

# CUSTOMER CHURN PREDICTION MODEL FOR SYRIATEL TELECOMMUNICATIONS COMPANY

## CONTRIBUTORS

1. Lisa Maina
2. Allan Omolo
3. Christine Gitau
4. Zacheaus Nyaga
5. Esther Gakio
6. Linet Wangui

# TABLE OF CONTENTS

- 1. Research Questions
- 2. Objectives
- 3. Business Understanding
- 4. Business Problem
- 5. Data Understanding
- 6. Modeling
- 7. Model Results
- 8. Gradient Boosting Classifier
- 9. Feature Importance
- 10. Limitations
- 11. Recommendations
- 12. Conclusion

# RESEARCH QUESTIONS

---

The project aims at answering the following:

1. What were the factors influencing customer churn?
2. What is the best model for predicting customer churn?
3. How can the insights from feature importance help improve customer churn?

# OBJECTIVES

---

1. To develop a binary classification, model that forecasts if a client will "soon" terminate their relationship with SyriaTel,
2. To determine what factors influence customer churn,
3. To determine the best model for predicting customer churn,
4. To evaluate how insights from feature importance can help improve customer churn.

# BUSINESS UNDERSTANDING



SyriaTel is a leading telecommunications company. Its stakeholders include the marketing and sales teams, customer service departments, and upper management, with the potential to significantly impact customer retention and overall profitability.



## BUSINESS PROBLEM

- SyriaTel, is facing challenges with customer churn, causing revenue loss and harming its reputation and competitiveness.
- To address this, SyriaTel seeks to analyse predictive patterns and build a reliable classifier to forecast customer churn, thereby mitigating the issue and enhancing customer retention strategies.

# DATA UNDERSTANDING



- The title of this dataset is called "[SyriaTel Customer Churn](#)"
- **Source:** [Kaggle](#)
- **Number of columns:** 21
- **Target variable:** Churn
- **Number of records:** 3333
- **The columns dropped:**
  1. Phone Number
  2. State

# MODELING

For this project we choose to use 4 different classification models:



