SYRIATEL CUSTOMER CHURN PREDICTION

Author: IMRAN AWADH

PROJECT OVERVIEW

- The objective of this project is to develop a machine learning model that can predict whether a customer is likely to discontinue their relationship with SyriaTel in the near future. By identifying at-risk customers, the model will enable SyriaTel to take proactive measures to prevent churn, improve customer retention, and optimize retention strategies, ultimately reducing customer loss.

BUSINESS UNDERSTANDING

- **Objective** :

1. Develop a predictive model to identify customers at risk of churn.
2. Provide actionable insights to reduce churn and improve customer retention.

- **Success Criteria** –

1. High accuracy and recall for the churn class.
2. Actionable insights for reducing churn.

Key Questions

1. What features most influence customer churn?
2. 2. How accurately can churn be predicted?
3. 3. What strategies can be implemented to reduce churn rates?

DATA EXPLORATION AND ANALYSIS (EDA)

- Data from Syriatel was imported and explored for analysis.
- Summary statistics and distribution checks were conducted to understand data quality and characteristics.
- Key insights about data imbalance and feature distributions were noted.

DATA CLEANING

- Key Steps:
- 1. Handle missing values and duplicates.
- 2. Ensure data consistency.

DATA PREPROCESSING

- The dataset was pre-processed to handle any missing values, remove uneccessary columns, and encode categorical data

- Key Steps:
- 1. Encode categorical variables.
- 2. Split into dependent (y) and independent variables (X).
- 3. Standardize numerical features.

FEATURE ENGINEERING:

Encoding Categorical Variables:

- Label Encoding was used for columns like `state`, `international plan`, and `voice mail plan`.

Scaling:

- The dataset was scaled using `StandardScaler` to ensure that numerical features are on a similar scale for model training.

MODELLING

- Two classification models were used:
- - Logistic Regression: A baseline model for comparison.
- - Random Forest: A more complex model with hyperparameter tuning.

RECOMMENDATIONS

- - Focus on at-risk customers identified by the Random Forest model.
- - Improve customer engagement by addressing common churn reasons.
- - Continuously monitor and refine the model with updated data.