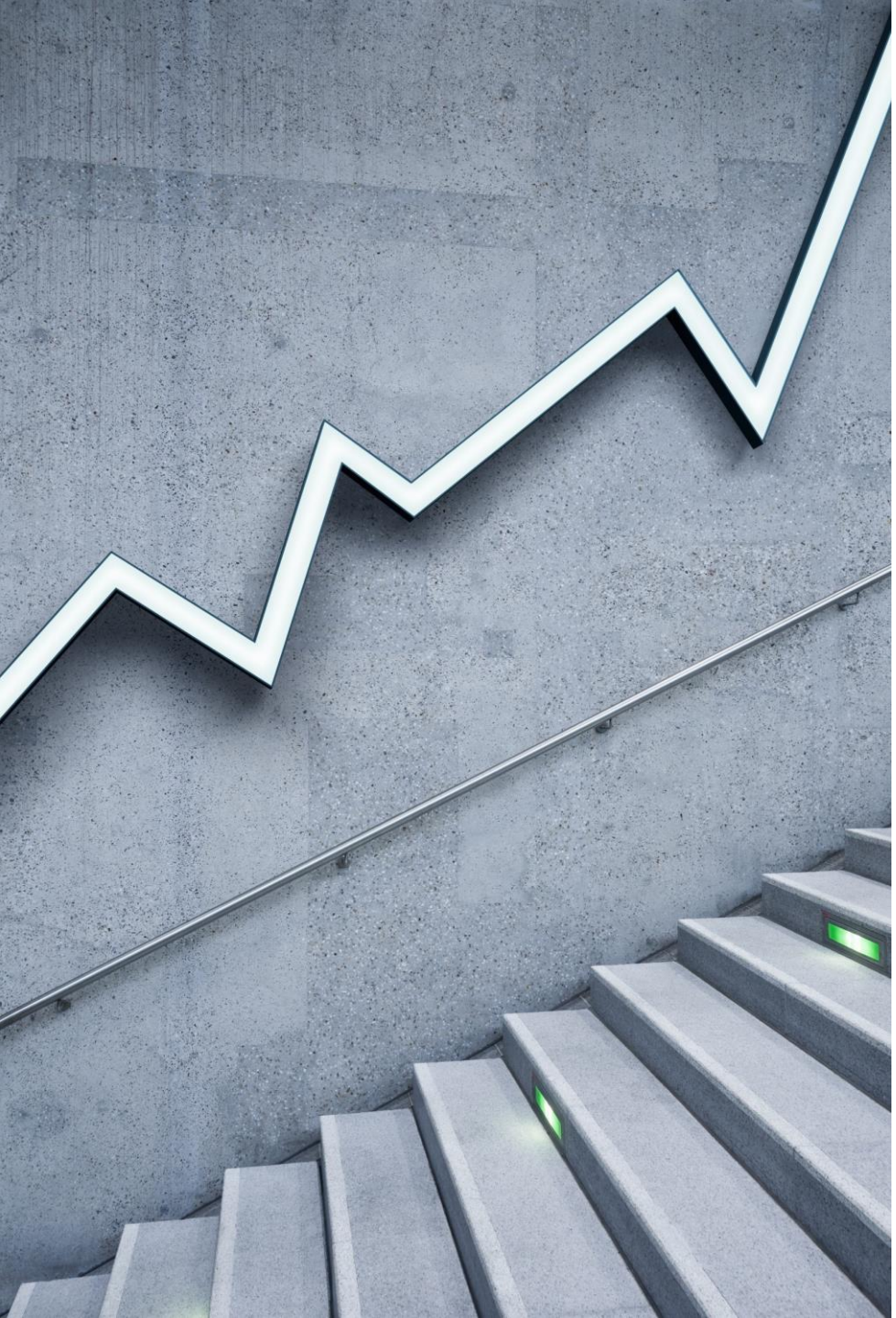


Predicting Customer Retention

Syriatel Mobile Telecom

Business overview

SyriaTel, a leading telecommunications provider, faces significant challenges in retaining customers.

The goal of this project is to predict customer churn using machine learning techniques and provide actionable insights to reduce churn rates.

By identifying customers likely to leave, SyriaTel can proactively implement targeted retention strategies, minimizing revenue loss and enhancing customer loyalty.

Problem statement

Customer Churn: SyriaTel customer churn is very unpredictable; therefore, the business struggles to anticipate which customers are likely to stop using its services and why. This unpredictability results in reactive strategies, where the company focuses on damage control after customers have already decided to leave, rather than proactively preventing churn

Data Understanding

-This dataset was sourced from kaggle and it has 3333 rows and 21 columns.

