SYRIATEL TELECOMMUNICATION COMPANY CUSTOMER CHURN PREDICTION

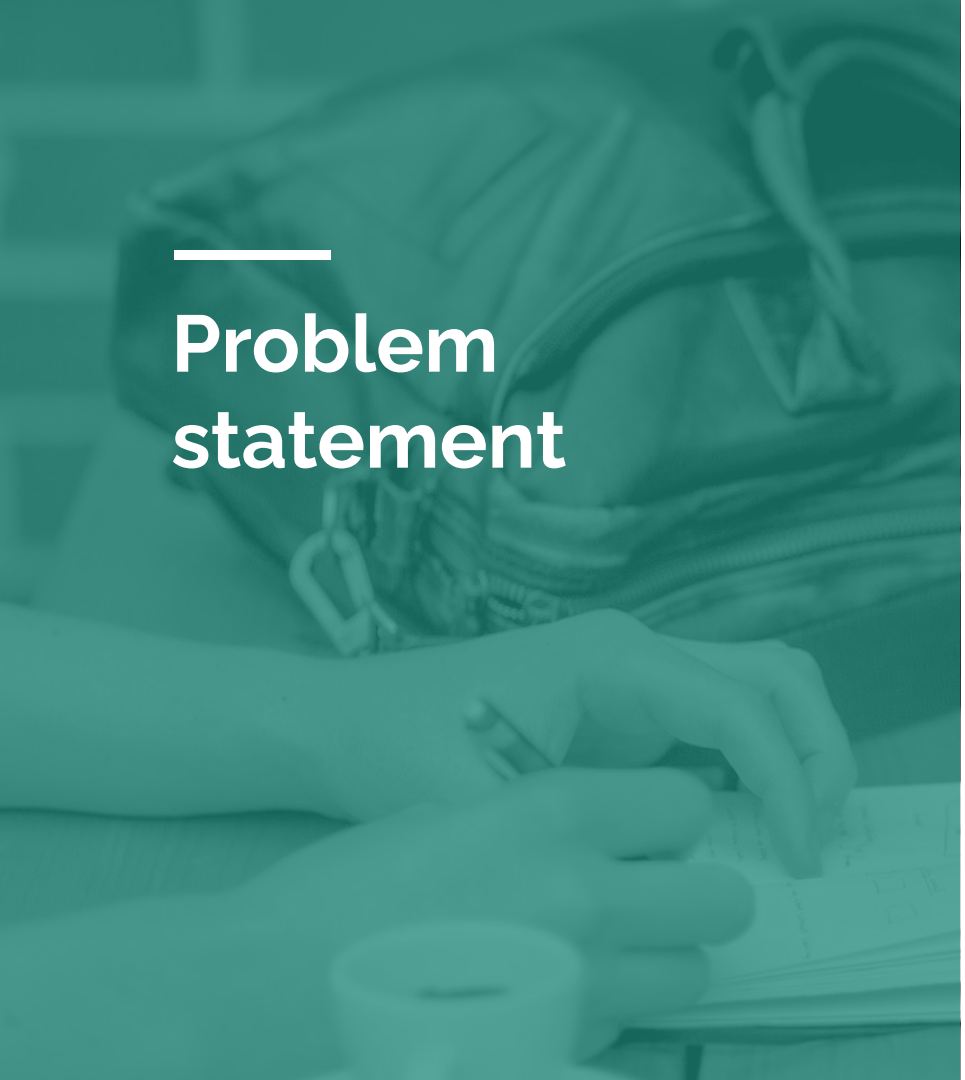
Phase 3 project
May 2023

Business overview

- The telecommunication industry is one of the most competitive industries nowadays thus the major problem that they all face is customer churn.

- Since losing customers would result in a large loss of revenue for the business,it is imperative to identify those clients who are likely to quit the organization in the near future in advance.

- Accurately predicting churn can lead to higher customer retention rates, increased market share and improved business performance.



Problem statement

- SyriaTel telecommunications company wants to take the required actions to retain customers in order to stabilize their market value .

- This is because the cost of recruiting new customers is significantly higher than the cost of maintaining existing customers and also in order to reduce losses incurred from customer churn.

