Customer Churn Prediction Based on Interpretable Machine Learning Algorithms in Telecom Industry

Liwen Ou
College of Arts and Science, New York University
New York, United States of America
lo2054@nyu.edu

***Abstract*—Research regarding the interpretability and feature importance of customer churn prediction is rare compared to study on its performance. Nevertheless, knowing what type of customer information are the most significant for predicting churn is critical for firms to layout cost-efficient investment strategies. In this work, feature importance for churn prediction in telecom industry is investigated using a dataset with nearly 40 customer features, and 3 interpretable, tree-based machine learning models. In particular, the data is trained in random forest, decision tree and extra tree classifier algorithms, and uses their built-in feature importance functions to interpret the significance of each feature. The feature importance scores from all three models are compared to analyze which features are the most significant overall. As the result, all models yield optimal prediction accuracy, with decision tree model's slightly lower than the others. The resulting top features from all three models have large overlaps, indicating all models are suitable for the purpose of this study. Random forest and extra tree classifier, as the more complicated models, produce the most satisfactory results, which demonstrate that features related to customer's tenure, consumption in the firm and referral amount are the most important to make accurate churn prediction. This analysis result is reasonable since all these features imply customers' personal preference towards the company.**

Keywords—Customer churn prediction, feature importance, random forest, decision tree

I. Introduction

A common issue that practically every business faces is customer churn. When a customer churns, it means this individual stop using certain product or service provided by a company. Since the cost to retain existing customers is usually lower than acquiring new ones, it is critical for firms to predict customers canceling their services before it happens, also known as churn prediction. Churn prediction utilizes machine leaning algorithms that learns from customer features, including but not limited to age, location, education, browsing behaviors, to classify which groups of customers are most likely to churn [1-3]. This allows them to act proactively to keep their customers from leaving by designing investment strategies accordingly. These strategies can be tailored to target specific groups, or individuals, to maximize the cost efficiency. As this is a universal problem for firms, churn prediction has very wide application and more and more firms are adopting this method, especially firms that provide subscription services.

The history of machine learning be traced back to the 1940s. In 1943, Walter Pitts and Warren McCulloch created the world's first neural network mathematic model that mimic how human think. In 1952, Arthur Samuel published the Checkers program that can play checkers and learn from playing, which is one of the earliest machine learning programs and initiated the researches in this field. In 1956, many well-known mathematician and researchers were invited to Dartmouth College, where they brainstormed on self-learning and thinking machine for weeks, and this event is later considered to be the birthplace of artificial intelligence. The development in machine learning and artificial intelligence were slowed down since 1973 as British government cut funding for research, known as "AI winter". But with the rapid development of computer system and game technology in the 1990s, machine learning developed very fast. For instance, Google's AlphaGo program can beats the world's top Go players. Today, machine learning is wildly utilized in many fields, including customer churn prediction in business.

Early work related to this topic mainly focused on testing and confirming that machine learning is a highly efficient way to predict customer churn [4]. And after that, research focused on comparing different machine learning algorism to yield optimal prediction accuracy. Commonly used models include but not limited to Naive Bayes, Logistic Regressions, Linear Regression, Decision Trees, Support Vector Machines. For instance, Bingquan Huang in his 2011 research found that prediction using Naïve Bayes modelling yields much better result when high dimension dataset is transformed to lower dimension, compared to unreduced dataset [5]. Research has also been done on comparing different customer feature sets [5]. Noticeably, there is a limitation on the type of industry that were studied about their customer churn trend. A handful amount of previous research studied on the telecommunication industry, and it is one of the early developed subscription-service-providing industry, which most directly benefit from lower customer churn rate and higher retention. In recent years, research has been done on other industries, such as streaming, banking and insurance industry. However, although plenty of research have been done on producing accurate prediction, fewer research studied on the interpretability and feature importance of churn prediction result. Being able to interpret prediction is important to firms, as it gives out what type of customer data need to be collected for making prediction.

To solve the limitation mentioned above, this study will conduct on a relatively high dimension dataset from a specific industry, and uses interpretable machine learning models to investigate what type of customer features are significant for making prediction customer churn.

