Telecom Customer Churn Prediction Using Machine Learning.

Amarachi Ezeji
Department of Business
Humber College Institute of Technology and Advanced Learning
Toronto, ON, Canada
amaraezeji@gmail.com

Aparajita Roy
Department of Business
Humber College Institute of Technology and Advanced Learning
Toronto, ON, Canada
aparajita8108@gmail.com

Ayesha Karadia
Department of Business
Humber College Institute of Technology and Advanced Learning
Toronto, ON, Canada
karadiaayesha@gmail.com

Gbenga Adebowale
Department of Business
Humber College Institute of Technology and Advanced Learning
Toronto, ON, Canada
babatunde.adebowale97@gmail.com

Ivin Alexander
Department of Business
Humber College Institute of Technology and Advanced Learning
Toronto, ON, Canada
ivinjacob97@gmail.com

***Abstract*— Customer churn forecast plays a role in the efforts of telecom businesses to retain their income streams. In order to identify the elements that affect turnover, customer data from a telecom firm was analysed in this paper. Neural networks, Random Forest and Association rules were used to create models. These models have shown to have a churn prediction accuracy rate of over 80%. Customer satisfaction with the offered service, monthly fees, and internet usage are notable aspects that have been discovered. The learnings from this study can be used to create retention tactics that are unique to each client.**

Keywords*— *Customer churn forecast, telecom businesses, income streams, turnover identification, customer data analysis, neural networks, Random Forest, Association rules, churn prediction accuracy, customer satisfaction, service quality, monthly fees, internet usage, retention tactics.

I. INTRODUCTION

In recent years, the telecommunications sector has experienced rapid growth and intense rivalry, with many brands vying for consumers' interest and allegiance. Customer churn, the act of losing subscribers to competing firms or service termination, poses a huge challenge for telecommunications brands in this highly dynamic environment. For a business to maintain long-term profitability and a devoted client base, the capacity to anticipate and reduce churn is essential.

The goal of this project is to create a churn prediction system for a well-known telecommunications company. Our goal is to analyse past customer data using data analytics and machine learning to find trends and signals that point to possible churn behavior. With the help of this predictive model, the company may reach out to at-risk clients in a proactive manner and offer tailored retention methods to reduce churn rates.

The telecommunications brand can obtain a competitive edge by improving customer retention and loyalty by successfully adopting a churn prediction system. The brand will also be able to better understand client preferences, problems, and behaviour thanks to the data gained from this project, enabling targeted marketing campaigns and enhanced customer support.

In the end, the project's successful completion will show the importance of data-driven decision-making and how it will significantly influence the development of customer-centric strategies in the telecom sector. The brand can cultivate long-lasting customer relationships and ensure sustainable growth and ongoing success in a market that is constantly changing by taking a proactive approach to churn prediction and retention.

II. LITERATURE SURVEY

The literature survey section offers a comprehensive survey of the current research landscape on customer churn prediction utilizing machine learning algorithms.

A. *Study on Predictive Model of Customer Churn of Mobile Telecommunication Company (Chen et al., 2011)*

The goal of this [1] study is to create a predictive model for customer attrition in the mobile communications sector. The writers address the idea of managing client turnover and pinpoint the major variables that affect it. After that, they create a new prediction model utilising data mining methods and use a case study to show how it may be used. The study comes to the conclusion that customer turnover can be effectively managed by data mining approaches. The suggested predictive model, which employs data mining techniques, aids in reducing customer churn by examining consumer behaviour and determining the likelihood and contributing elements to a customer's decision to quit the network. The model predicts churn probability and assesses the efficacy of various models using algorithms such as decision trees, logic regression, neural networks, and others.

The results of the case study showed that the predictive model was effective in identifying customers at risk of churn and implementing appropriate retention strategies. The company implemented measures such as adjusting service combinations, introducing new services, and conducting targeted marketing activities based on customer characteristics identified through cluster analysis. These efforts resulted in a significant decrease in churn rate, from 27.5% to 13.1% within two months.