Objectives

Main objective:

- To predict customer churn using a classification algorithm model

Specific objectives:

- To do exploratory data analysis on the data
- To fit different classification algorithm models to determine which one works best for churn prediction
- To select the best model
- To make predictions using the selected model
- To check the accuracy of the predicted variables

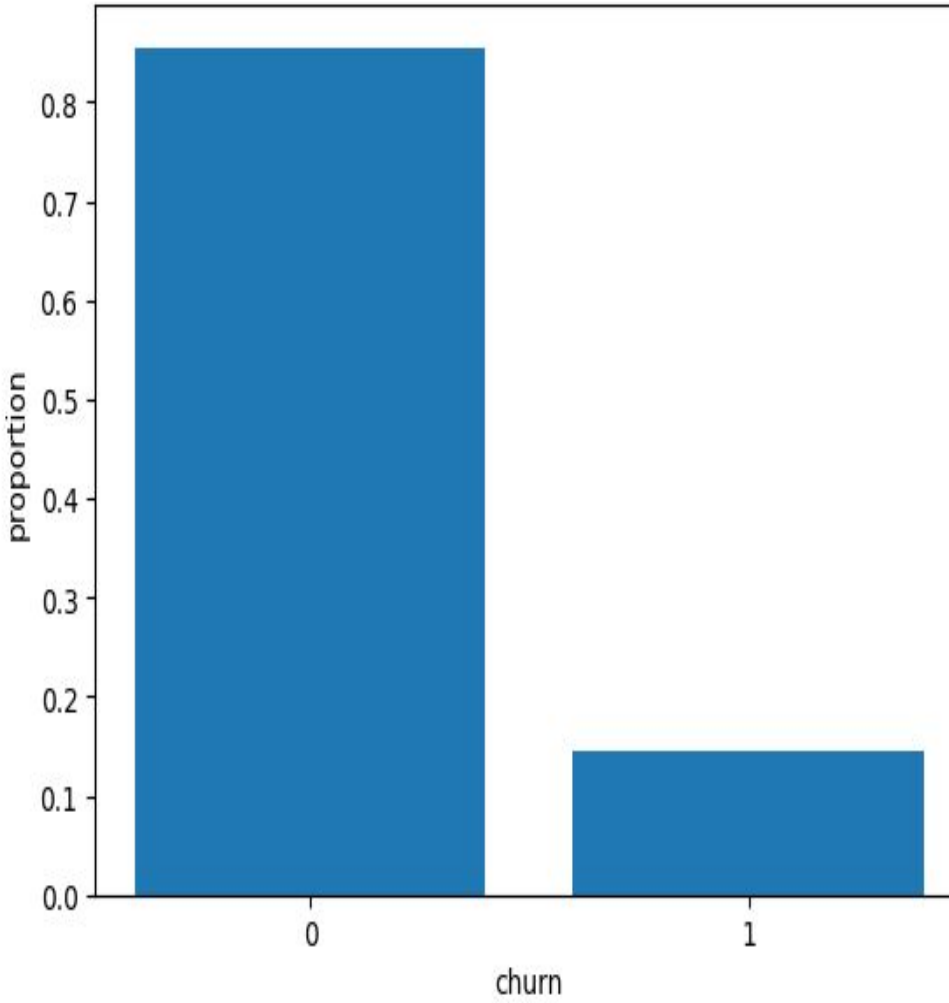
Data understanding

- This data is from syriatel telecommunication company and it was obtained from kaggle.
- The data has 21 columns and 3333 rows, each row representing a customer.
- The target column, churn, is a bool column where True means the customer did churn and False means the customer did not churn, making this a binary classification problem.

Data Analysis:

Distribution of the target variable, churn.

From the plot, we can see that most of the customers in this data did not churn.



Modelling

- Since this data was a binary classification problem, different classification models were fit to the data and the model with the highest accuracy was selected.

- The classification models that were used were; logistic regression, K-nearest neighbors , decision trees and random forest.

- The best model in terms of accuracy was the decision tree classifier.