II. Methods

A. Dataset and Preprocessing

A customer information dataset of a Telecom company

979-8-3503-4372-4/23/$31.00 ©2023 IEEE
DOI 10.1109/CSMIS60634.2023.00120

from Kaggle was employed to conduct this research [6]. The original dataset contains 7, 043 rows, which represent all of the customers this company has in from Q2 2022. Each column of the dataset represents one of the 38 features of customers, including the customers' location, gender, age, educational background, tenure etc.

While some of the data are unimportant or missing, pre-processing is necessary for later analysis. Firstly, the features that is intuitively unrelated to customer churn were dropped. The location features including Longitude, Latitude and Zip Code since locational information is loosely correlated with churning were not considered. Then, this study also dropped Customer ID since this is just random string of digits given to every customer. Churn Category and Churn Reason is also dropped since the data is very sparse, with only 1869 existing data. Also, this type of feature which is purely words and sentences is hard to be translate into meaningful numerical data. There might be some other features that is knowingly unimportant for our prediction purpose.

In addition, after a brief view of the dataset's statistical summary, it can be found that there are a significant amount of missing data. Both Avg Monthly Long Distance Charges and Avg Monthly GB Download are missing about 1000 customer data. The averages were used to fill those, as deleting all these customers will loss a large proportion of data. Also, about half of the features are categorical data, such as features Married, Offer and InternetType. Thus the LabelEncoder from Sklearn was utilized to change categorical data to numerical. For example, for Married feature, replace 1 for Yes, 2 for No. The NA was also replaced with 3. The resulting dataset is in the shape of 7043 x 32. This means 31 columns of features and 1 column as label, which is the churn status. Finally, the train_test_split function from Sklearn was also employed to randomly split dataset into 20% testing set and 80% training set.

B. Proposed Approach – tree-based Models

For model selection, since this study emphasized on the interpretability and feature importance analysis of churn prediction not just for the performance of models e.g. accuracy, tree-based models were considered in this research as they are the most easy and intuitive to interpret how significant each feature is for predicting a target variable. In particular, the random forest, decision tree and extra trees classifier were chosen as the models for making prediction. All three of them were imported from Sklearn, all of which has built-in feature importance function and thus will be easy for the research purpose.

Random forest and decision tree are both supervised machine learning algorism for classification and regression purposes [7-10]. The way a decision tree works is iteratively split the data based on certain features and reach a specific result. A random forest algorism, as the name suggests, is a collection of multiple randomly created decision trees and combine the outputs of each tree to generate a final result. Extra tree is another algorism that put together multiple decision trees, similar to random forest, but with different cut-point selection and faster run-time.

For random forest, the n_estimators parameter represents the number of trees that will be used in the model. 1000 trees is usually considered a standard number. After some experimenting, it can be found that the resulting prediction accuracy from a small number (less than 500) is not significantly lower than the result from a large number (3000). Thus, 1000 was considered in the model as it gives us a decent accuracy while keeping the runtime relatively short.

After fitting the 3 models with training dataset, the feature importance function was utilized as mentioned earlier, as well as matplotlib to draw a bar plot for the significance of each feature in each model. Then, for each model, this study used the testing dataset to make prediction with the trained model, and used it with the actual data to create a Confusion Matrix to assess the churn prediction of our models.

III. RESULTS AND DISCUSSION

A. Performance and Feature Importance

The result of all three models showed fairly good accuracy for predicting customer churn. The top features of all three model have high overlap, which confirms the significance of those features on making predictions. For random forest and extra tree classifier, the significance of the features is more even and spread. The top 8 feature all showed decent significance with feature importance score of more than 0.05 shown in Fig.1, Fig. 2 and Fig. 3 . However, for decision tree, the most significant feature accounts for almost 0.35 of the result and the importance of following features dropped rapidly, as shown in Table I.

