•TELECOM CHURN PREDICTION FOR SYRIATEL SUBTITLE: IDENTIFYING PATTERNS FOR CUSTOMER RETENTION

Mary Gaceri

AGENDA

INTRODUCTION

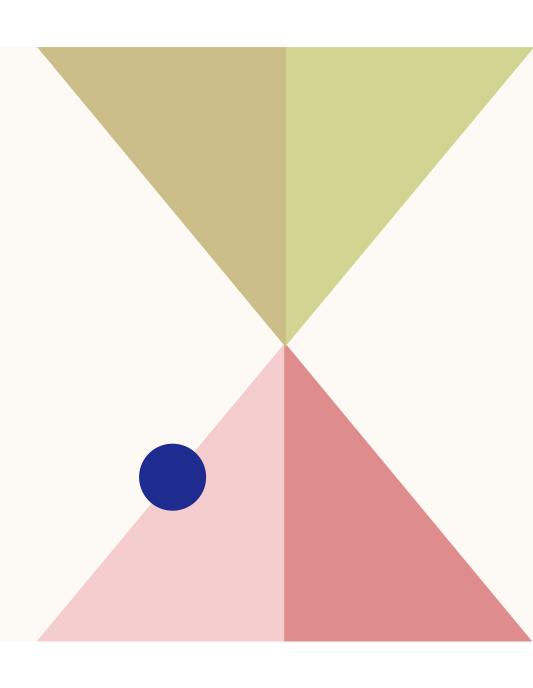This project centers on SyriaTel, a prominent telecommunications company, seeking to predict and reduce customer churn. Customer churn, defined as customers discontinuing services, poses a significant challenge to businesses. Our focus is to leverage data analysis and machine learning techniques to unveil patterns indicative of potential customer exits. By identifying these patterns, SyriaTel can proactively implement measures to retain customers and mitigate revenue loss. The project involves thorough data analysis, the development of predictive models, and the formulation of actionable recommendations to enhance customer retention strategies. The overarching goal is to augment SyriaTel's understanding of customer behaviors, ultimately contributing to more effective retention initiatives.

OBJECTIVES.

1.To apply classification modeling techniques to analyze customer churn data for Syria Tel, aiming to identify and quantify the influential factors contributing to customer churn.

2. Explore the Data for Classification Explore the dataset to understand the relationships between different variabl and the target variable (customer churn)

3.Develop accurate machine learning models for predicting customer churn based on historical data.

4.Identify and analyze patterns and trends in customer behavior that contribute to churn.4.Evaluate model performance and fine-tune hyperparameters for optimal predictive accuracy.

5.Provide actionable insights and recommendations based on churn predictions to aid SyriaTel in implementing effective customer retention strategies.

DATA UNDERSTANDING

During the data understanding phase, the Telecom Churn dataset underwent a detailed examination, focusing on customer-related features and the target variable, churn. Activities included identifying data types, addressing missing values, and generating descriptive statistics for numeric variables. Exploratory Data Analysis (EDA) provided valuable visual insights into data distributions and relationships. Categorical variables were assessed for potential one-hot encoding, and exploration of the target variable illuminated the distribution of churn instances. Correlation analysis identified potential multicollinearity, and integration of domain knowledge enhanced understanding of variable significance. This comprehensive data quality assessment established a robust foundation for subsequent preprocessing and model development stages.
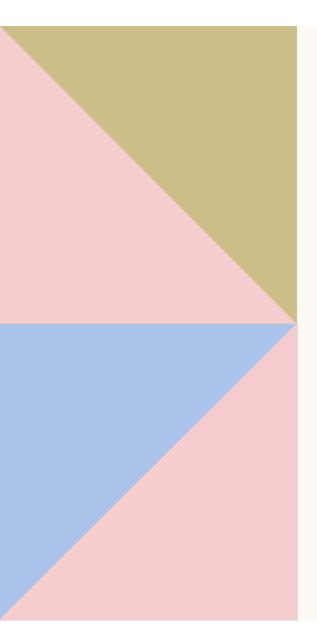
RESULTS AND INTERPRETATION

a)Baseline Model

Logistic Regression:

Accuracy: Around 87% Precision for Churn: 68% Recall for Churn: 18% Findings: The Logistic Regression model performs reasonably well in predicting non-churn instances but struggles to identify churn cases, as indicated by its low recall score. It's somewhat accurate but needs improvement in capturing churn instances effectively
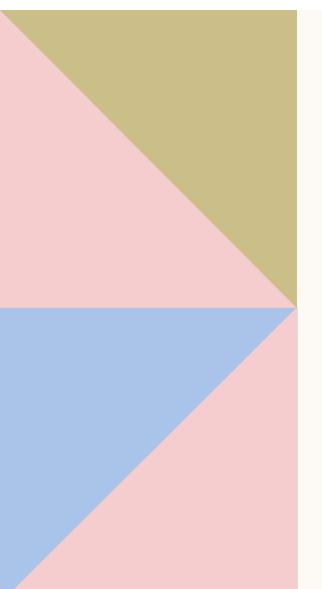
b) More Complex Model

In the Random Forest model Accuracy: Approximately 89% Precision for Churn: 84% Recall for Churn: 32% Findings: The Random Forest model shows improvement compared to Logistic Regression. It achieves better balance in precision and recall for churn. However, it still faces challenges in identifying churn cases effectively.

c) XGBoost Classifier:

Accuracy: Roughly 91% Precision for Churn: 85% Recall for Churn: 47% Findings: The XGBoost model demonstrates the highest accuracy among the three models and significantly improves identifying churn instances. While the balance between precision and recall is better, there's still room for enhancing the recall score for churn.

OVERALL

Best Model:

XGBoost Classifie Reasoning: XGBoost stands out by being the most accurate and striking a good balance between finding all actual churn cases and minimizing incorrect predictions. It excels at identifying customers who are likely to churn compared to Logistic Regression and Random Forest models.Improvement Areas:All models face challenges in accurately pinpointing instances of churn, suggesting the necessity for improvements. This could involve delving deeper into the data, enhancing features, or exploring alternative modeling techniques to better grasp and predict patterns associated with customer churn.

CONCLUSION

All models exhibit predictive abilities, with XGBoost showing particular promise in predicting customer churn due to its elevated accuracy and improved ability to identify churn cases. Nevertheless, there is room for improvement in refining the models to enhance churn prediction, with a focus on increasing recall while maintaining precision.

RECOMMENDATIONS

1.Improve Customer Service Experience

2.Communicate oactively

3.Understand Churn Indicators

THANK YOU

Mary Gaceri Mutuma

gacerimutuma0@gmail.com