

SYRIATEL CUSTOMER CHURN PREDICTION.

Leveraging Machine Learning for Improved
Customer Retention





TABLE OF CONTENT

- OVERVIEW
 - BUSINESS AND DATA UNDERSTANDING
 - MODELING
 - EVALUATION
 - RECOMMENDATIONS
 - NEXT STEPS
- 

PROJECT OVERVIEW

- Customer Churn Challenge: Syriatel is struggling with a high rate of customer churn, a critical issue in the telecom industry.
- Cost-Effectiveness: Retaining customers is more cost-effective than acquiring new ones, essential for profitability.
- Predictive Model: The project aims to develop a model to predict customers likely to churn.
- Proactive Retention: Identifying at-risk customers enables Syriatel to implement strategies to reduce churn and revenue loss.

BUSINESS UNDERSTANDING

- Problem Definition:

- Predict which customers are likely to churn (leave the service provider)
- Identify the factors or patterns that contribute to customer churn

Business Objectives:

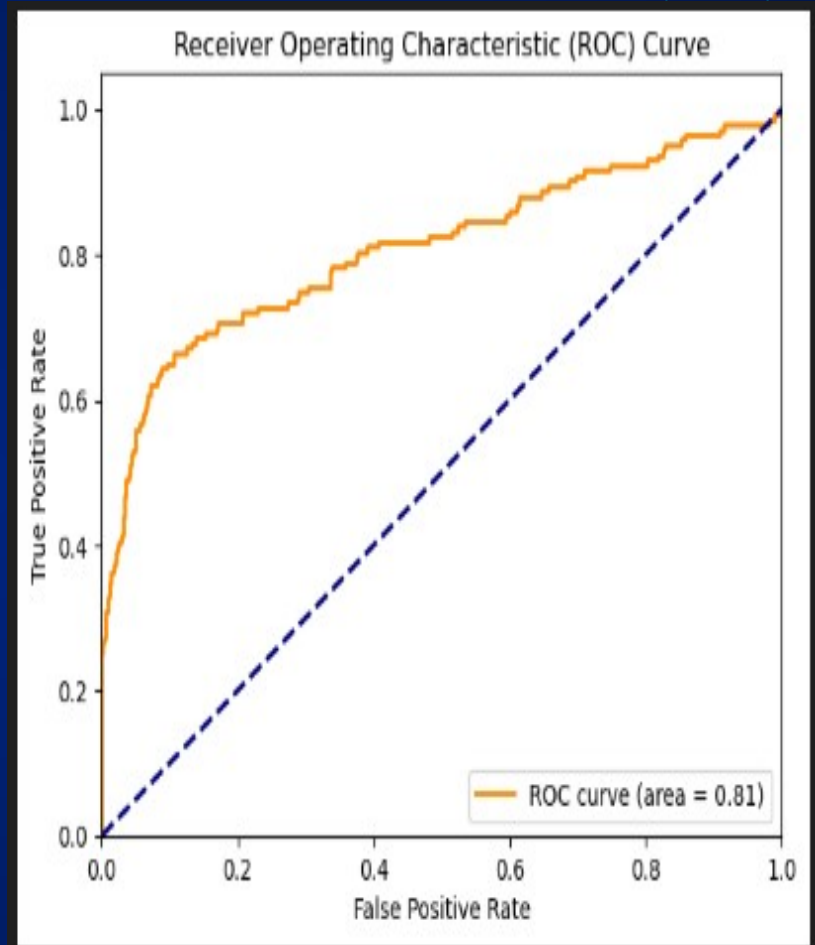
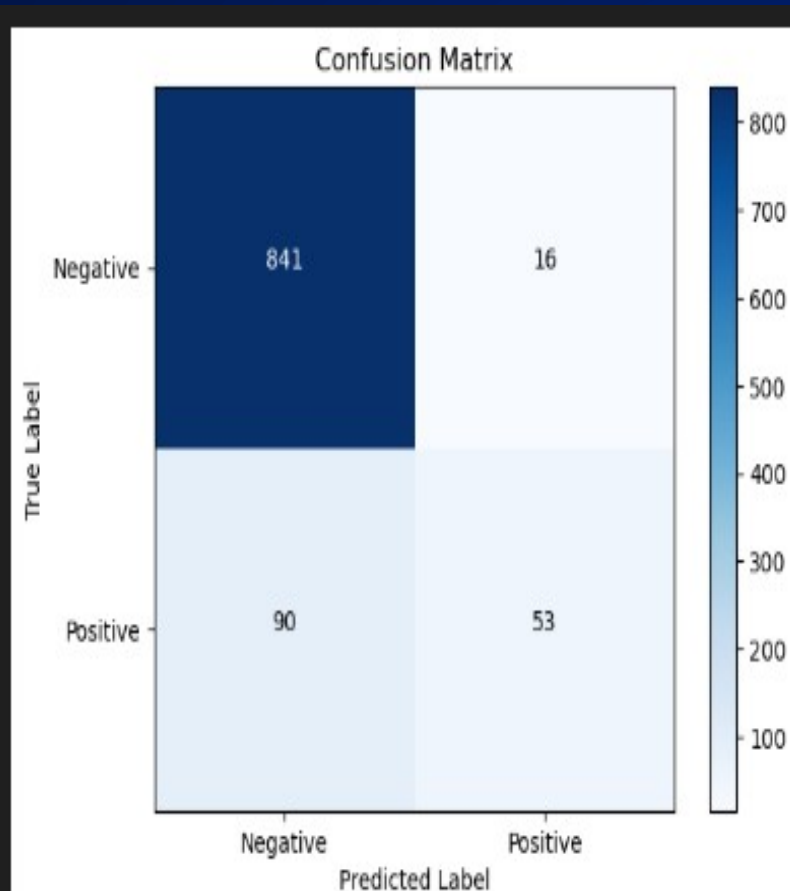
- Reduce customer churn rate
- Increase customer retention and loyalty
- Improve customer satisfaction and overall profitability

