLORATORY DATA ANALYSIS (EDA)

CHURN VS NON-CHURN DISTRIBUTION

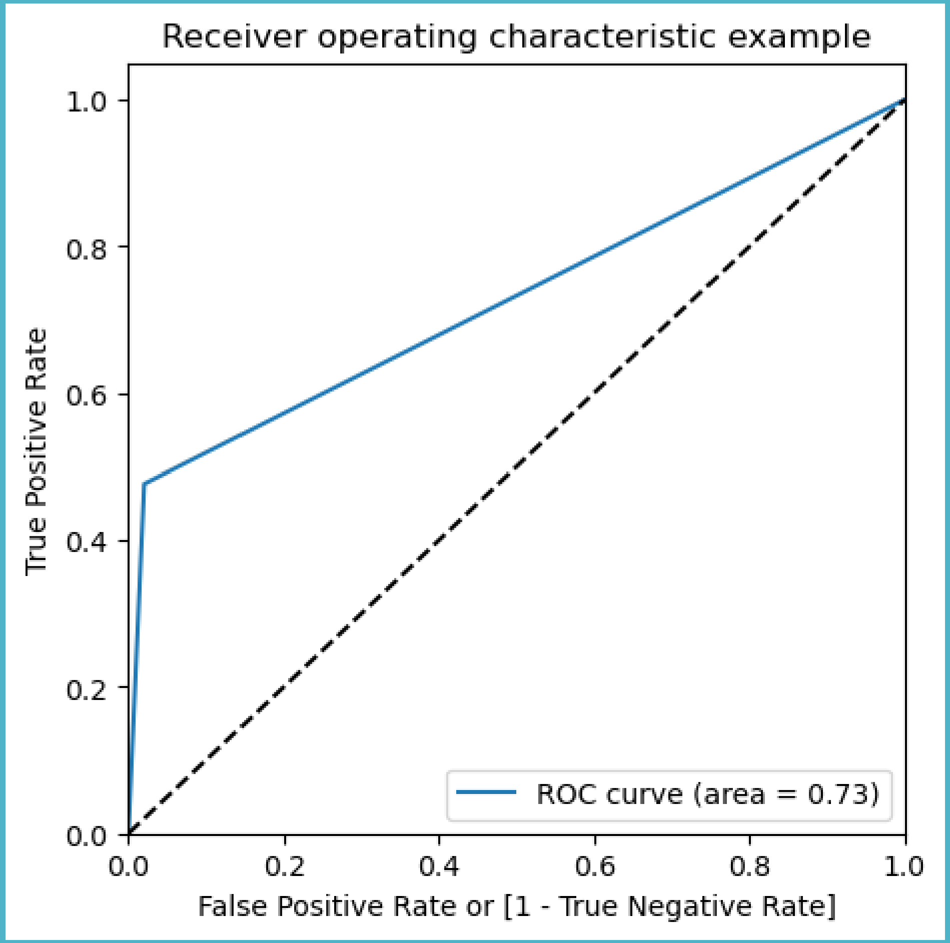




EXPLORATORY DATA ANALYSIS (EDA)



