PREDICTION OF CUSTOMER CHURN IN TELECOM INDUSTRY

Akshay Shaju (22152563), Caroline Vincent (22153926), Karthika Nair (22105522)

National College of Ireland, Dublin

Abstract – **The telecom industry is one of the fastest industries in the world which is adjusting and adapting to all the modern era innovations which are being launched. It is of prominent importance to retain customers than acquiring fresh customers for these telecom industries because of complex competition and customer churn, which then might significantly impact their profitability and revenue margins. The aim of this project is to use advanced machine learning algorithms and forecast customer churn with maximum accuracy and discover key factors influencing it. By doing this, we aim to improve our understanding on customer retaining strategies and minimize losses.**

Keywords – **Customer churn, Gradient boost, customer retention, telecom, Predictive Analysis**

1. INTRODUCTION

In this dynamic era of technological giants, the telecom industries are having a very tough time to follow up and retain their customers as people intent to switch from one company to another within a very short period. This is affecting their revenue streams and is also making them spent an additional budget for acquiring new customers. Hence, predicting the factors underlying customer churn and minimizing it has become a prominent priority for the telecom firms.

This project is focusing on solving this problem faced by telecom companies by making a customer churn prediction system using the power of advanced machine learning algorithm. We are aiming this prediction system to have the potential to predict the churn with high value accuracy and efficiency and find the major weak reasons that are making people feel decide to depart from their beloved companies. By the successful implementation of this model, our goal is to build a dynamic strategy to retain the customers, revenue protection, increasing profitability and making new investments to maintain the customer satisfaction to maximum level. For this project, we are using a delicate and quality customer churn telecom dataset and employing gradient boosting algorithm in order to attain very accurate results. The dataset contains various customer usage patterns and demographics of customer choices. As the increasing customer churn is having negative effects on the company's reputation and market ranking, it has become a matter of paramount importance for them to find a solution for this and maintain their position in this most edge-competitive field to challenge other big companies as customers have always got a full bag of options to switch from one to another.

2. LITERATURE REVIEW

In telecom sector, customer churn is a major concern because it is more economical to keep existing customers than to acquire new ones. Due to the potential to support telecom businesses in identifying and retaining at-risk consumers, predictive modelling for customer churn has attracted a lot of attention. This literature review contains main studies and developments in predicting customer turnover in the telecom industry.

The literature [1] emphasizes the fundamental importance of customer churn analysis and forecast in the telecom sector, due to its crucial role in maintaining the customers. As the competitions in this sector increases, understanding customer behaviour is essential for telecom business to prevent future subscription cancellations. In the competitive environment of today's businesses, modern data mining techniques and sophisticated algorithms have become crucial tools for predicting client churn. The main topic of this paper is the use of machine learning models for the churn prediction. The study aims to contribute considerably to the customer retention efforts of the telecom sector by offering insights into effective tactics for forecasting and reducing customer churn through a comparison of machine learning models.

Customer retention has become a top priority for telecom firms due to the rapid improvements in the industry and strengthening rivalry among operators. These businesses have created a variety of predictive tactics to anticipate churn and its implications. When used in actual situations, however, existing churn predictions do have some drawbacks. The study [4], which is specifically adapted for the telecom sector, fills the gap. The work contributes by improving churn prediction solutions through the suggestion of the Half Termination Dynamic Label in conjunction with the XGBoost algorithm. The effectiveness of the solution which greatly outperforms conventional approaches in F1-score-based churn prediction, is demonstrated by empirical results. Thus, this study represents an important step toward improving customer churn prediction tactics in the telecom sector.

The paper [2], examines the use of machine learning models within the telecom industry, ranging from prediction of customer turnover to broader field of customer relationship management. The study explores how machine learning methods help telecom industry overcome the difficulties, with a focus on the IEEE paper on Machine Learning and Applied Network Technologies. By analysing the predictive nature of the churn and its implications for enhancing the customer relationships, the article gives ideas into the shifting environment of customer retention. The study proposed valuable viewpoints on the incorporation of machine learning for improving operational approaches in the telecom sector, helping as a reference for the research who is in search of a comprehensive understanding of the advancements in telecom industry.