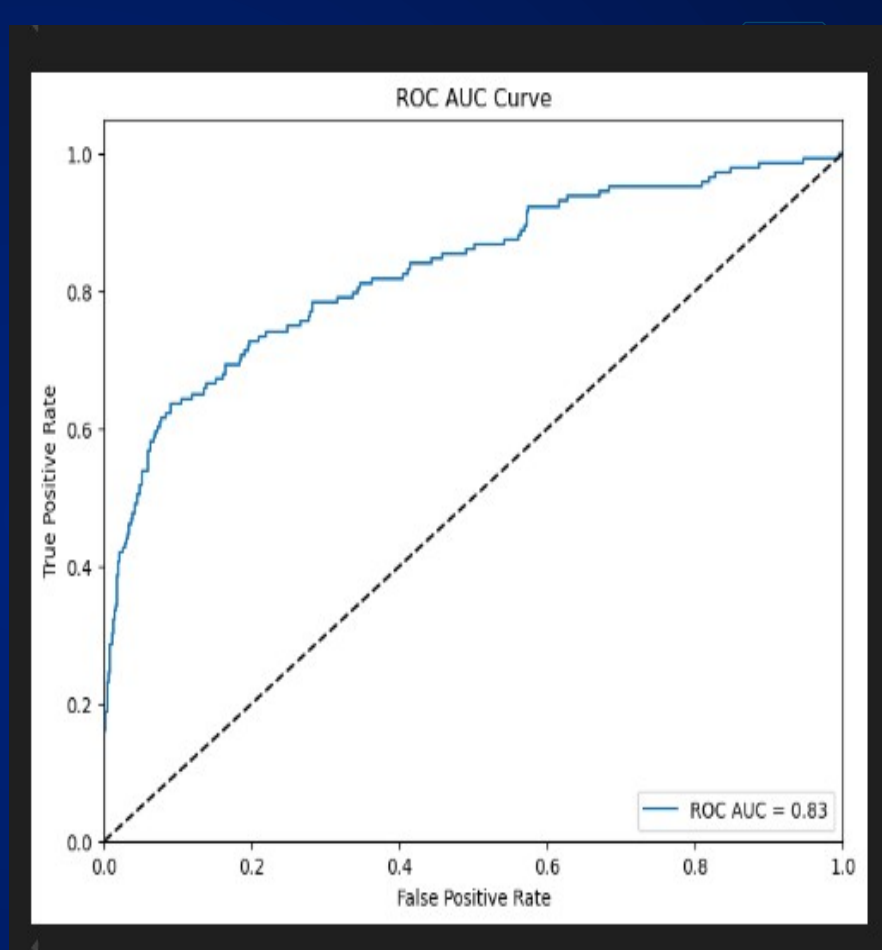
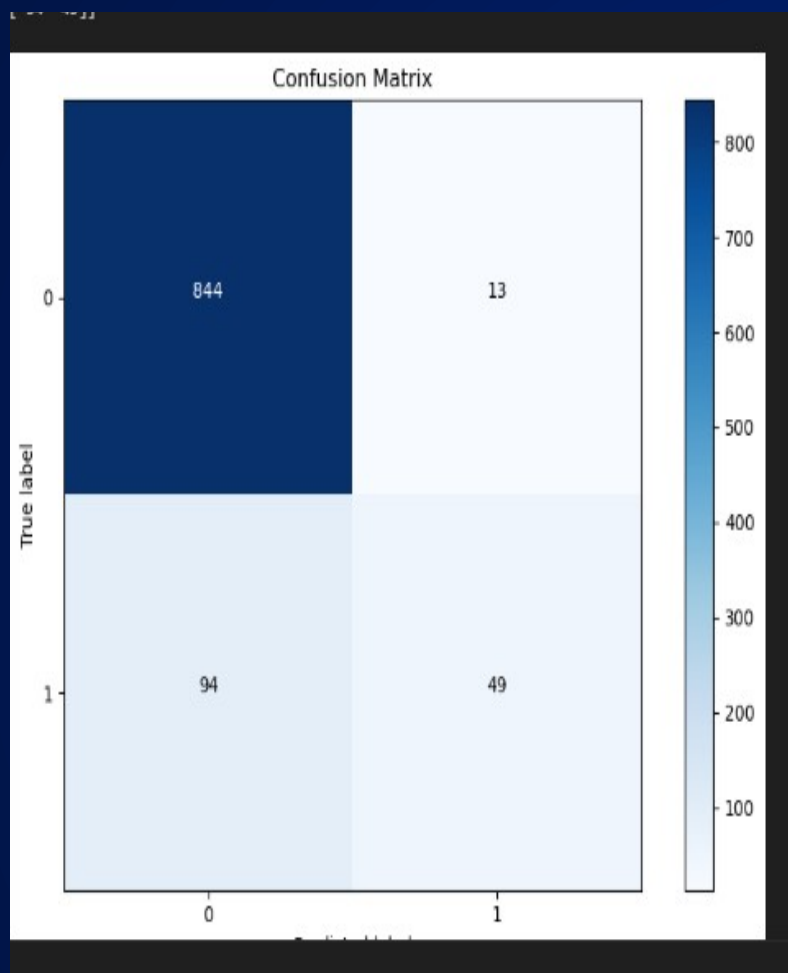
DATA UNDERSTANDING

