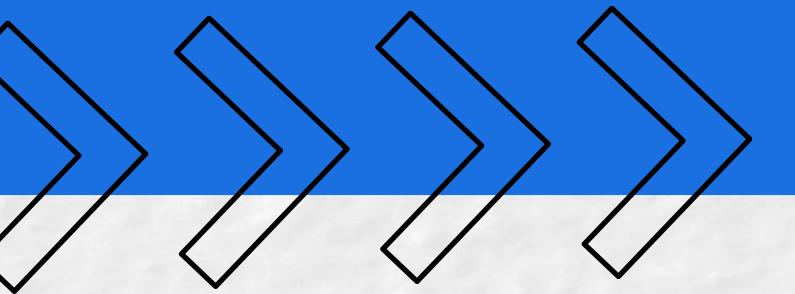


SYRIATEL CUSTOMER CHURN



PROJECT OVERVIEW

- The project's aim is to develop a machine learning model to predict customer churn for SyriaTel, a telecommunications company.
- The dataset included various features such as customer service calls, total day charge, and total day minutes.
- The objective is to create a model that will accurately identify customers who are likely to churn, allowing the company to implement targeted retention strategies.

BUSINESS UNDERSTANDING

- The project's primary objective is to predict customer churn for SyriaTel, a telecom company, to maintain revenue and customer satisfaction.
- Stakeholders, including SyriaTel itself, will benefit from proactive strategies to address churn and improve service quality.
- The customer service department can utilize the churn prediction model to provide timely support to at-risk customers, potentially preventing them from leaving.
- Marketing and sales teams can leverage the model to focus on retaining existing customers, thereby maximizing revenue and improving customer loyalty.

DATA UNDERSTANDING

- The dataset is from SyriaTel Telecommunication company and was obtained from Kaggle .
- It comprises of 21 columns and 3333 rows. The columns have various attributes related to customer demographics, service usage, and churn behavior.
- The rows correspond to a recorded customer. The dataset encompasses both continuous and categorical variables.
- The target variable identified is "churn," with the remaining variables serving as predictors.

OBJECTIVES

Develop an accurate model for predicting customer churn

Identify the factors that significantly affect customer churn

Provide valuable insights that will help reduce churn

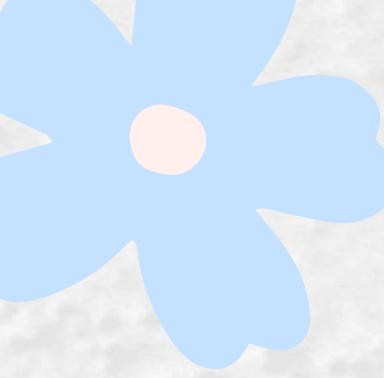
Optimize revenue generation by focusing marketing efforts on retaining existing customers.

MODELING

Four models were employed in the analysis:

- Logistic Regression:
 - Precision: 39% predicted churn cases identified correctly.
 - Recall: Captured 77% actual churn cases.
 - Accuracy: 78% overall accuracy.

- Decision Tree:
 - Precision: 96% for class 0, 64% for class 1.
 - Recall: Achieved 92% for class 0, 78% for class 1.
 - Accuracy: 90% overall accuracy.