B. *Machine Learning Based Telecom-Customer Churn Prediction (Bhuse et al., 2020)*

The case study [2] focuses on implementing a number of methods, such as Random Forest, SVM, XGBoost, Ridge classifier, K-nearest neighbours (KNN), and Deep Neural Networks, in order to predict customer turnover in the telecoms sector. The dataset that was used has columns for customer churn and variables for customer attributes. The Random Forest model has the highest prediction accuracy in the testing data used before grid search, at 90.96%. Older, simpler models like Random Forest, according to the study, can be helpful in predicting consumer attrition.

The findings also demonstrated that ensemble-based classifiers, including bagging, boosting, and random forest, had lower error rates, higher sensitivity, and higher accuracy when compared to other classifiers, including Naive Bayes, Decision Tree, and SVM. The study also discovered that overfitting occurred because the models' accuracy on the training set was marginally greater than on the testing set. Overfitting was avoided using strategies including reducing tree depth or increasing the number of trees, although careful parameter optimisation was required to strike a compromise between accuracy and overfitting.

The paper also discussed the research's future directions, which include employing various hyper-parameter optimisation strategies, investigating various sets of criteria for client retention policies, and enhancing performance parameters with deep learning techniques. Overall, the case study showed how various algorithms and approaches might be used to predict customer turnover in the telecom sector, with Random Forest proving to be the most effective model in this particular application.

III. Data Description

The "Telco Customer Churn" dataset consists of 7043 items and 21 columns. The columns include client demographics, service use facts, contract information, and the goal variable 'Churn'. The dataset comprises both numerical and category characteristics. Some of the important aspects are:

- ‘SeniorCitizen’: A binary variable showing if the consumer is a senior citizen.
- ‘Partner’ and ‘Dependents’: Binary variables showing if the client has a partner or dependents.
- ‘tenure’: The number of months the customer has been with the firm.
- ‘PhoneService’, ‘PaperlessBilling’: Binary variables showing if the client has phone service and paperless billing, respectively.
- ‘MonthlyCharges’, ‘TotalCharges’: The monthly and total costs incurred by the customer, respectively.
- ‘gender’: A binary variable showing if the customer is a male or female.
- ‘InternetService’: The type of internet service a customer is using.
- ‘MultipleLines’, ‘OnlineSecurity’, ‘OnlineBackup’, ‘DeviceProtection’, ‘TechSupport’, ‘StreamingTV’, ‘StreamingMovies’: categorical variables indicating whether or not a consumer is using particular services.
- ‘Contract’: The type of contract made between the customer and the firm.

IV. Data Preparation

The data preparation stages included addressing missing values, encoding categorical variables, feature scaling and data splitting.

- Missing Values: The ‘TotalCharges’ column and ‘AvgMonthlySpending’ column in the dataset both lack values. We dealt with the missing numbers by substituting the mean for the missing values. We took that action since there were few missing values and the column was crucial for predicting future churn. Additionally, by substituting the mean for missing values, we hope to maintain the feature's average value and lessen the effect of missing data on the feature's overall distribution.
- Encoding Categorical Variables: One-hot encoding was used to encode categorical variables as binary or dummy variables. The mapping was done to binary values (0 and 1) for binary features such as ‘Partner’, ‘Dependents’, ‘PhoneService’, and ‘PaperlessBilling’.
- Feature Scaling: To guarantee that all features had comparable sizes, numerical features (such as ‘tenure’, ‘MonthlyCharges’, and ‘TotalCharges’ were standardised using the StandardScaler.
- Data Splitting and Modeling: The dataset was split into two core components: the feature set ('X') and the target variable ('y'), a fundamental step in building an effective machine learning model. To ensure a comprehensive evaluation of performance, the data was partitioned into training and validation subsets via the 'train_test_split' function. Guided by the Scikit-learn library, predictive modeling was facilitated using a Multi-Layer Perceptron (MLP) Classifier. Training on the training data was performed using the 'fit()' method, the classifier acquired predictive capabilities. This enabled predictions on both the training and validation sets. Thorough assessment included key metrics (accuracy, precision, recall, and F1-score) for both training and validation sets, offering a comprehensive view of the classifier's performance across various data.

V. Methodology

Two basic methodologies were employed for prediction: Neural Network Classification and Association Rule Mining.

1. Neural Network Classification:

In this analysis, a predictive model for customer churn was developed using a Multi-Layer Perceptron (MLP) architecture, a form of artificial neural network.