- The data used is the "Churn in Telecom's dataset" from Kaggle, which contains customer information and behavior data along with a target variable indicating whether the customer churned or not
- The primary goal of this analysis is to identify the best-performing classifier(s) for predicting customer churn in the telecommunications industry.
- The insights gained from this analysis can help service providers take proactive measures to retain customers by targeting those at high risk of churning and offering them incentives or improving their service experience.



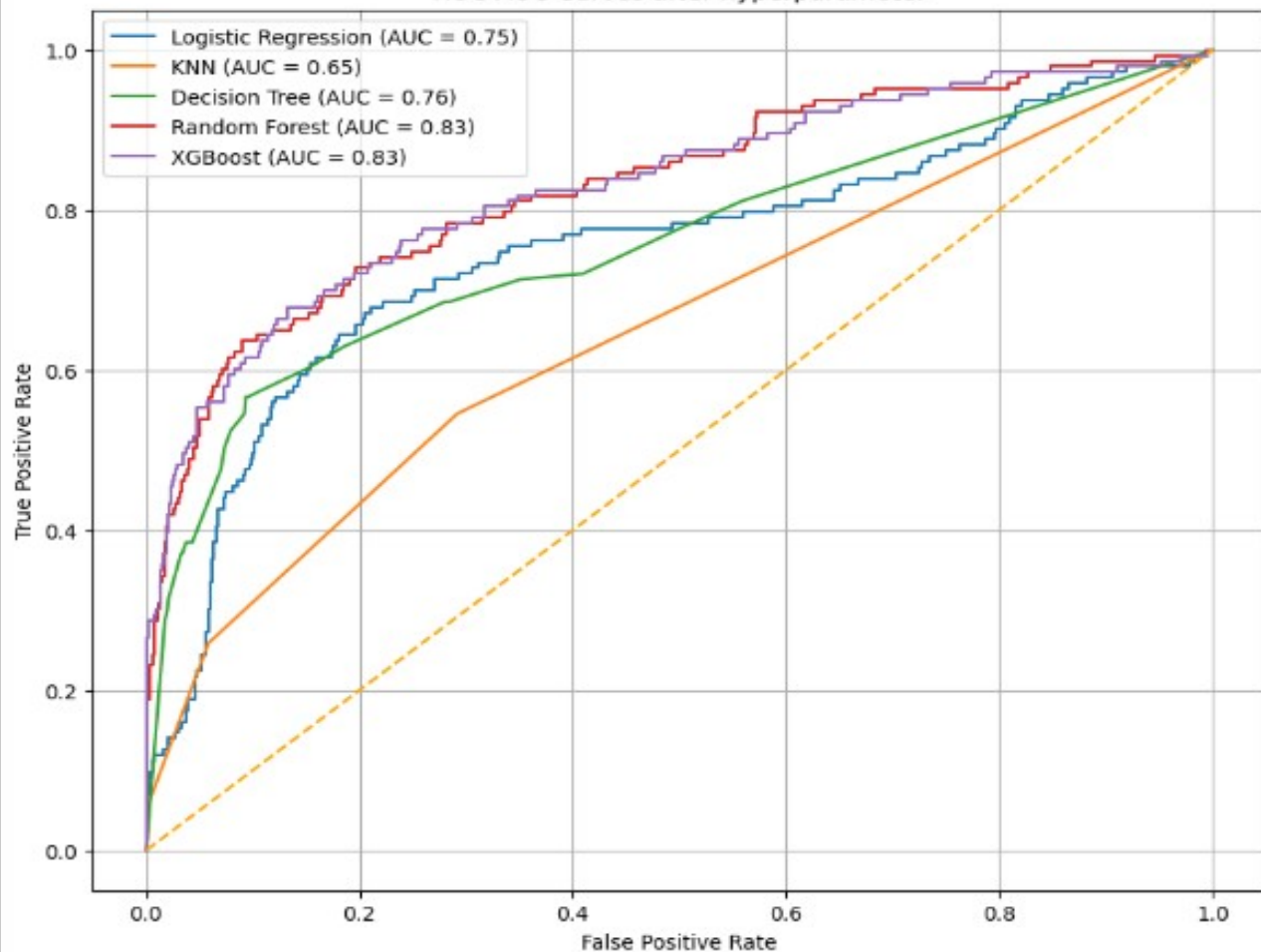
MODELLING





- The above plots show the accuracy of the two best models using a confusion matrix and an ROC AUC curve
- The random forest classifier and the XGBOOST
- XGBOOST has an accuracy of 89% while the Random forest has an accuracy of 85%
- We deployed the XGBOOST as it was the best model of the two

ROC AUC Curves after hyperparameter



- The plot above is just a comparison and representation of the accuracy scores of all the models.
- The accuracy of the models is as follows from the least accurate to the most accurate:
- KNN Model , Logistic Regression, Decision Tree , Random Forest and the XGBOOST.
- Below you'll see a table comparing the accuracy scores of the models

..