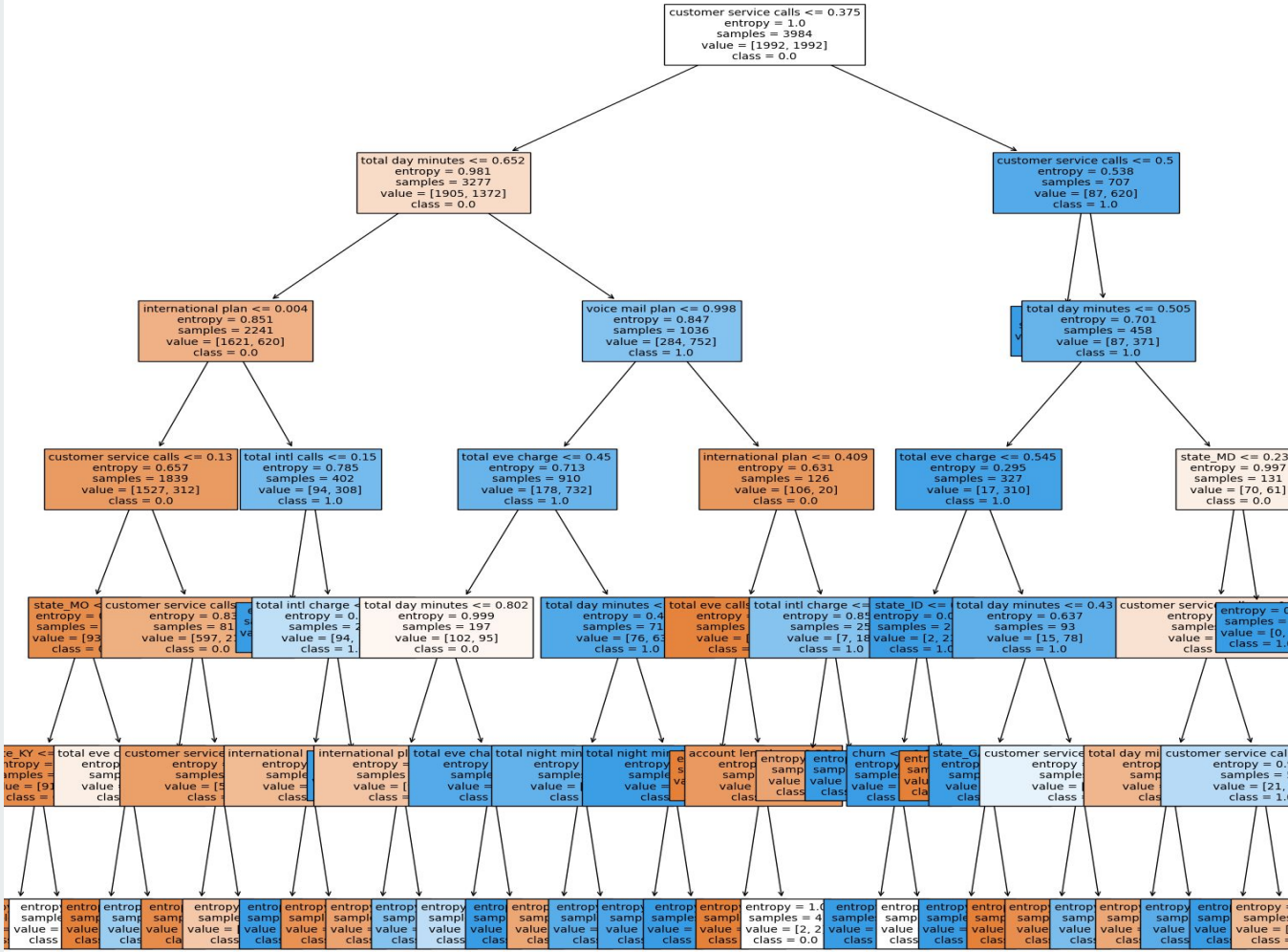
Modelling

The features important to the selected model were:

- customer service calls
- total day minutes,
- International plan
- voicemail plan
- total eve charge and calls
- total intl charge and calls
- total night minutes
- state
- account length

Modelling

Decision tree:



Model evaluation

The selected model was used to predict customer churn and the accuracy of the predicted values was assessed .

The table shows the metrics of the model.

Metric	**Score**
Accuracy	92.6%
Precision	72.3%
Recall	77.5%
F1- score	74.8%

Conclusions

- ➔ Wv state seems to be having most of the customers and CA has the least customers.
- ➔ Most of the customers don't have an international plan and also most of them don't have a voicemail plan, explaining why the majority of the data in the number of voicemail messages is 0.
- ➔ The most influential predictors were; customer service calls, total day minutes, International plan, voicemail plan, total eve charge, total intl charge, total intl calls,total night minutes, total eve calls, state LA, state MS, state NJ, state ME,state IL, state HI , state AL and account length(They are listed according to their impact in descending order)
- ➔ The model has 92.6% accuracy thus the model is suitable for churn prediction

Recommendations

- Focus on training customer service representatives to handle customer issues efficiently and provide proactive support to enhance customer satisfaction and loyalty.
- Offering discounts and offers to customers likely to churn.
- Most customers do not have an international plan or voicemail plan. The company should consider promoting these services to customers who currently don't have them but may benefit from them.
- The company should consider focusing retention efforts in states with more customers and marketing campaigns on states with fewer customers.
- Consider reaching out to loyal customers with special offers or personalized assistance to address their concerns and improve their experience.

Challenges

- Due to the presence of high multicollinearity among several predictor variables, we had to remove some of those columns from our analysis.

- The model has some limitations, since decision trees can be biased towards the majority class if the dataset used for modeling has imbalanced class distribution, we had to use SMOTE to solve the class imbalance problems thus any new data fed to the model has to undergo the same preprocessing technique

Future work

- While the model currently has good accuracy and performance, we should continue monitoring and evaluating its performance on new data. As customer behaviors and preferences change over time, it's important to ensure that the model remains effective and up-to-date.

- Exploring advanced techniques like ensemble methods, gradient boosting methods, XGboost and Adaboost , or deep learning to further improve churn prediction performance.

Questions?

Thankyou

Name: Sharon Wanjiku Kimani

Email: Kimanisharon23@gmail.com

Linkedin: Sharon kimani (https://www.linkedin.com/in/sharon-kimani-b4b84624b)

Github:(https://github.com/Sharonkimani)