

MARCH 07, 2024

# INTRODUCTION

A Project by  
Alice Mumbi.

# Overview

Predicting Customer Churn in a  
Telecommunications Company



# Importance of customer retention and reducing churn for telecom businesses

## *Dataset Used*

Kaggle Dataset

## *Objectives*

- Identify factors influencing churn
- Build a classifier to predict customer churn



## BUSINESS UNDERSTANDING

The primary goal is to develop a predictive model that can accurately identify customers who are likely to churn or discontinue their services with SyriaTel in the near future. By doing so, the company aims to proactively address customer retention, minimize revenue loss associated with churn, and improve overall customer satisfaction and loyalty

# DATA UNDERSTANDING

- Overview of the dataset's structure, including the features and target variable.
- Description of the features available in the dataset.
- Importance of understanding the data before proceeding with analysis.





# *Data Preparation*



## ***Exploratory Data Analysis (EDA)***

Summary statistics

Identification of any data quality issues: Missing values

Visualization of the distribution of numerical variables, correlation matrices

## ***Preprocessing***

Distribution of categorical variables

Handling of categorical variables: One-hot encoding

Feature Scaling

Feature Engineering



# *Modeling and Evaluation*

A top-down view of a white laptop on a light-colored desk. To the left of the laptop is a smartphone with a brown leather-like case. To the right is a small black rectangular device with a triangular logo. The laptop screen shows a grid of images. A semi-transparent pink box with white text is overlaid on the bottom half of the image.

## Train Test Split

Define features ( $X$ ) and target variable ( $y$ )

Split the dataset into training and testing sets



*Selection of classification algorithms*

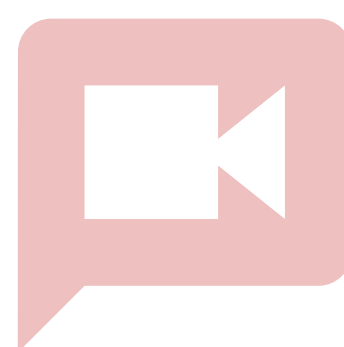
*Comparison of model performance*

# *K-Nearest Neighbors*

METRICS PERFORMANCE EVALUATION: ACCURACY, PRECISION, RECALL, AND F1-SCORE

*The model correctly predicts the class label for about 88.2% of the instances in the testing dataset*





# Hyperparameter Tuning and Optimization

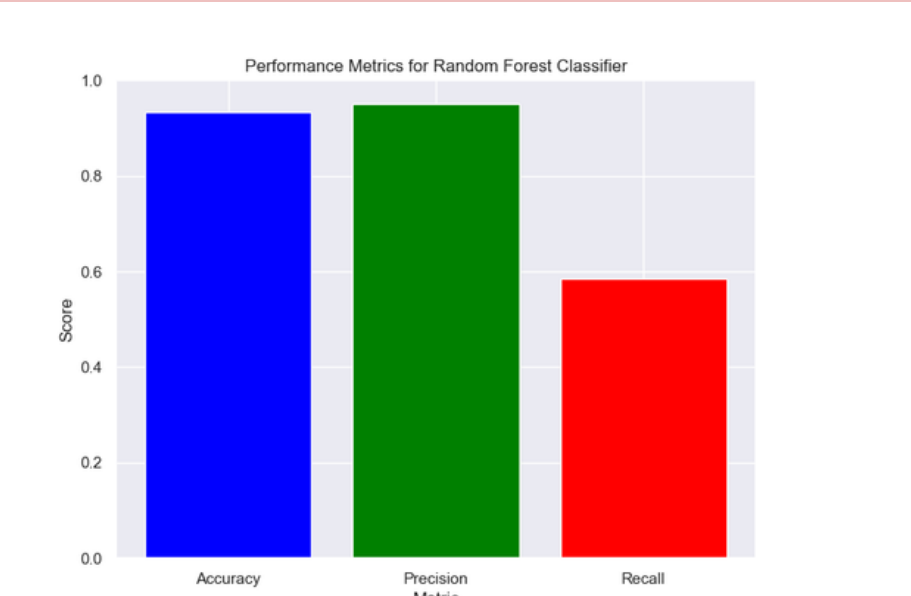
## Identification of best hyperparameters

