CUSTOMER CHURN PREDICTION

BY: TRACY GWEHONA

OVERVIEW

- Customer churn (or the loss of customers) is one of the biggest challenges for telecommunications companies because acquiring new customers can be much more expensive than keeping them. High churn rates affect profitability, market share, and long-term expansion.
- I aim to use historical customer information to anticipate who is destined to exit. If the company knows the churn pattern, it will be able to apply proactive measures to increase retention and customer satisfaction and keep up with the competition.

BUSINESS PROBLEM

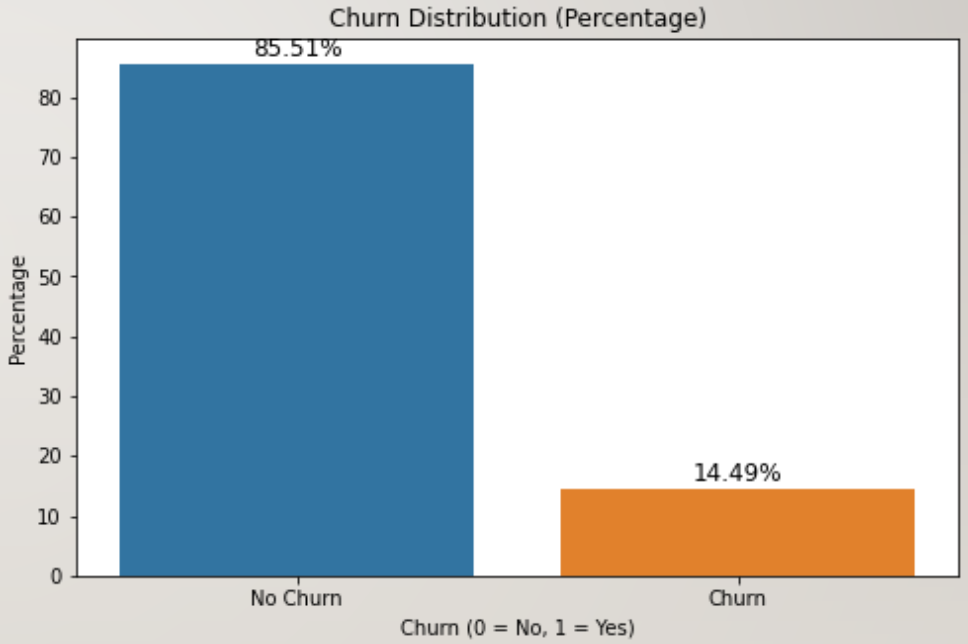
- SyriaTel is having customer churn which is directly impacting revenue and profit.
- The company can't develop retention plans without understanding churn and high-risk customers, leading to continued customer loss and a decline in market share.
- There is a need to predict customers likely to leave.

OBJECTIVES

1. Predict customer churn.
2. Identify key factors influencing customer churn.
3. Provide actionable insights to enable SyriaTel to implement targeted interventions.

DATA UNDERSTANDING

- The dataset contains information on customer behavior, usage metrics, and plan details to predict **customer churn**.
- **Target variable**: churn.
- Other features include account length, international plan, voice mail plan, number of voice mails, total minutes, calls and charges, and customer service calls.



MODELING

I tested two models:

1. Logistic Regression Model
2. Random Forest Classifier

EVALUATION

Logistic Regression results

- Accuracy = 72%

The model correctly predicted 72% of the cases overall

- ROC-AUC score = 0.8052

It is closer to 1 (ideal) and suggests the model performs well at ranking predictions.