-The dataset has been used by other data scientists before and is accessible for anyone to look through it, it is updated often.

-The data is in csv format inside a folder named data. I then read through the data using the panda's library in order to get a data frame as our output .

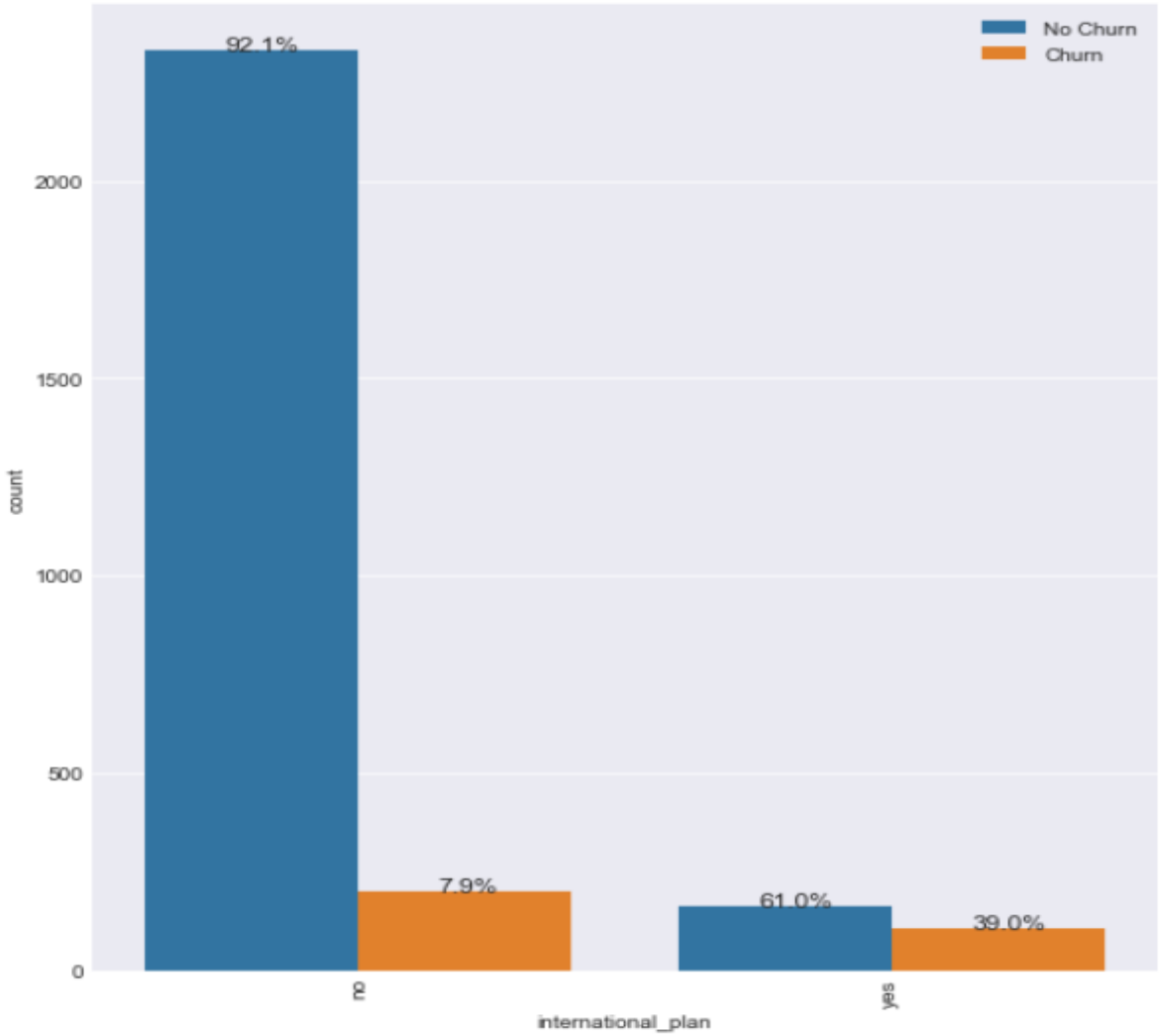
-The dataset has data recorded in different data type including float, integers and objects