3. METHODOLOGY

This section will emphasize on the techniques that have been taken to perform model building for the analysis of customer churns in the telecom industry. The purpose of is to successfully build a prediction model that generates a complete overview of the patterns and factors that could lead a customer to discontinue or move over to another telecom service provider.

a) *Data Collection*

In this study, by thorough investigation conducted on collection of data, we have considered obtaining the dataset from Kaggle. This dataset contains key components or information of customers registered with the particular telecom company. These key factors include customer demographics, the type of contract the customer has undertaken, pattern of usage by the customer and their tenure with the company.

b) *Data Integration*
For an effective analysis of this study, the modelling should be carried out on a rather larger dataset. Through this, we can observe a wide range of patterns and information that help evaluate the potential cause of churning in the company. To achieve this, Python libraries have been used to combine the multiple datasets obtained from Kaggle.

c) *Data Preprocessing*
After collecting the dataset, it is important to furbish the quality of the data. This maximizes the efficiency of analysing the data. Preprocessing of data includes elimination of null values which is essential as its presence can cause variations in the analysis. Using Python libraries, null values in columns with numerical data have been replaced with the mean value of the entire column. Accordingly, the categorical data column with null data is substituted with most frequently occurring value in that column.

```
Total night charge        0
Total intl minutes        0
Total intl calls          0
Total intl charge         0
Customer service calls    0
Churn                     0
```

Fig. 1. Snippet of data after removal of null values

d) *Data Encoding*
For compatibility of the algorithm with the dataset, transformation of categorical variables is needed as some machine learning algorithms require numerical variables only as their input. The processing of this encoded data helps develop meaningful predictions.

The data under "Churn" variable contains categorical information. It indicates whether the given customer has decided to discontinue the telecom service or not and hence, the values are of type True or False. We have performed label encoding to denote True values as 1 and False values as 0.

Total intl calls	Total intl charge	Customer service calls	Churn
4	2.35	1	0
6	3.43	4	1
9	1.46	4	1
6	2.08	2	0
1	3.00	1	0

Fig. 2. Values after performing label encoding on "Churn" variable.

e) *Data Visualisation*
Visualisation of data helps in identifying trends and highlight the features that have higher potential to influence the prediction of churning. In this study, visualisations have been shown in the form of bar graphs and pie charts.

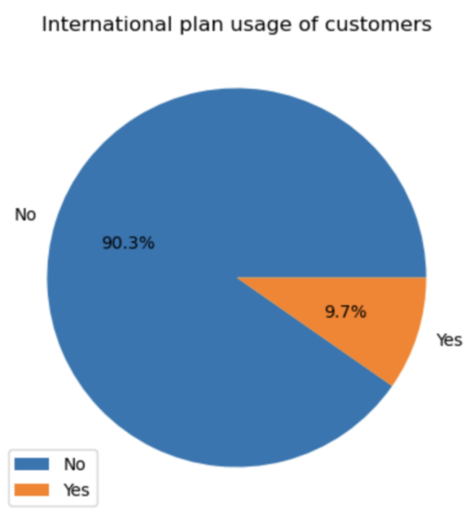

Fig. 4. International plan usage of customers

The Fig. 3 shows a pie chart depicting the percentage of customers using an international plan. 9.7% of the customers use an international plan.

Fig. 4 shows a bar graph representing a total of churn count corresponding to the plan the customer uses.

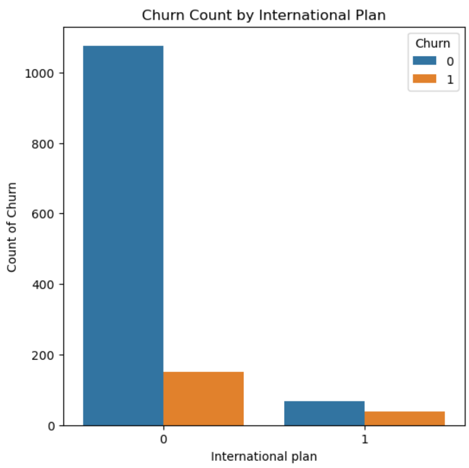

Fig. 5. Churn count by international plan

f) *Model Building*

Generating a prediction model to create accurate predictions on customer churning is the main goal of this study. We have focused on using Gradient Boosting Classifier as an algorithm to train our model.

