FINAL PROJECT SUBMISSION

- Student name:Muli Brian
- Students pace: Full time
- Scheduled project review date/time:24 /05/2023
- Instructor name:Lucille Kaleha, Simon Karu, Nikita Njoroge

BUSINESS UNDERSTANDING

INTRODUCTION

In today's highly competitive telecommunications industry, customer retention is of paramount importance for sustained business success. Customer churn, which refers to the phenomenon of customers discontinuing their services with a company, can have significant financial implications and impact the overall growth and profitability of a telecommunications company. Understanding the factors that contribute to customer churn and developing an effective predictive model to identify potential churners is crucial for proactively addressing customer retention strategies and implementing targeted retention initiatives.

Business Problem

The business problem is to reduce customer churn and retain valuable customers for SyriaTel. By addressing the following questions, the telecom business can develop effective strategies to mitigate churn:

1.What are the key factors or patterns that contribute to customer churn in SyriaTel? Can we predict, with reasonable accuracy, which customers are likely to churn in the near future?

2.How can SyriaTel proactively identify and target customers who are at high risk of churning?

3.What actions or incentives can be offered to customers at risk of churn to increase their loyalty and retention?

4.How can SyriaTel allocate resources effectively to customer retention initiatives and minimize the financial impact of churn?

PROBLEM STATEMENT

The problem at hand is to develop a predictive model that can accurately identify customers who are likely to churn from SyriaTel . By analyzing historical customer data and identifying patterns or indicators of churn, the goal is to build a classification model that can predict whether a customer is likely to stop doing business with SyriaTel in the near future.

MODEL INTERPRETATION

We build several classification models to classify whether our customers will churn or not.Some we found to be well performing and others were poor in accurately predicting whether a customer will churn on not.

MODEL EVALUATION

Our best performing model was a Logistic Model which produced an accuracy of 89% and a recall of 1.0 this shows the model predicted all the positives accurately but their is a limitation to this as the classes were imbalance.

After balancing the classes we build another model of Random Forest and this acquired an accuracy of about 67% and a recall of 0.77 this signifies that the model was able to identify the positives 77% accurately thus it is good for reducing the False Negatives.

The other model we build with balanced classes was a DecisionTree Classifier Model this performed quite moderately getting an accuracy of about 54% and a recall of 0.62 which shows that it identified the positives at an accurate rate of 62% which is a moderate performance with a loy of room for improvement.

Our poorest performing model was a KNN Model which achieved an accuracy of 49% and a recall of 0.37 which translates to 37% correctness in identifying our positives we can therefore conclude that this model is not good for the prediction and classification we are trying to do.

INSIGHTS

1.We looked at the churn level by area code and we came to a realization that Most customers who churn are from area 408 followed by area 510 and area 415 has the lowest churn level.

2. We also checked at the distribution of voice mail messages by area code and we realised that area code 415 has the highest amount of voice mail messages does this mean this might mean the higher the voice mail messages the lower the churn level.

3.When it comes the number of international calls area code 415 still has the highest number of international calls which translates to higher revenue for the company and we can also say that the number of international calls in an area will affect the churn rate .

CONCLUSIONS

The project was aimed at determining whether a customer will churn given multiple factors like the international plan,area code ,state ,voice mail plan and other factors.

Firstly, we performed exploratory data analysis (EDA) and found the churn of a customer may have a relation to the area code the customer is in and the area code may also affect whether a customer will have international plan or not.We also noticed that West Virginia state has the highest number of customers.

We then developed several classification models to check whether the customer will churn given certain factors.Our main evaluation metrics were accuracy and recall as we were trying to reduce the number of False Negatives.Some models performed well others were moderate and others were poor as you can see from the model interpreatatio above.

RECOMMENDATIONS

1. If the company wants to improve its international calls revenue,it should focus more on the area with the most international calls for successful campaigns.

2. The company should focus on minimizing the number of customer service calls as the more the customer service calls the higher the churn level this will be done by putting inplace more efficient and fast way of assisting customers in addressing their issues instead of calling.

3. Looking at the distribution of customers according to state ,the company should target certain campaigns for example if the campaign is focused on customer retention,it should do it in states with most customers like West Virginia and if the campaign is aimed at customer acquisition it should be done in states with few customers like Virginia.

NEXT STEPS

Further research and modelling can be done to improve the accuracy of and recall of the models and get optimal predictions and a better classification of our data.