The MLP architecture was utilized for its ability to capture complex relationships within the data. The model was trained using a range of hyperparameters that significantly influence its performance. These hyperparameters included the size of hidden layers, activation functions, optimization algorithm, maximum training iterations, and batch size. By training on a combination of these hyperparameters, the neural network effectively learned patterns in the data. Predictive capabilities were subsequently evaluated on both the training and validation datasets.

The performance evaluation of the neural network revealed its predictive prowess in identifying customer churn. On the training dataset, the model demonstrated an accuracy of 79.99%, with a precision of 71.52%, a recall of 40.97%, and an F1-score of 52.10%. These metrics illustrated the model's effectiveness in correctly identifying churn cases while minimizing false positives. The validation dataset also exhibited strong predictive outcomes, with an accuracy of 80.06%, with a precision of 71.30%, a recall of 45.30%, and an F1-score of 55.40%. The utilization of the 'classification report' and 'confusion matrix' provided detailed insights into the model's performance across different evaluation dimensions. Overall, the MLP neural network demonstrated its capacity to make accurate predictions and offer valuable insights into customer churn behavior, which holds potential for informed decision-making in the business context.

2. Random Forest:

In this analysis, a Random Forest Classifier was employed as a predictive model to anticipate customer churn. Similar to the MLP architecture, the Random Forest model was selected for its ability to handle intricate relationships within the dataset. This ensemble learning technique leveraged decision tree classifiers to collectively make predictions. The model was instantiated using the 'RandomForestClassifier' class from the 'sklearn.ensemble' module and was trained on the designated training dataset.

Through the utilization of the trained Random Forest model, we made accurate predictions regarding churn outcomes for both the training and validation datasets. This enabled us to evaluate the model's performance through fundamental metrics such as confusion matrices, accuracy, precision, recall, and the F1-score. These metrics provided a comprehensive understanding of the classifier's effectiveness in predicting customer churn while minimizing false positives.

The integration of the Random Forest Classifier enriched our analysis by uncovering intricate relationships and patterns within the data. This facilitated a deeper comprehension of customer churn behavior. Our evaluation demonstrated an exceptional level of accuracy in training results (99.8%), paired with commendable precision, recall, and F1-score values. Notably, the model showcased its generalization ability to new data in the validation phase, with an accuracy of (79.7%), while precision, recall, and F1-score metrics emphasized its proficiency in predicting customer churn trends. The metrics implies great performance on training data but lesser generalization on validation data. This suggests some overfitting.

A confusion matrix is constructed to examine the number of true positives, true negatives, false positives, and false negatives.

In summary, random forest shows potential for churn prediction but needs additional tuning and regularization to reach optimal results on new data. The model outputs and metrics give a thorough mechanism to examine and modify model performance.This amalgamation of techniques affords valuable insights into the factors driving customer churn, thereby empowering businesses to make informed decisions.

3. Association Rule Mining:

In this study, we used Association Rule Mining to explore the intricate aspects of customer churn from a different perspective. By focusing on relevant categorical features, the dataset was meticulously prepared for association rule mining. One-hot encoding was employed to encode categorical variables, transforming them into a suitable format for the Apriori algorithm. The Apriori algorithm was applied to unearth intriguing relationships between these categorical features. This process involved the generation of frequent item sets, which highlighted patterns with a minimum support threshold. The strength of these associations was quantified using the lift metric, enabling the identification of substantial relationships. To ensure relevance and quality, the resulting association rules were meticulously filtered based on predefined metrics.

This methodology allowed the extraction of valuable insights into potential correlations between various features and the phenomenon of customer churn. Ultimately, the Association Rule Mining approach provided a novel perspective that contributed to a more comprehensive understanding of the intricate factors influencing customer churn behavior.

VI. Exploratory Data Analysis (EDA) and Visualization Interpretation

EDA was undertaken to acquire insights into the data and find trends connected to churn. Visualizations were utilized to better comprehend the distribution of the target variable and correlations between various characteristics and churn.

- Churn Distribution: The graphic demonstrated that there was a class imbalance, with a greater proportion of non-churned consumers compared to churned customers.