The algorithm is fed with the training data and the training process is initiated. The internal parameters are tweaked so that the algorithm learns the data and makes accurate predictions accordingly. Through the training data, the relationships between the target variable and the input features is identified by the algorithm.

g) *Hyperparameter Tuning*

The algorithm might not be able to learn some parameters during the training. These parameters are called hyperparameters and they can be set before the training starts. Learning rate is an example. We have hyper-tuned the algorithm and the results will be mentioned under the results and evaluation section.

4. Results and Evaluation

A classification report is generated that gives information on the precision, recall, F1-score and accuracy.

This report helps us understand whether the model fitted will provide accurate predictions on the customers churning. With higher precision and accuracy values, it can be concluded that the model chosen is the ideal one and can be used for further analysis.

The figures below show the classification report before and after hyper-tuning.

	precision	recall	f1-score	support
0	0.96	0.99	0.97	854
1	0.91	0.74	0.82	146
accuracy			0.95	1000
macro avg	0.93	0.86	0.89	1000
weighted avg	0.95	0.95	0.95	1000

Fig. 6. Classification report before hyper-tuning.

The accuracy score is 0.95 which indicates a good prediction potential. The precision and recall values are 0.96 and 0.99 respectively.

	precision	recall	f1-score	support
0	0.96	0.98	0.97	854
1	0.89	0.74	0.81	146
accuracy			0.95	1000
macro avg	0.92	0.86	0.89	1000
weighted avg	0.95	0.95	0.95	1000

Fig. Classification report after hyper-tuning.

The classification report after hyper-tuning suggests that there is not much difference when compared to the report recorded before hyper-tuning. In other words, the model chosen concludes to be the ideal model in this case to predict precisely the customer churns.

5. CONCLUSION

To conclude, in the project we have used the potential of data analysis and machine learning algorithm to solve a major concern that was affecting the revenue of the telecom companies that is customer churn. As we have already seen, using gradient boosting algorithm, we managed to attain exceptional results and most importantly, the overall accuracy and f1 score of 0.96 and 0.97 respectively shows the proficiency and the balance of the model in making correct forecasts.

The churn prediction system has a potential to build high business values to the company's revenue margins by minimizing churn and implementing modified retention strategies, thereby, maintaining customer satisfaction to a good level. The valuable insights we managed to achieve from this project will help to mitigate the factors affecting the churn and help companies maintain their reputation in this era of competitive business and find new innovations for customer dynamics.

6. FUTURE WORK

There are several opportunities to make improvements to our project and enhance the prediction system and impact the business in a more positive manner by attaining more accurate results.

a) Exploring more complex algorithms like Neural network or other advanced ensemble methods could help us get more accurate values and thus have a better understanding about the customer business patterns.

b) Involving real time data instead of preset data will help to achieve super accurate results, as the prediction system could rapidly adapt to updated changes of the customer dynamics and make new decisions.

c) Also building a platform to compare the present evaluations of the performance of the model with the real time data and integrating social media activities of the customers and their usage and shopping patterns will help the prediction system to evolve through the new choices and make far more better decisions.

7. REFERENCES

[1] A. Gaur and R. Dubey, "Predicting Customer Churn Prediction in Telecom Sector Using Various Machine Learning Techniques," 2018 International Conference on Advanced Computation and Telecommunication (ICACAT), Bhopal, India, 2018, pp. 1-5.

[2] F. Alhaqui, M. Elkhechafi and A. Elkhadimi, "Machine learning for telecoms: From churn prediction to customer relationship management," 2022 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT), Soyapango, El Salvador, 2022, pp. 1-5.

[3] P. Bhuse, A. Gandhi, P. Meswani, R. Muni and N. Katre, "Machine Learning Based Telecom-Customer Churn Prediction," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp. 1297-1301.

[4] Y. Zhang et al., "Telecom Customer Chum Prediction based on Half Termination Dynamic Label and XGBoost," 2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Wuhan, China, 2022, pp. 1563-1568.

[5] L. Liu and H. Ding, "Modeling China Telecom customer churn prediction based on CRISP_DM," 2011 International Conference on E-Business and E-Government (ICEE), Shanghai, China, 2011, pp. 1-4.