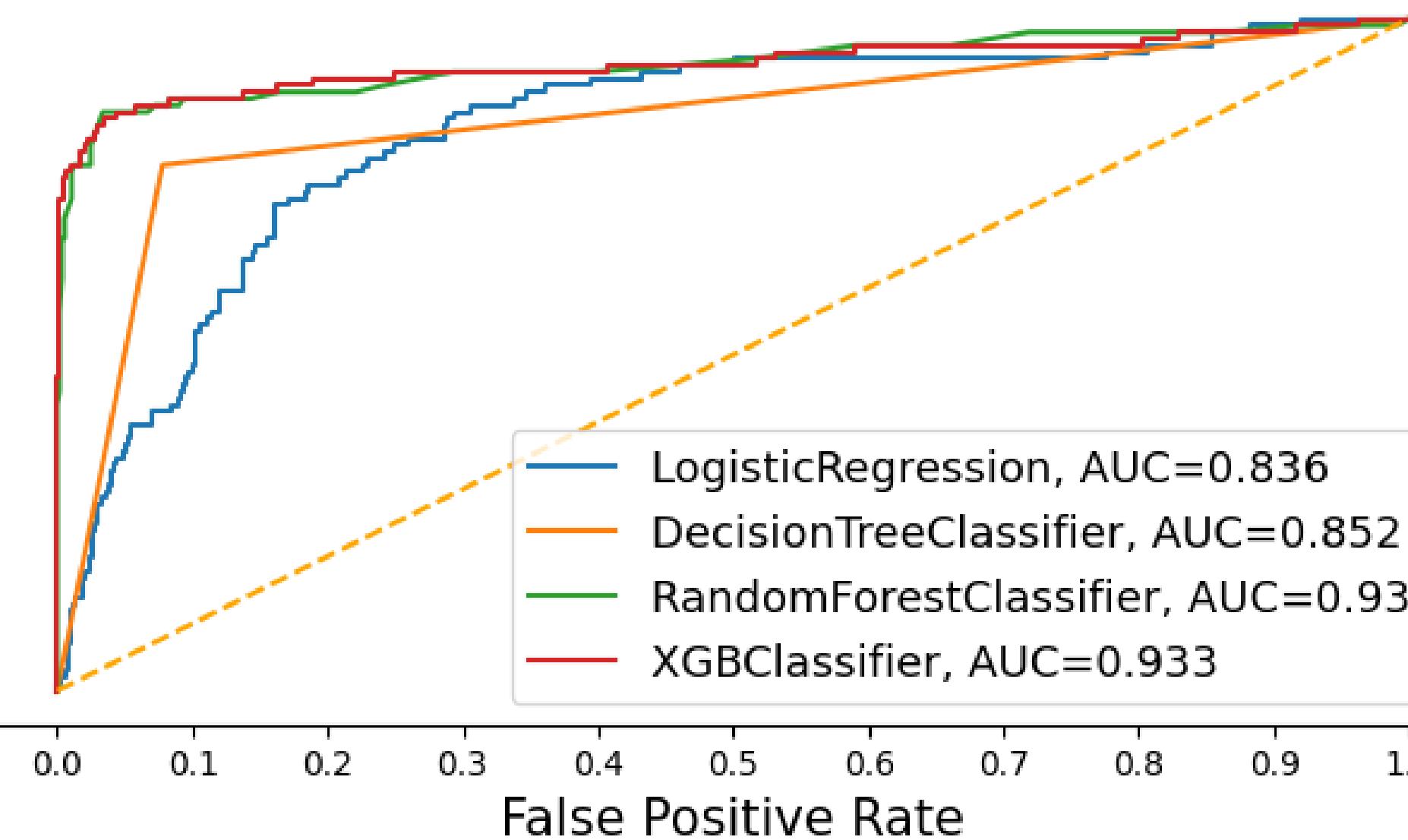


MODELING

- Random Forest:
 - Precision: 96% for class 0, improved to 85% for class 1.
 - Recall: Improved to 98% for class 0, remained stable at 79% for class 1.
 - Accuracy: Increased to 95%.
- XGBoost:
 - Precision: 96% for class 0, improved to 89% for class 1.
 - Recall: Remained 98% for class 0, stable at 79% for class 1.
 - Accuracy: Maintained at 95%.
- Model Tuning: Considered for performance enhancement.

AUC ROC

ROC Curve Analysis



The AUC (Area Under the ROC Curve) measures the ability of a classifier to distinguish between positive and negative classes across all possible thresholds.

Logistic Regression: AUC = 0.836, good but lower than other models.

Decision Tree: AUC = 0.852, slightly better than logistic regression.

Random Forest and XGBoost: AUC = 0.933, strong performance likely due to ensemble nature.

MODEL TUNING

Random Forest and XGBoost models were tuned as they exhibited the highest performance compared to other classifiers.

Tuned Random Forest Model Metrics:

- Precision for class 0 (non-churn): 0.97, recall: 0.97, F1-score: 0.97
- Precision for class 1 (churn): 0.84, recall: 0.82^{y_t}, F1-score: 0.83
- Overall accuracy: 95%

Tuned XGBoost Model Metrics:

- Precision for class 0: 0.97, recall: 0.98, F1-score: 0.97
- Precision for class 1: 0.85, recall: 0.80, F1-score: 0.83
- Overall accuracy: 95%

EVALUATION

- Recall (true positive rate) is the main metric of success.
- Maximizing recall ensures that the model effectively captures as many actual churn cases as possible, minimizing the risk of losing customers.
- The aim is to achieve a recall of 80%. Other performance metrics such as Precision, Accuracy, F1 score were used.
- After tuning, Random Forest achieved a recall of 82%, while XGBoost achieved a recall of 80%. Therefore, Random Forest emerged as the best performing model.
- the top three important features identified as customer service calls, total day charge, and total day minutes.

EVALUATION

- These features reflected customer behavior and usage patterns.
- Higher numbers of customer service calls and total day minutes, along with higher total day charges, suggested potential dissatisfaction or heavier service usage, which could have led to churn.

CONCLUSION

- Customers with three calls to customer service showed the lowest churn rate, indicating potential satisfaction.
- Increased customer service calls correlated with higher churn rates, suggesting dissatisfaction
- Churned customers had higher average total day minutes and charges, indicating higher usage and expenses.
- Having an international plan was associated with lower churn rates.
- The tuned Random Forest model performed the best with a recall of 82%, exceeding the 80% recall target.

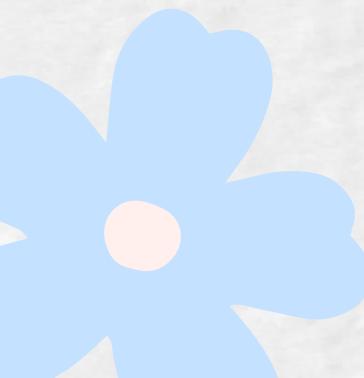
RECOMMENDATIONS

To boost customer retention at SyriaTel:

- Integrate the machine learning model for real-time predictions to continuously enhance accuracy.
- Implement loyalty programs and incentives for high-usage customers.
- Proactively address customer issues identified through feedback forms, particularly for frequent service callers.
- Consider adjusting call charge rates to stay competitive and prevent churn.

RECOMMENDATIONS

- Focus on improving features with lower importance but still relevant to reducing churn.
- Encourage customers to opt for international plans, which show lower churn rates.
- Leverage the tuned Random Forest model for effective churn prediction and mitigation.



THANK YOU