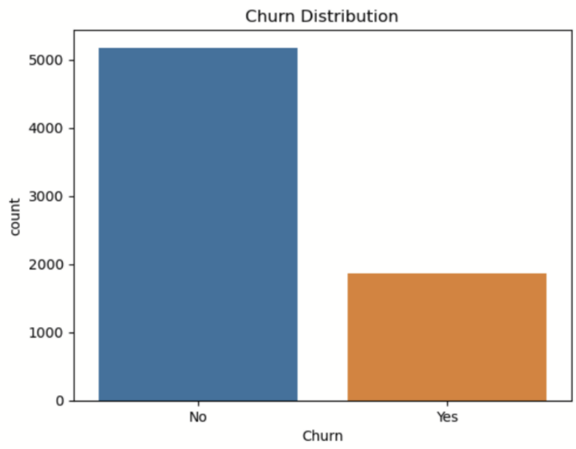

Figure I: Churn Distribution

- Churn by Contract Type: Customers with shorter contract periods are more likely to churn, as shown in the above figure. Customers with month-to-month contracts left the company at a rate that was 7-8 times higher than those with one- or two-year contracts.

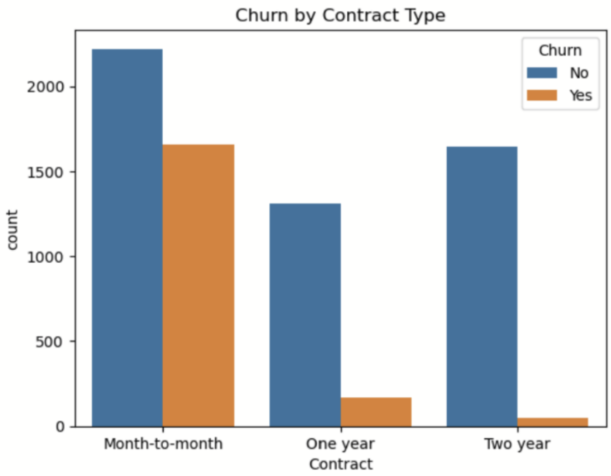

Figure II: Churn by Contract Type

- Churn by Internet Service Type: We deduced from the analysis of the aforementioned visualisation that consumers with fibre optic internet service were more likely to leave than those with DSL or no internet access. When compared to customers with DSL or those without internet connection, customers with fibre optic internet service had a churn rate that was four times higher.

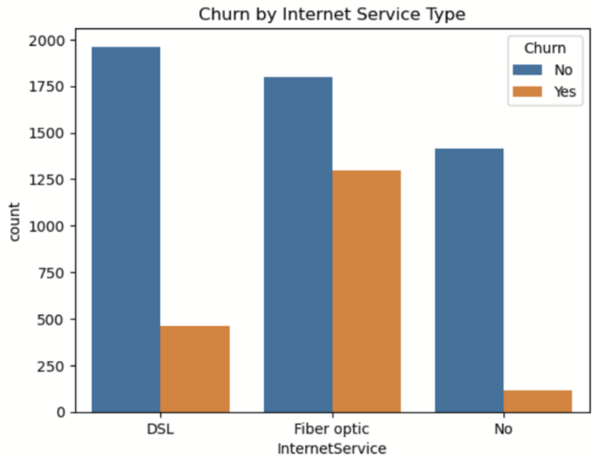

Figure III: Churn By Internet service

- Churn by Payment Method: Customers that pay by electronic cheque have a higher churn rate, as seen in the aforementioned statistic. those who pay with an electronic cheque have a higher churn rate, which is five times higher than those who pay in another way.