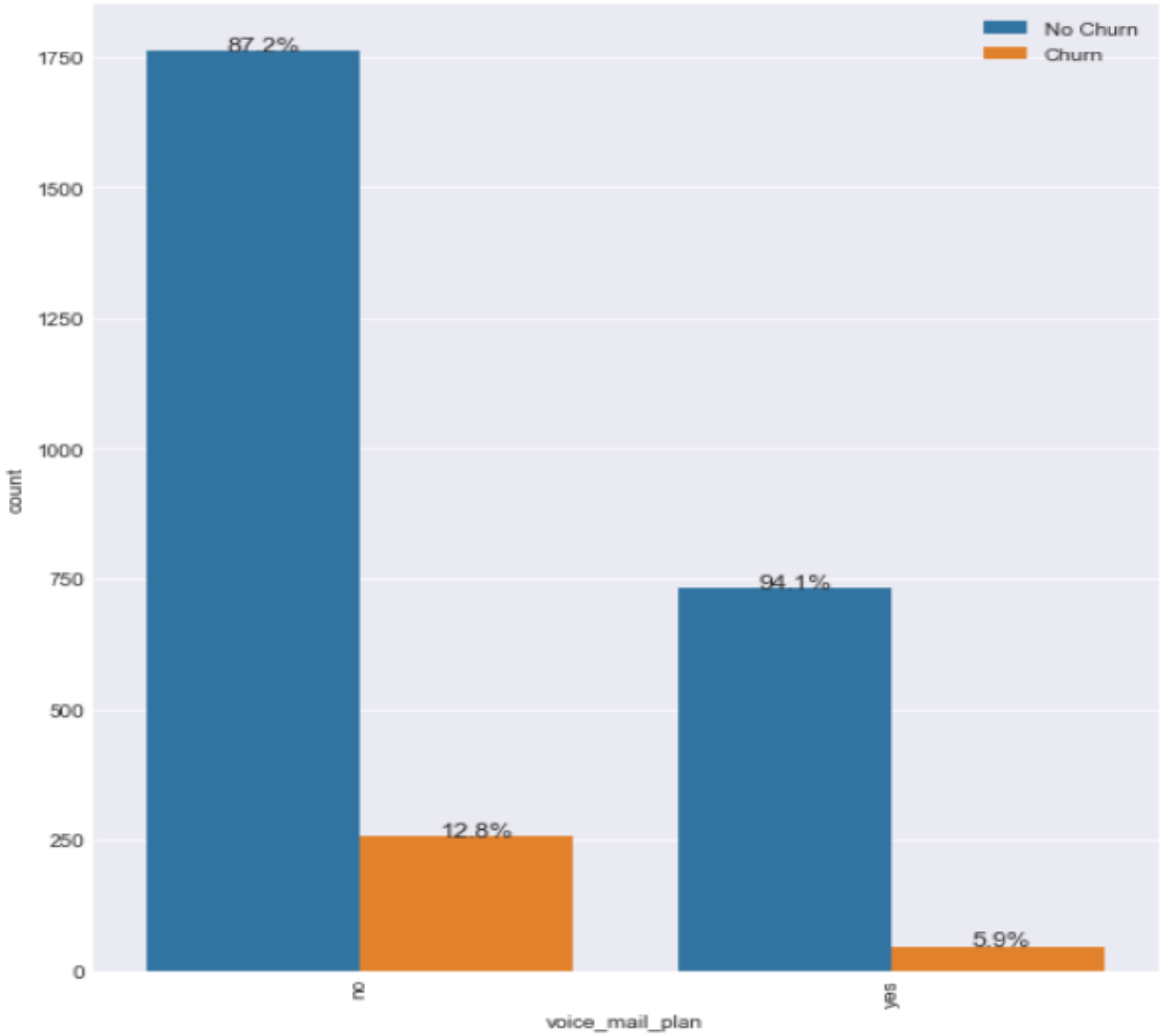
Data Analysis

We will use univariate, bivariate, and multivariate analysis to perform a thorough investigation of the data in this section.

Finding potential correlations between the features and variable distribution is the goal of this kind of data exploration, which will be crucial for feature engineering and modelling. Features that have a high correlation with the target are often good for building baseline models.

