Literature Review

Contents

Churn Prediction

Problem definition:

Customer churn is defined as when customers or subscribers discontinue doing business with a firm or service. Customers in the telecom industry can choose from a variety of service providers and actively switch from one to the next. The telecommunications business has an annual churn rate of 15-25 percent in this highly competitive market. Because most businesses serve a big number of clients and cannot afford to spend much time with each one, individualized customer retention is challenging. The added money would not balance the expenditures, which would be too high. However, a business might target its customer retention efforts only on these "high risk" consumers if it could foresee which clients are most likely to depart. The ultimate objective is to increase client loyalty and the company's coverage area. The client themselves is the key to success in this business.

Objectives:

- Analyzing the data in terms of various features responsible for customer churn.
- Finding a suitable model for churn prediction with highest accuracy.
- Churn prediction.

Dataset:

- The Customer Churn table contains information on all 7,043 customers from a Telecommunications company in California in Q2 2022
- Each record represents one customer, and contains details about their demographics, location, tenure, subscription services, status for the quarter (joined, stayed, or churned), and more!
- The Zip Code Population table contains complimentary information on the estimated populations for the California zip codes in the Customer Churn table.

Our findings:

We trained our model using four different machine learning algorithms; Random forest classifier, decision tree classifier, logistic regression, and Knn(nearest neighbors).

Upon building the model using these four algorithms these were the results:

- Decision tree model showed an accuracy of 0.80181
- Random forest model showed an accuracy of 0.82112
- Logistic regression model showed an accuracy of 0.7961
- Knn(nearest neighbor) model showed an accuracy of 0.77455

After these findings we tried another approach, rather than multi class we targeted binary class which showed different accuracies.

Here are the accuracies after trying this approach:

- Decision tree model showed an accuracy of 0.8598300
- Random forest model showed an accuracy of 0.8725728
- Logistic regression model showed an accuracy of 0.834336
- Knn(nearest neighbor) model showed an accuracy of 0.786407

From these findings, we can conclude that Random forest with targeted binary class gives the highest accuracy of 0.8725728, while Knn(nearest neighbor) gives the lowest accuracy of 0.786407.

While comparing our findings to other contributors it was found that:

One contributor used the following machine learning algorithms.

1 -Random Forest Classifier
2 – Logistic Regression
3 – Gaussian NB
4- Decision Tree Classifier
5 - XGBClassifier

And the best accuracy concluded was 0.8066

Another contributor used the Ada boost classifier machine learning algorithm and got an accuracy of 0.85

Another contributor used the gradient boosting classifier and got an accuracy of 0.84

Reference list:

https://www.kaggle.com/code/hiimanshuagarwal/customer-churn-prediction#Data-Model-Building

https://www.kaggle.com/code/davideluise93/telecom-churn-prediction#Model-Comparison-and-Final-remarks

https://www.kaggle.com/code/aviraljain58/telecom-customer-churn#Model-Comparison-and-Final-remarks