B. Discussion

The results are reasonable, since both random forest and decision tree classifier are more complicated machine learning algorithms compare to decision tree, and thus is capable of producing more accurate prediction. Tenure in month is on average the most significant feature, as it is always the most important features of any model. This make sense since intuitively, a large number of customer churn is based on their previous experience with the firm. Whether a customer has been purchasing product and service in the firm for too long or too little of a time indicates strong tendency of churning. Similar reasoning can be used with the feature Contract, which is another top feature on average. In addition, Total Charges, Total Revenue, Monthly Charge and Total Long Distance Charges are all related to customers' consumption level in the firm, which are all among the top features in three models. This is reasonable since a customer might want to leave if the charges are too high. And also, if a customer is spending too little, this might indicate that he/she does not like the service that the firm provide, which can be a plausible cause for churning. Essentially, whether customers choose to stay or leave really depend on if the value of the services they receive matches the amount they are charged. Thus, related features are likely significant indicators for churn prediction. Finally, Number of Referral is also an somewhat important feature as indicated by all three model. This make sense since the number of times a customer has referred a friend to the company is strongly positively related to this customer's personal preference towards the firm. Knowing whether a customer like or dislike the firm imply a lot on his/her possibility to stay or leave.

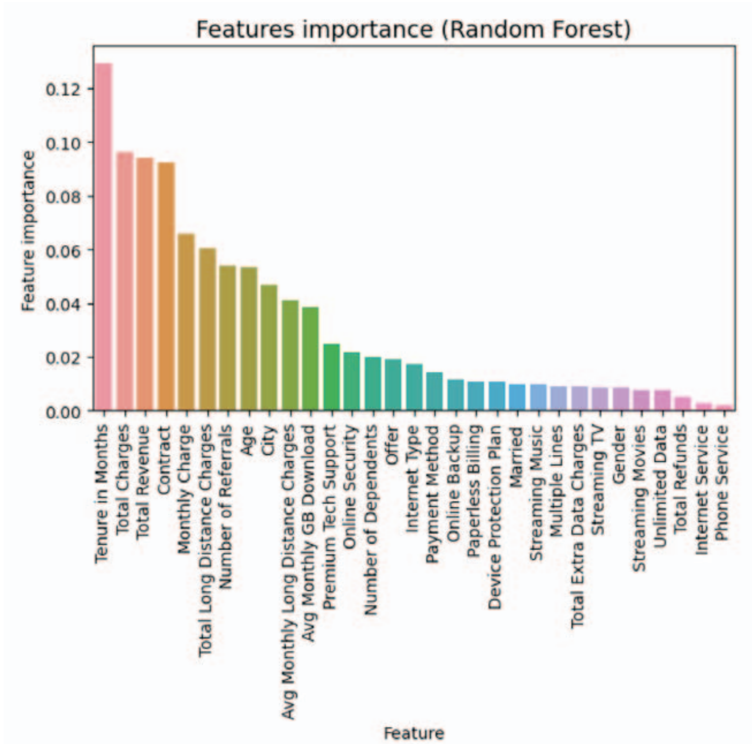

Fig. 1. Feature importance for random forest model

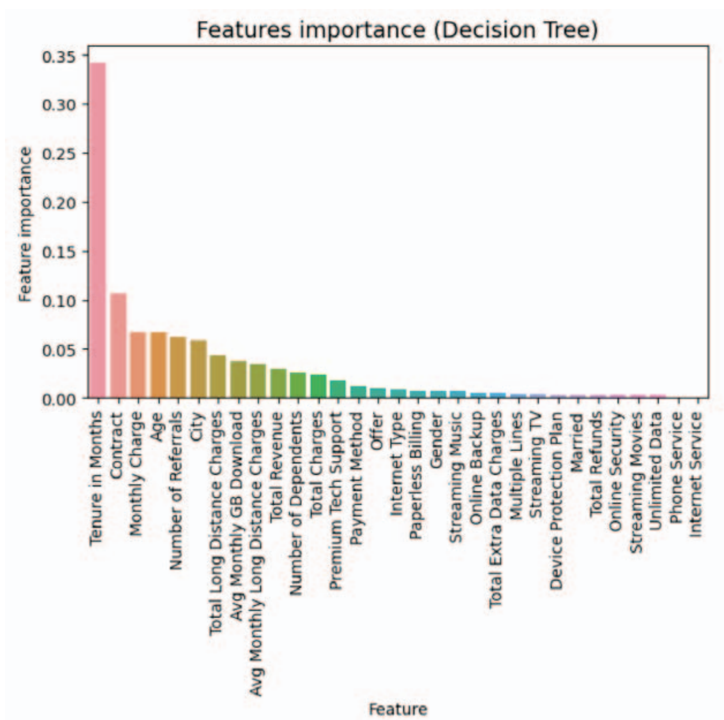

Fig. 2. Feature importance for decision tree model

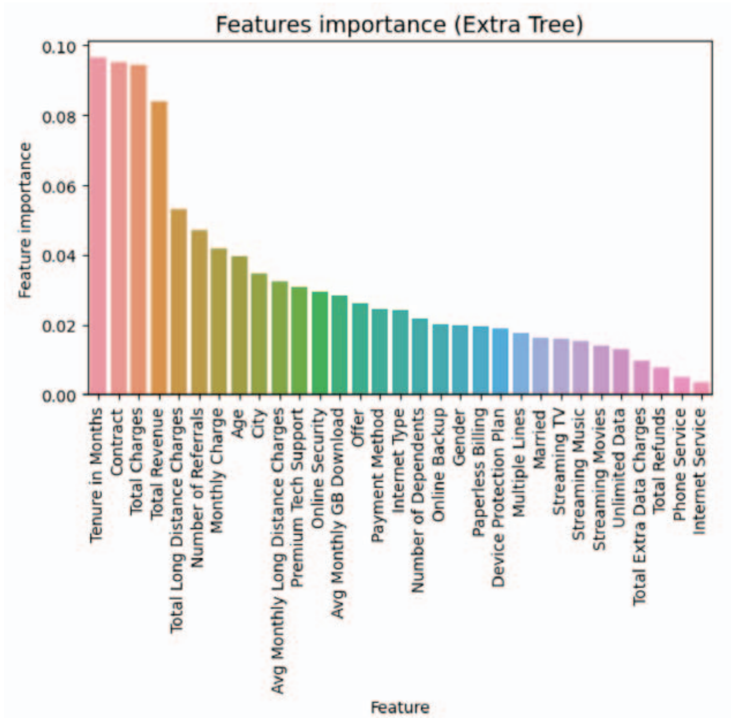

Fig. 3. Feature importance for extra tree model

TABLE I. THE PERFORMANCE OF MODELS

Performance	Classification Models		
	Random Forest	Decision Tree	Extra Tree
Accuracy	0.82328	0.77502	0.82541

IV. CONCLUSION