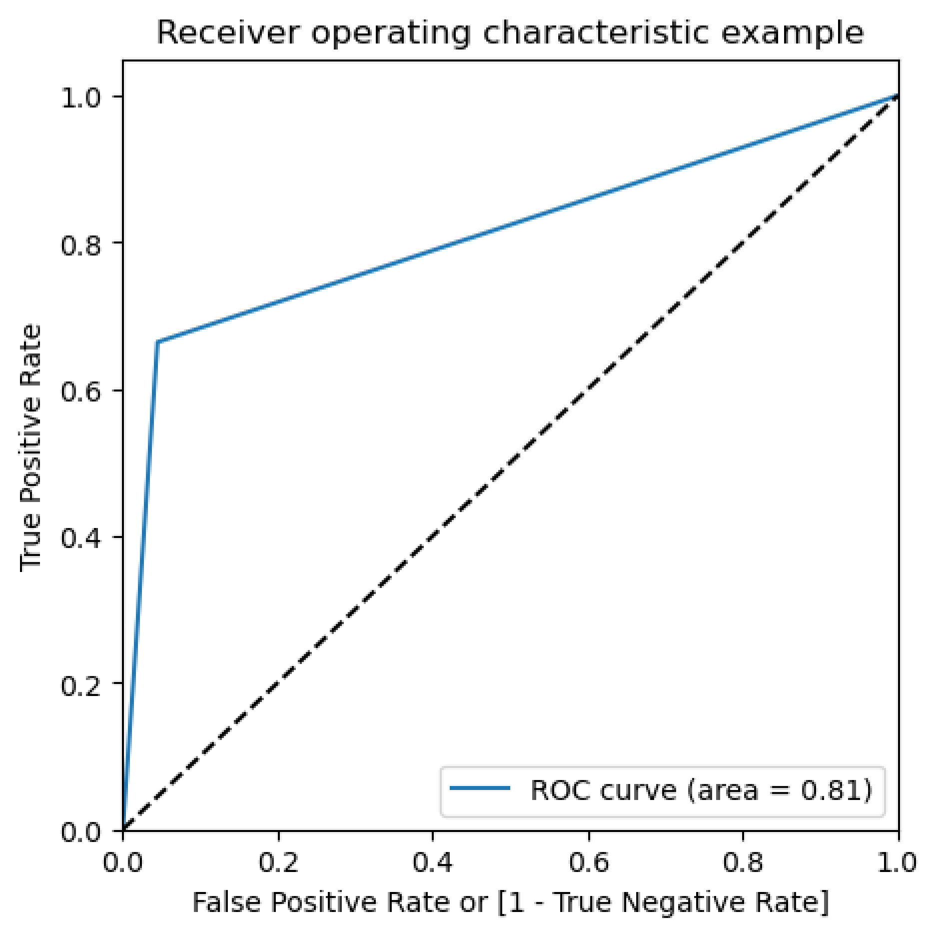
PREDICTING MODELING USING LOGISTIC REGRESSION



Inference:

- In summary, the basic logistic regression model achieved an accuracy of 92.7%. However, the other metrics, such as recall at 47.5% and precision at 71.2%, indicate that the model's performance is not balanced. Additionally, the ROC curve's area under the curve (AUC) is 0.73.
- To address the issue of class imbalance and potentially improve model performance, we plan to implement a more advanced model, specifically the Random Forest algorithm. This advanced model will be used to enhance the predictive capabilities and overall performance of the classification task.