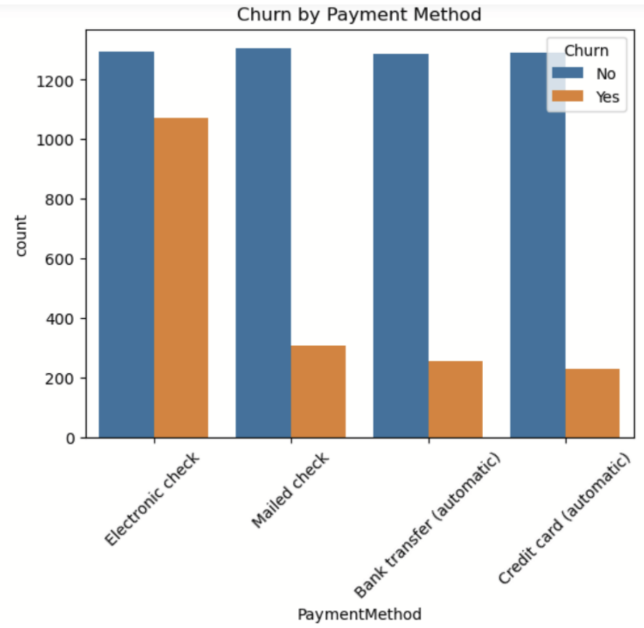

Figure IV: Churn by Payment Method

- Correlation Heatmap: Below is the correlation heatmap depicting the correlation between 3 variables 'SeniorCitizen', 'tenure' and 'MonthlyCharges'. MonthlyCharges and tenure variables have low correlation with each other implying that higher tenure is associated with lower monthly charges. MonthlyCharges and SeniorCitizen variables have low correlation with each other implying that Senior Citizens have to pay lower monthly charges.

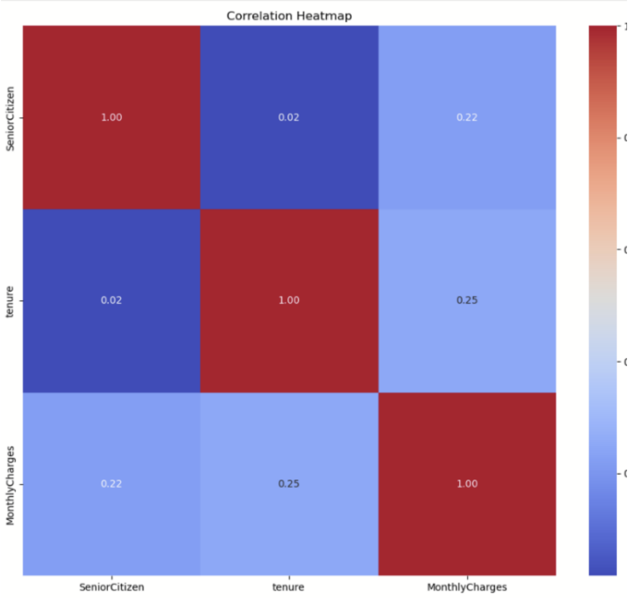

Figure V: Correlation Heatmap

VII. Model Optimization and Recommendations

To increase model performance, we may study different machine learning methods like Decision Trees, Random Forest, and XGBoost. Additionally, feature engineering and hyperparameter tweaking might be further researched to boost model accuracy and interpretability.

Based on the insights acquired from the investigation, the telecom operator may execute targeted retention measures to lower churn rates. For example, giving incentives to consumers on longer contract terms or promoting other payment options to limit electronic check usage.

VIII. Conclusion

In this research, we successfully created a telecom churn prediction model employing a Neural Network classifier, Random forest classifier and Association rule mining. The model demonstrated promising results in forecasting customer attrition, and the association rules gave useful insights into consumer behavior. By employing this predictive model, the telecom business may take proactive efforts to retain consumers and enhance overall customer happiness. Hyperparameter adjustment using grid search cross-validation can assist decrease overfitting and enhance validation performance. Feature selection and ensemble approaches like bagging can also aid.

The feature significance from the model may be retrieved to discover the strongest predictive variables for churn. These business insights can flow into client retention campaigns.

IX. Future Work

Future work may comprise researching additional machine learning models, undertaking feature significance analysis, and implementing the final model in a production setting. Regular monitoring and retraining of the model with fresh data will be crucial to maintain its continuous accuracy and efficacy in minimizing churn.

References

[1] Chen, Y., Li, B., & Ge, X. (2011). Study on Predictive Model of Customer Churn of Mobile Telecommunication Company. 2011 Fourth International Conference on Business Intelligence and Financial Engineering, 114–117. https://doi.org/10.1109/bife.2011.112

[2] Bhuse, P., Gandhi, A., Meswani, P., Muni, R., & Katre, N. (2020). Machine Learning Based Telecom-Customer Churn Prediction. 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), 1297–1301. https://doi.org/10.1109/iciss49785.2020.9315951

[3] Telecom Customer churn Prediction. (2022, July 6). Kaggle. https://www.kaggle.com/datasets/shilongzhuang/telecom-customer-churn-by-maven-analytics