This study investigated on what type of features are the most significant for customer churn prediction in the telecon industry. The interpretable tree-based machine learning models, including random forest, decision tree and extra tree classifier were utilized to calculate the feature importance of almost 40 customer features. The results demonstrated both random forest and extra tree model can achieve satisfactory performance. In addition, the information related to the feature importance showed that features that indicate customer's tenure, consumption level and referral amount are most significant for making prediction. All of those can show customer's preference toward the company. In the future, research on different industry is necessary for purposing a universal way to determine feature importance for churn prediction on all industries.

REFERENCES

[1] T. Chih-Fong, and Y. H. Lu. "Customer churn prediction by hybrid neural networks." Expert Systems with Applications 36.10, 2009, 12547-12553.

[2] Y. Qiu, et al. "Clustering Analysis for Silent Telecom Customers Based on K-means++." 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). Vol. 1. IEEE, 2020.

[3] W. Verbeke, et al. "Building comprehensible customer churn prediction models with advanced rule induction techniques." Expert systems with applications 38.3, 2011, 2354-2364.

[4] V. Umayaparvathi, and K. Iyakutti. "A survey on customer churn prediction in telecom industry: Datasets, methods and metrics." International Research Journal of Engineering and Technology (IRJET) 3.04, 2016.

[5] H. Bingquan, M. T. Kechadi, and B. Buckley. "Customer churn prediction in telecommunications." Expert Systems with Applications 39.1, 2012, 1414-1425.

[6] Kaggle. Telecom customer churn by maven analytics, 2022. https://www.kaggle.com/datasets/shilongzhuang/telecom-customer-churn-by-maven-analytics

[7] M. Anthony J., et al. "An introduction to decision tree modeling." Journal of Chemometrics: A Journal of the Chemometrics Society 18.6, 2004, 275-285.

[8] Q. J. Ross. "Learning decision tree classifiers." ACM Computing Surveys (CSUR) 28.1, 1996, 71-72.

[9] B. Gérard, and E. Scornet. "A random forest guided tour." Test 25, 2016, 197-227.

[10] R. Steven J. "Random forest." Journal of Insurance Medicine 47.1, 2017, 31-39.