	Model	Accuracy	Precision	Recall	F1 Score	Sample Size
0	Logistic Regression	0.84	0.64	0.56	0.58	1000
1	KNN	0.84	0.65	0.60	0.62	1000
2	Decision Tree	0.88	0.77	0.67	0.71	1000
3	Random Forest	0.89	0.85	0.66	0.71	1000
4	XGBoost	0.89	0.84	0.68	0.72	1000

RECOMMENDATIONS

- ● Optimize subscription plans, especially for international and voice mail services, to ensure competitive pricing, quality, and attractive features, retaining customers on these plans.
- Develop customer loyalty programs with tailored incentives and benefits based on tenure to enhance satisfaction and reduce churn among long-term customers.
- Monitor call usage patterns, including duration, frequency, and international destinations, to identify abnormal behavior as an early warning sign for potential churn, enabling proactive retention efforts.

NEXT STEPS

- Develop a user-friendly interface or dashboard for marketing and customer service teams to access model predictions and take appropriate retention actions.



The background of the slide features a dark blue color with a network of light blue lines and dots, resembling a circuit board or a digital network. These lines originate from the bottom and left sides, extending upwards and outwards across the slide.

THANK YOU!!!!

For more information you can reach
out to me on my personal email:

novatrading.edu@gmail.com