Logistic Regression



Decision Tree



Random Forest  
Classifier



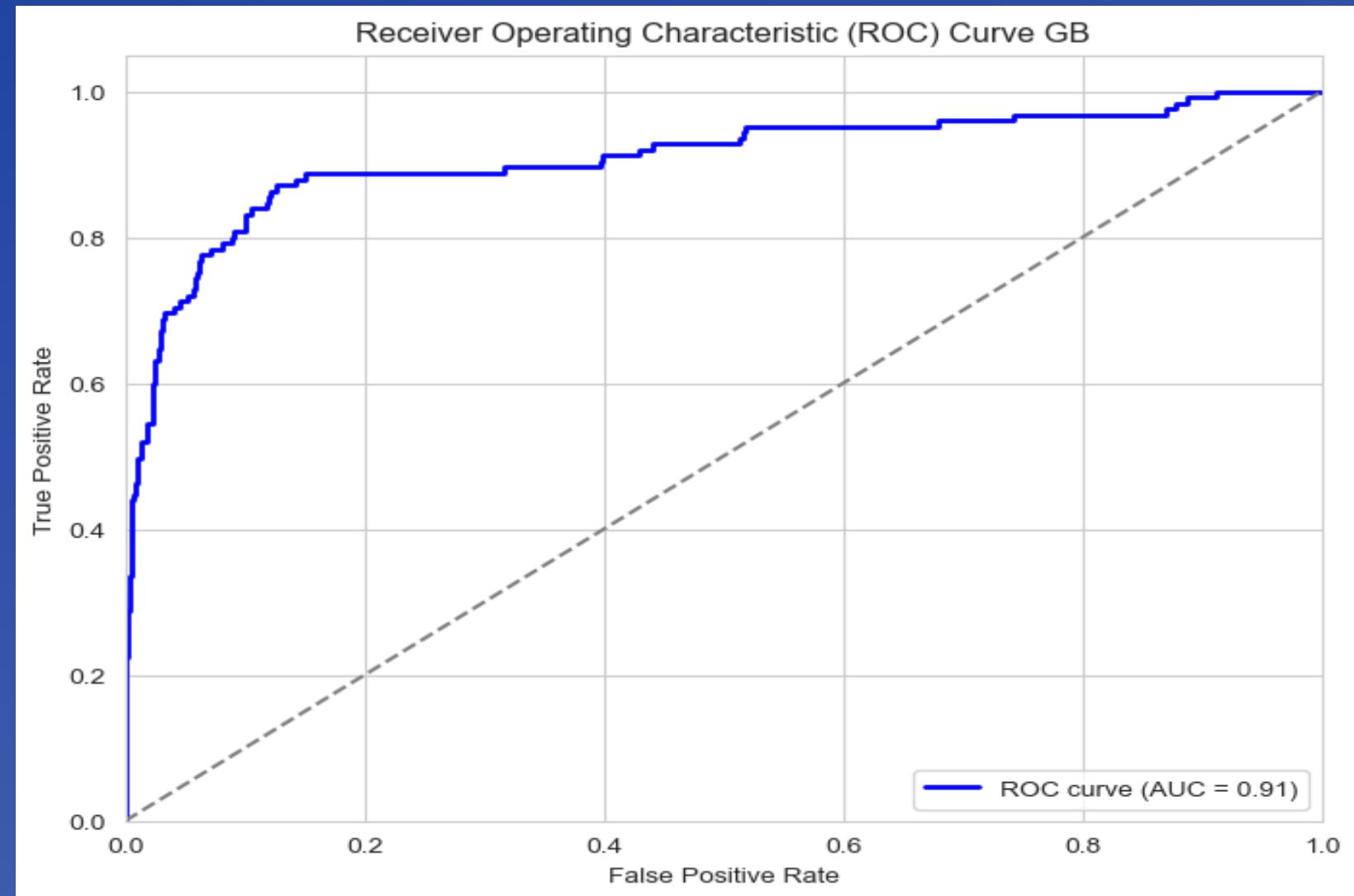
Gradient Boosting  
Classifier

# MODEL RESULTS

MODEL	TECHNIQUE	TRAINING ACCURACY	TESTING ACCURACY	AUC
<b>Logistic Regression</b>	Technic Imbalanced	0.8626	0.8501	0.7609
	SMOTE	0.7274	0.7122	0.7774
	SMOTE + Tuning	0.7279	0.7170	0.7775
<b>Decision Tree</b>	SMOTE	1	0.8333	0.7800
	SMOTE + Tuning	1	0.8549	0.8026
<b>Random Forest</b>	SMOTE	1	0.9208	0.9132
	SMOTE + Tuning	0.9915	0.9220	0.9097
<b>Gradient Boosting</b>	SMOTE	0.9098	0.9100	0.9091

# GRADIENT BOOSTING CLASSIFIER

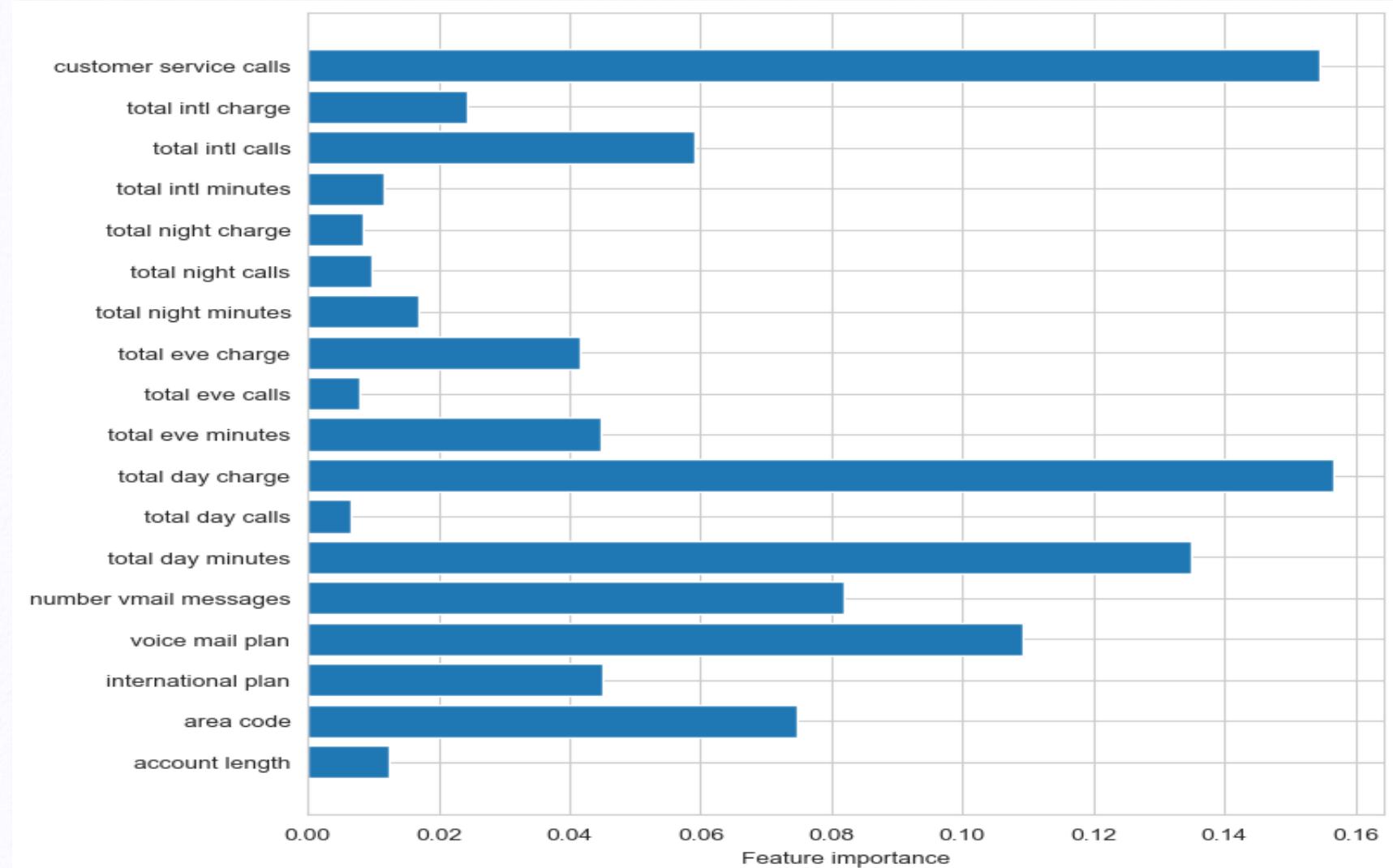
The Gradient Boosting Classifier outperforms other models with an accuracy of 0.91 for testing and 0.90 for training. It's less prone to overfitting and exhibits strong generalization capabilities.