RANDOM FOREST - SMOTE

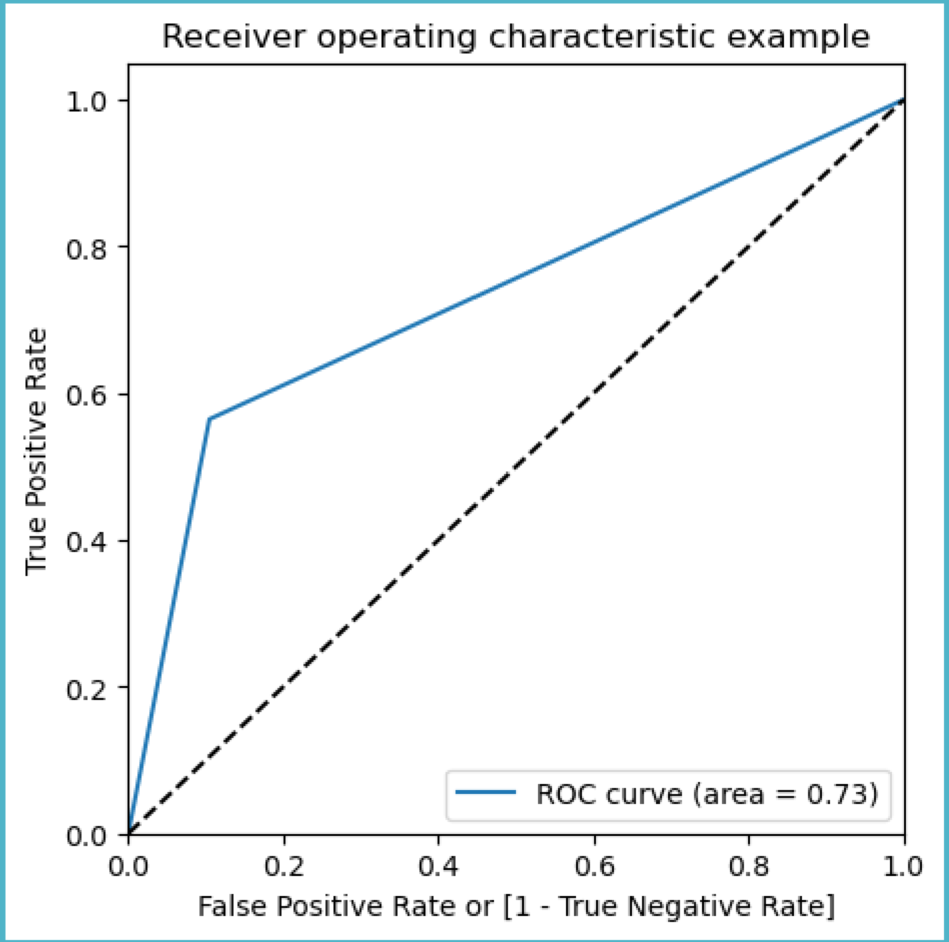


Inference:

- The ROC curve has shown a significant improvement with an AUC (Area Under the Curve) of 0.82, indicating better overall model performance.

- Additionally, both recall and precision metrics have experienced a notable increase in their values, suggesting improved accuracy and the model's ability to correctly identify positive cases (recall) and avoid false positives (precision).

DECISION TREES - SMOTE



Inference:

- When comparing the performance of Decision Trees with SMOTE to that of Random Forest with SMOTE, the Decision Tree model exhibits inefficiency, as indicated by a degradation in the ROC curve value. In other words, the Decision Tree model does not perform as well as the Random Forest model when applied in conjunction with SMOTE for handling class imbalance.

CONCLUSION

- Random Forest model demonstrates significant improvement.
- It achieves a higher ROC curve value of 0.81, indicating enhanced predictive capability.
- The accuracy of the model also stands at an impressive 92%.
- The key features that have a major influence on predicting customer churn are as follows:
  - 1. Outgoing calls in the eighth month (arpu_8)
  - 2. Local incoming calls in the eighth month (loc_ic_t2m_mou_8)
  - 3. Total incoming calls in the eighth month (total_ic_mou_8)
  - 4. Percentage of local outgoing calls in the eighth month (loc_og_mou_8_perc)
  - 5. Local outgoing calls to mobile numbers in the eighth month (loc_og_t2m_mou_8)
  - 6. Total recharge amount in the eighth month (total_rech_amt_8)
  - 7. Local incoming calls to other telecom operators in the eighth month (loc_ic_t2t_mou_8)
  - 8. Local outgoing calls to fixed-line numbers in the eighth month (loc_og_t2t_mou_8)
  - 9. Maximum recharge amount in the eighth month (max_rech_amt_8)
  - 10. Percentage of local incoming calls in the eighth month (loc_ic_mou_8_perc)
  - 11. Last day recharge amount in the eighth month (last_day_rch_amt_8)
  - 12. Total outgoing calls in the eighth month (total_og_mou_8)
  - 13. Local outgoing calls to mobile numbers in the eighth month (loc_og_t2t_mou_8)
  - 14. Off-network outgoing calls in the eighth month (offnet_mou_8).
  - 15. These features collectively provide valuable insights into customer behavior and serve as key indicators for predicting churn.

RECOMMENDATIONS

- To improve customer retention and address potential issues, it's essential to closely monitor month-on-month trends in ARPU (Average Revenue Per User), recharge behavior, and call activities. If a declining trend is identified, the following actions can be taken:
- Offer Recharge Discounts: Implementing special discounts or promotional offers on recharges can incentivize customers to top up their accounts more frequently.
- Introduce Value-Added Packs: Introducing value-added packs for recharges, such as bonus data or additional talk time, can attract and retain customers who are seeking extra benefits from their recharges.
- Evaluate Network Quality: Investigate potential concerns regarding network performance that may be affecting call activities. Addressing network quality issues can improve the overall customer experience and encourage more usage of call services.
- By proactively monitoring and addressing these aspects, telecom providers can enhance customer satisfaction, reduce churn, and ensure the sustainability of their business.