Random Forest results

- Accuracy = 95%

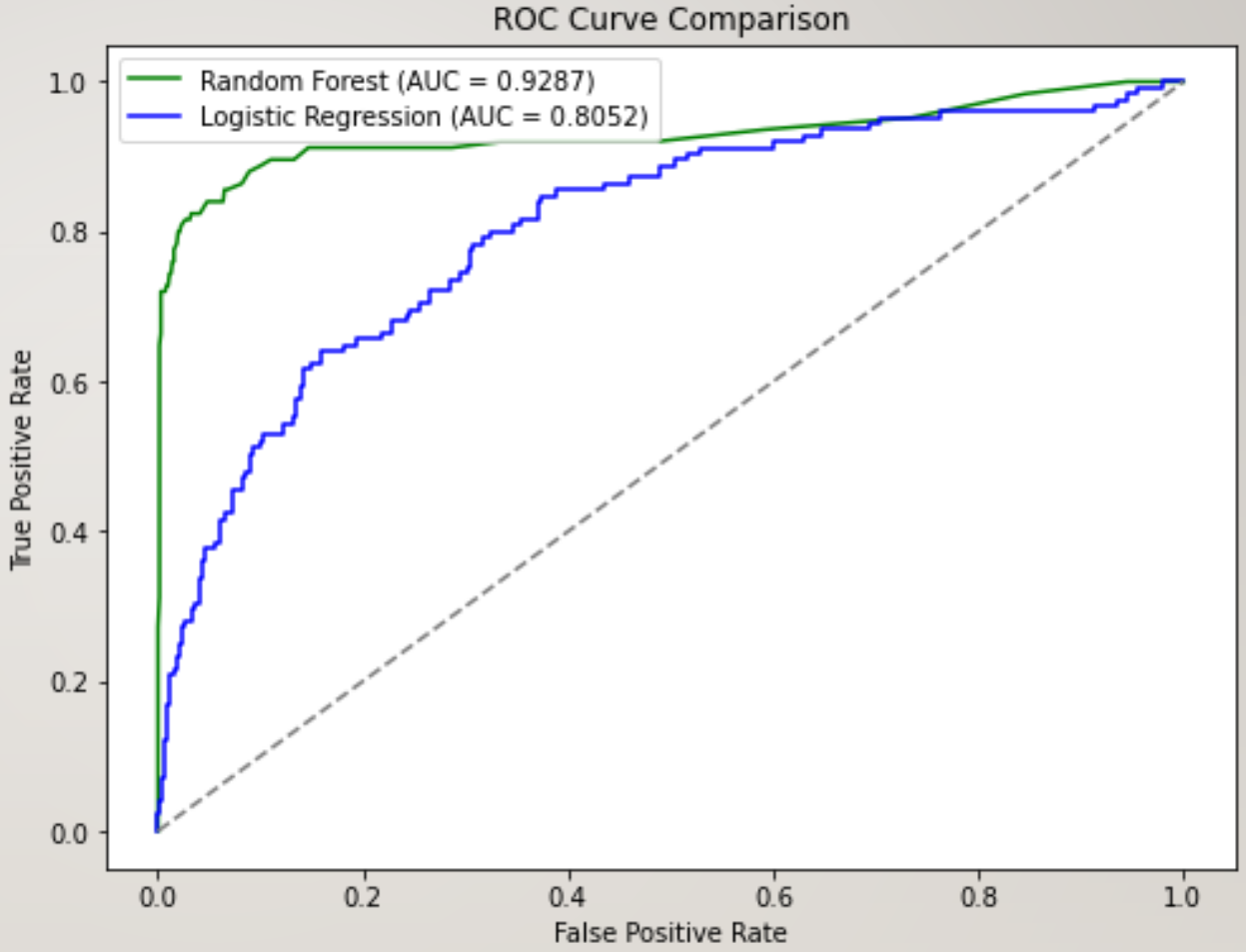
The model correctly predicts whether a customer will churn in 95% of cases.

- ROC-AUC score = 0.9287

Indicates excellent model performance.

EVALUATION

The closer the ROC curve is to the top-left corner, the better the model's performance.



CONCLUSION

- **Logistic Regression** performs adequately on non-churn cases but struggles with churn detection.
- **Random Forest**, on the other hand, performs significantly better than logistic regression for churn detection

RECOMMENDATIONS

- For **better performance** on both classes, **Random Forest** is the better model, especially if your goal is to achieve a more robust performance for detecting churn
- If **interpretability and computational efficiency** are key, Logistic Regression might be preferred, but additional statistical techniques could help improve its performance.

NOTE: However, for churn prediction in this case, Random Forest should be prioritized, especially when dealing with imbalanced data.

RECOMMENDATIONS

- On the features that affect churn:

1. Focus on Daytime Users: Monitor customers with high daytime usage and proactively offer support or incentives.
2. Improve Customer Service: Address issues related to customer service interactions to reduce churn.
3. International Plans: Evaluate and improve international plan offerings to reduce dissatisfaction.

NEXT STEPS

- Deploy the Random Forest model in production.
- Monitor model performance over time.
- Further refine the model with new data.
- Develop a customer retention strategy based on predictions.

THANK YOU!

Email: tracy.gwehona@gmail.com

GitHub: @tracy-gwehona