Bivariate Analysis





Modelling

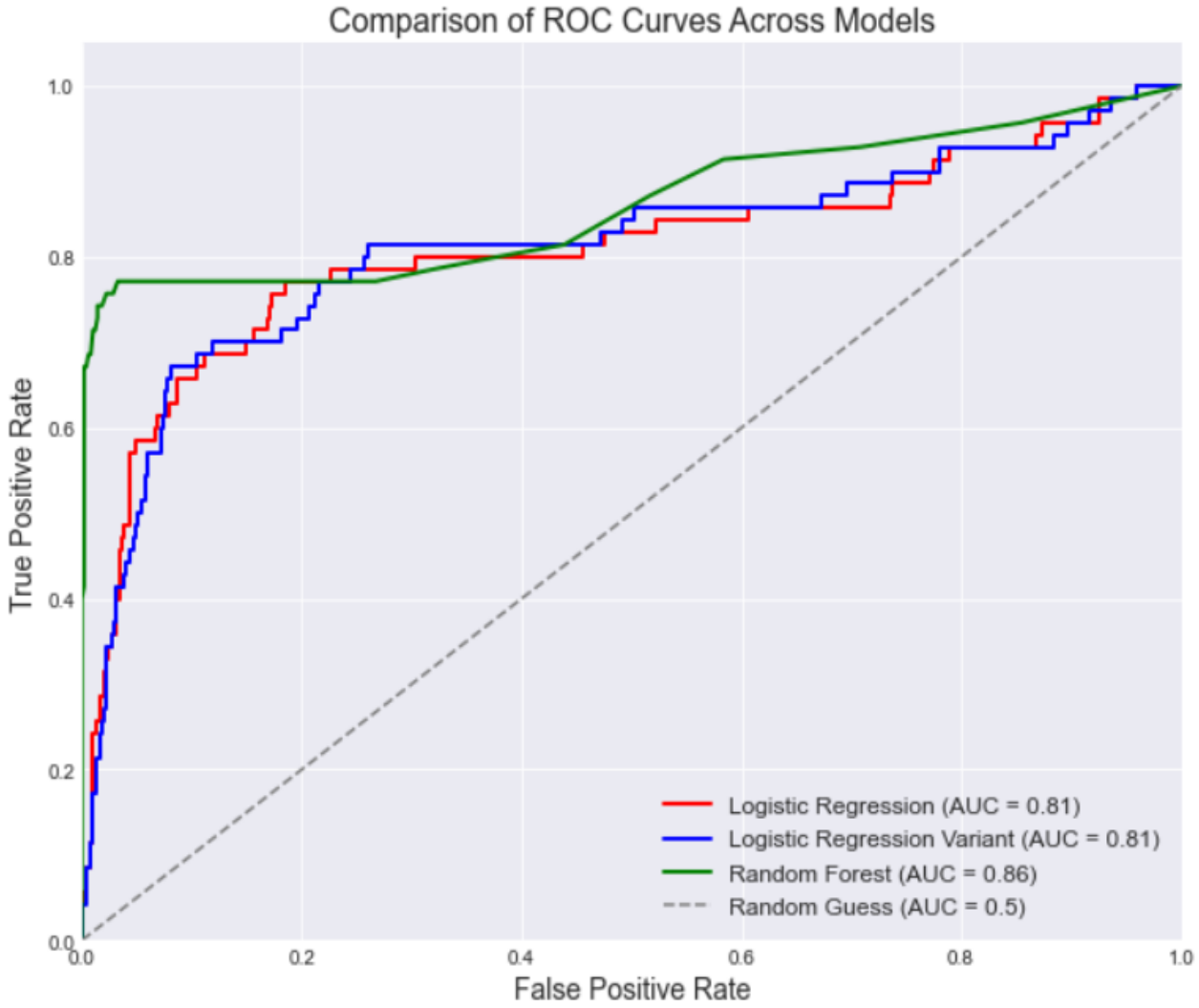
We develop and evaluate multiple models, followed by hyperparameter tuning to enhance their performance. Our goal is to identify the model and parameter configuration that delivers the best results.

The models we train and evaluate include:

1)Logistic Regression,

2)K-Nearest Neighbors (KNN),

MODEL PERFORMANCE



The Random Forest model provides the best classification performance among the evaluated models, as it has the highest AUC and its ROC curve is closest to the top-left corner.

Model Selected

1.LogiLogistic Regression (Red curve, AUC = 0.81): The baseline logistic regression model.

logistic Regression Variant (Blue curve, AUC = 0.81): A slightly modified version of logistic regression with similar performance.

Random Forest (Green curve, AUC = 0.86): The best-performing model in this comparison, with the highest area under the curve (AUC).

Random Guess (Dashed grey line, AUC = 0.5): Represents a random classifier, indicating no discriminatory ability.

2,Key Metric (AUC):

The Area Under the Curve (AUC) quantifies the overall ability of a model to distinguish between classes.

Higher AUC values indicate better model performance.

In this case, the Random Forest model (AUC = 0.86) outperforms the logistic regression models (both with AUC = 0.81), while the random guess (AUC = 0.5) serves as the baseline.

Conclusion:

Recommendations

1)Targeted Customer Retention Strategies: Identify customers who are likely to churn and offer them personalized deals or services to retain them.

2)Resource Allocation: Focus retention efforts on customers who are at risk of leaving, improving efficiency in marketing and customer support.

3)Revenue Growth: By reducing churn, Syriatel can increase customer lifetime value (CLV) and stabilize revenue growth.

Thank you!

Thank you to our audience for staying through the presentation!

Q&A

Dorothy Awino
Github-
LinkedIn -

-