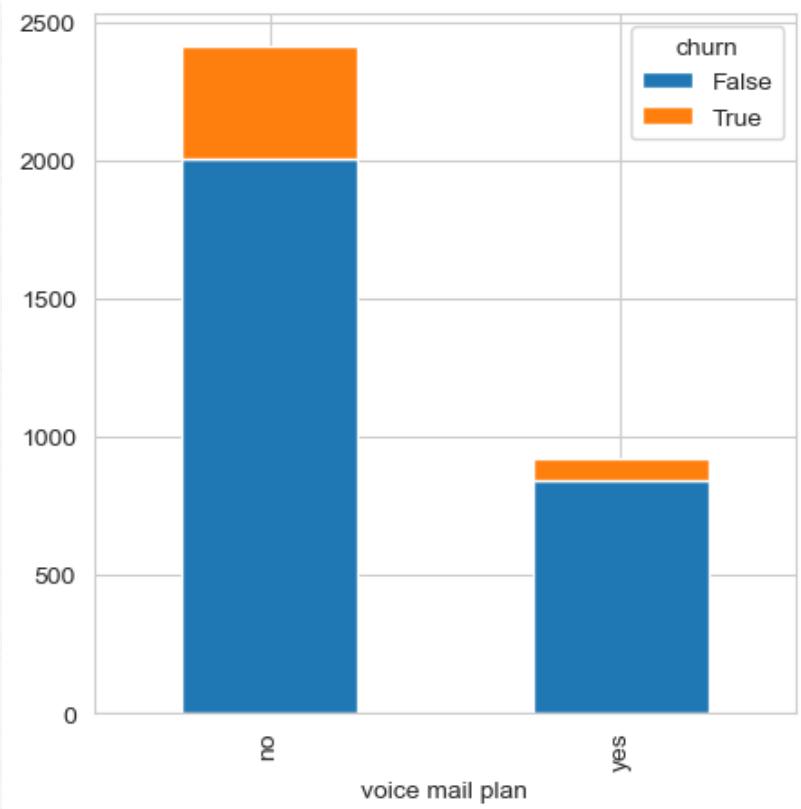
# FEATURE IMPORTANCE

## TOP 5 MOST CRITICAL DETERMINANTS OF CHURN:

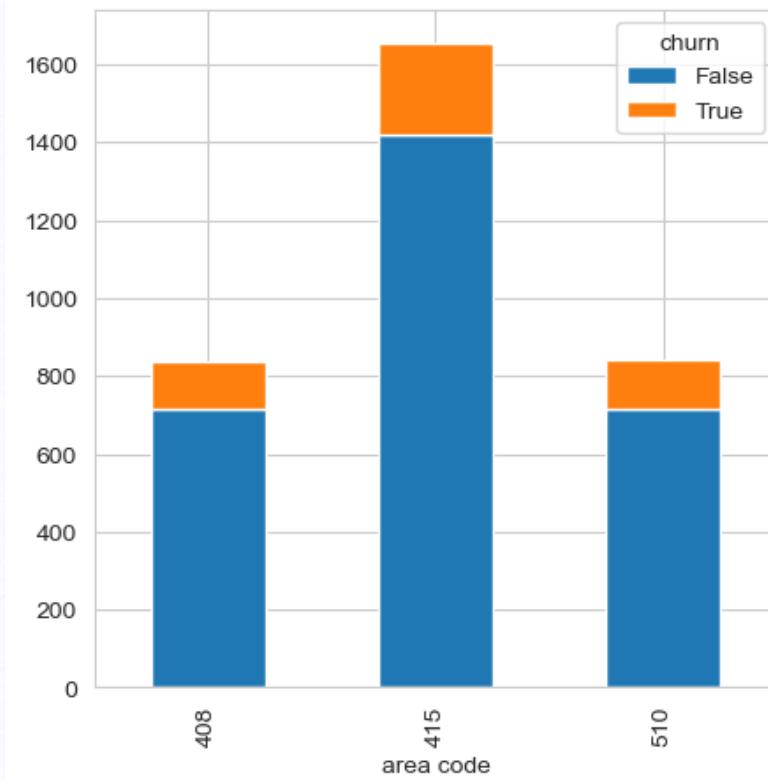
- a) Customer Service Charge,
- b) Total day minutes,
- c) Total day charge,
- d) Voice mail plan, Area code.