Techniques: GridSearchCV

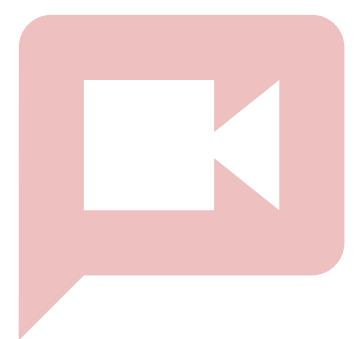
# Random Forest

## METRICS PERFORMANCE EVALUATION: ACCURACY, PRECISION, RECALL

*The precision score of 0.9516 indicates that 95.16% of the instances predicted as churn were actually churn.*  
*The recall score of 0.5842 means that the model correctly identified 58.42% of the actual churn instances*







# Hyperparameter Tuning and Optimization

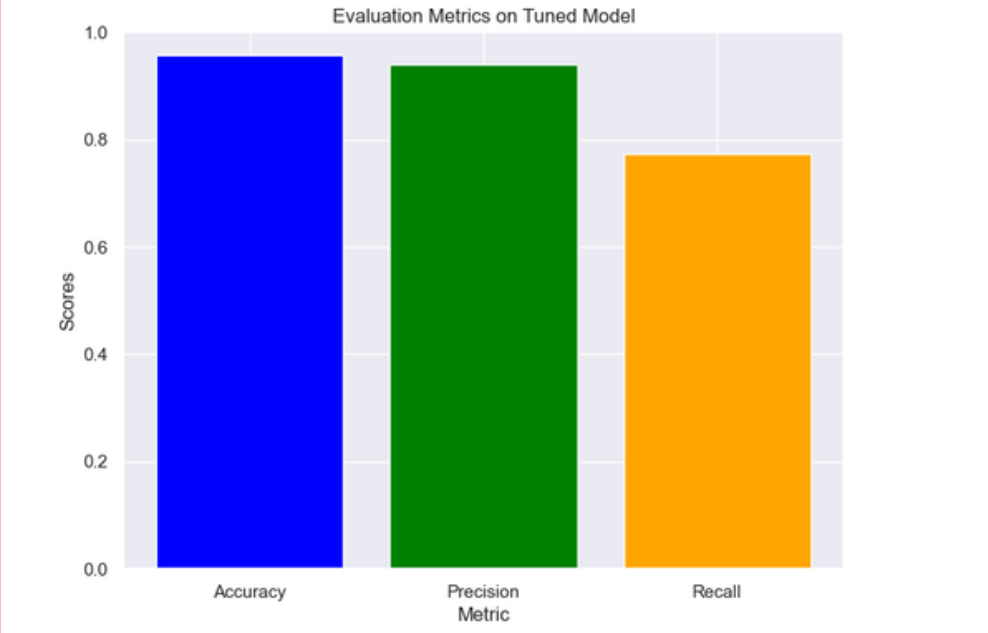
## Identification of best hyperparameters

Techniques: Regularization

# XGBoost

## METRICS PERFORMANCE EVALUATION: ACCURACY, PRECISION, RECALL

*We have successfully improved the Recall value therefore an improvement in correctly identifying all churn instances*



# *Impact on model performance*



## **GridSearchCV**



The model achieves an improved accuracy of approximately 88.4% on the training data





## **Regularization**

Model improvement with an accuracy of approximately 98.46% on the training dataset and 95.80% on the testing dataset

# *Conclusion*

- 
- 
- After thorough data exploration, preprocessing, and model development, I have successfully built a predictive model to forecast customer churn for SyriaTel, a telecommunications company.
  - The final model, an optimized XGBoost classifier, achieved an impressive accuracy of 95.8% on the testing dataset, indicating its effectiveness in predicting customer churn.
  - Through feature engineering, model tuning, and evaluation, I have identified key factors influencing customer churn, including customer service calls, account length, and international plan subscription.

*Recommendations*

- 
- 
- **Proactive Customer Engagement** Utilize insights from the predictive model to proactively engage with at-risk customers, offering targeted promotions or incentives to encourage retention.
  - **Enhanced Service Quality** Focus on improving service quality and customer satisfaction, particularly in areas identified as significant predictors of churn, such as customer service interactions and plan subscriptions.
  - **Personalized Marketing Strategies** Leverage the predictive model to tailor marketing campaigns and communication strategies based on individual customer profiles and behavior patterns, thereby increasing engagement and loyalty.
  - **Continuous Model Monitoring and Improvement** Establish a framework for ongoing model monitoring and evaluation to ensure its continued effectiveness over time. Regularly update the model with new data and refine its parameters as necessary to adapt to changing customer dynamics and market trends.

*THANK YOU*

[Github:/liciemw](#)

Email

Mumbiwm@gmail.com