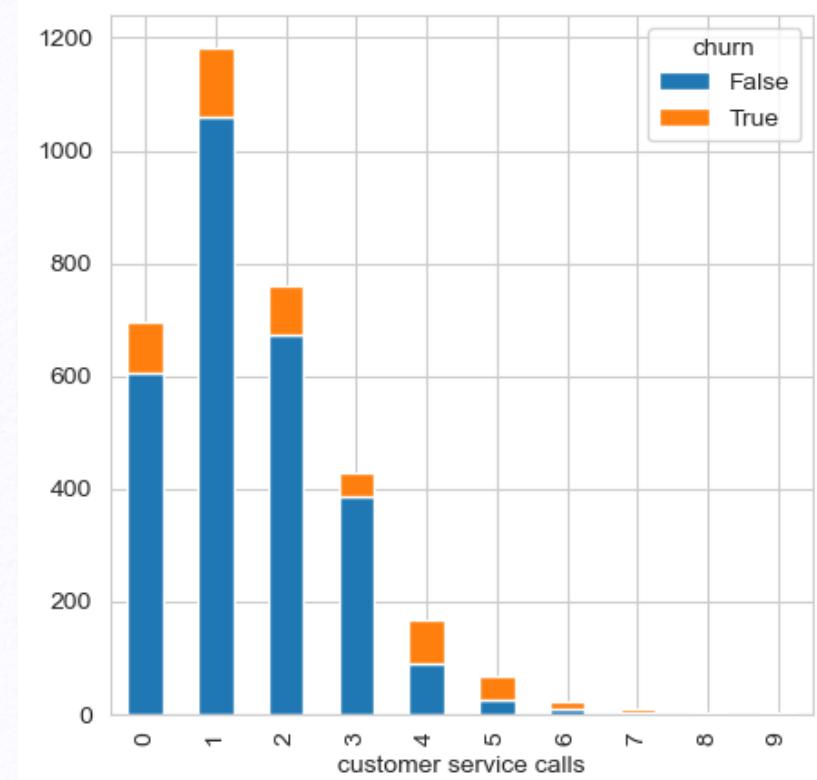
## GRAPH OF CHURN AGAINST VOICE MAIL PLAN



## GRAPH OF CHURN AGAINST AREA CODE



## GRAPH OF CHURN AGAINST CUSTOMER SERVICE CALLS





## LIMITATIONS

- Fine Tuning Constraints.
- Computation Costs of Models with Large Parameter Spaces.
- Need for Comprehensive Pre-Modelling Analysis.

# RECOMMENDATIONS



01

Customer segmentation using identified key features.

02

Gradient Boosting classifier as the model of choice for forecasting.

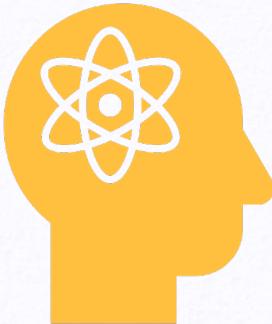
03

Perform comprehensive exploratory data: more Insights – Intensive exploratory data.

04

Tailor retention strategies and marketing campaigns to address the specific needs and behaviors of each segment, e.g:

- Discounts to area code 415,
- Increasing marketing campaigns to area code 415,
- Improve customer service,
- Include voice mail plan,
- To provide loyalty rewards,
- Offer specialized plans that provide discounted rates for calls made throughout the day.



# CONCLUSION

In conclusion, based on the model's **accuracy** and **AUC scores**, **Gradient Boosting classifier** is the best model to predict churn of SyriaTel's customers towards strategizing, saving costs, and prioritizing resources to increase profits. Also, customer service call, total day charge, total day minute, voice mail plan, and area codes are the most important that determine whether a customer will churn or not.



**THANK YOU**