Customer Churn Prediction Using Data Mining Techniques for an Iranian Payment Application

Olya Rezaeian
Department of Industrial Engineering and Management Systems
Amirkabir University of Technology
Tehran, Iran
elia.rezaeian95@aut.ac.ir

SeyedHamidreza Shahabi Haghighi
Department of Industrial Engineering and Management Systems
Amirkabir University of Technology
Tehran, Iran
shahabi@aut.ac.ir

Jamal Shahrabi
Department of Industrial Engineering and Management Systems
Amirkabir University of Technology
Tehran, Iran
jamalshahrabi@aut.ac.ir

2021 12th International Conference on Information and Knowledge Technology (IKT) | 978-1-6654-0849-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/IKT54664.2021.9685502

***Abstract*—Customer Relationship Management (CRM) and data-driven marketing have become of paramount importance in this age of evolved markets and fierce competition among businesses. One of the most important branches of CRM is retaining existing customers. Since customer acquisition is about 5 to 6 times more costly than retaining customers, achieving an accurate model for customer churn prediction is essential to devise marketing retention strategies. Therefore, in this study, ensemble models are proposed to predict customer churn. Since customer churn is a rare occurrence in an organization and causes an imbalanced distribution in the target variable, ensemble learning algorithms, one of the most efficient and widely used methods, have been used to deal with this problem. With regard to the case study, the dataset was generated on demographic and 13-month transactions of users of an Iranian payment application. In this study, the best model to predict customer churn is the bagging version of Decision Tree, reaching the highest accuracy, f-measure and AUC.**

Keywords— Customer Churn, Data Mining, Imbalance Data, RFM Model

I. INTRODUCTION

Currently, with regard to fierce competition between businesses, reduced product lifespan, and reduced customer loyalty to the brand, there are concerns about losing customers in organizations. This has led them to turn to Customer Relationship Management (CRM) and especially data-driven marketing. One of the most important branches of CRM is customer churn prediction (CCP) because it is more beneficial to retain existing customers and make them satisfied for the following reasons [1]:

- Successful companies have a long-term relationship with their customers, which helps them better understand customers' needs.
- Customers who have left the organization have a significant impact on others through their social networks.
- Long-term customers have a significant impact on both profits and costs. From a profit point of view, they tend to buy or use more products or services and convey their positive opinions through word-of-mouth marketing. In terms of cost, it is easier to provide services and products to these customers because their needs are known, and a lot of information is available about them, the service cost is reduced.
- Competitors' marketing practices have less impact on the long-term customers of organizations.
- Losing customers increases the need of customer acquisition and reduces profits by reducing sales and losing the opportunity of up-selling and cross-selling.

When a customer stops his/her relationship with a company, and it is possible that the customer also joins the competing company is called customer churn. It occurs in three ways [2]: 1) Voluntary, it can be due to new technology, economic,... and this category is difficult to identify. 2) Non-voluntary, it happens for some reasons such as committing a crime, non-payment of bills, etc. 3) Silent, for no reason, the customers leave the organization, in which case the organization can not do anything prevent. Based on this classification, it can be said that researchers place more emphasis on identifying the first category.

The main step in Customer Churn Management is to provide a model that can help organizations predict customer churn. Data mining techniques are well-known tools for this purpose. These techniques use various methods to extract patterns in large datasets [3]. The basic idea of data mining for CRM is that old data contain information that will be useful in the future because the customer behavior shown in these data is not random but informative, so customer data collection has become a must [4].

On the other hand, customer churn is a rarity, meaning that the number of people who leave an organization is much lower than those loyal and do not leave. Customer churn prediction is a binary classification with an imbalanced target variable. This class imbalance negatively affects classification algorithms [5], contributing to bias towards the majority class, high accuracy but low precision.

In this study, considering the gap in previous research of customer churn in financial services, especially payment applications that recently have risen widely to the surface, a

978-1-6654-0849-3/21/$31.00 ©2021 IEEE

fledgling Iranian payment application has been investigated. Due to the novelty of these applications, few loyal users are active in them, so it is necessary to prevent them from the churn and try to turn them into loyal users. There are many reasons for users to leave these organizations. For example, failed transactions, bank account changes, aggressive marketing strategies of competitors, etc. Predicting customer churn help this app achieve its goals. One of the most important challenges for these organizations can be solved, and less money can be paid for attracting the same number of customers by anticipating people prone to leave and using appropriate strategies to prevent them from leaving.

For the purpose of reaching the above-mentioned objectives, some informative variables are required to be used in CCP models[6]. Here, an appropriate dataset can be created by adding variables such as Failure Rate, Tenure, etc, to RFM model parameters extracted from the transactional data. Therefore, in this research, in addition to basic machine learning methods and the ensemble version of them, some advanced ensemble learning methods will be applied, and their ability to predict churners will be compared.

The remaining paper is structured as follows. Section II provides related work. Section III gives information about data and presents the proposed customer churn prediction models. Section IV describes the results derived from models, and section V conclude the discussion and presents future work.

II. RELATED WORK

CCP's goal is to identify customers who want to leave the organization, and this provides the proper insight to apply customer retention strategies. CCP has been a major issue in many areas such as telecommunication[7-9], e-commerce retail [7], search ads [8], banking[4], insurance, and financial services, which little research has been done in this field.

It is stated in the literature that retaining existing customers is more beneficial than acquiring new ones. Regarding customer acquisition, in addition to accepting its high cost (5 to 6 times more than customer retention[5, 9]), it needs to spend time to gain trust and inspire loyalty [10]. In some industries, a 5% reduction in customer churn has doubled their profitability [11]. In order to keep these customers in the organization, several researchers have tried to increase the accuracy of their data mining model to target leaving customers correctly in order to devise proper marketing strategies.

Researchers used many data mining techniques such as Decision Tree[1, 7, 8, 12], Support Vector Machine[8, 10, 12], Logistic Regression [1, 12], k-Nearest Neighbor [13], Artificial Neural Network [7, 9, 12], and Naive Bayes[14] beside some advanced techniques like Random Forest[12, 15], AdaBoost [12, 14], XGBoost [12], Gradient Boosting Decision Tree [8] to reach this goal. Furthermore, a lot of research has been conducted to handle the class imbalance problem. To be more in-depth, four approaches have been considered to deal with this issue:

1) Data level methods[5, 16]: Modify existing data by deleting and adding observations to balance the data distribution. For example, Random Over Sampling[9], Random Under Sampling [9, 17], SMOTE[9], DSMOTE[2], ...

2) Algorithm level methods[5, 16]: Methods that directly modify existing algorithms to adapt them to learn the minority class, such as Cost-Sensitive Approaches [17], One-Class Learning, and Active Learning[9].

3) *Hybrid method[16]s:* Methods take advantage of the above two methods, leading to a stable and effective model.

4) *Ensemble methods[5]:* These methods try to improve the performance of classifiers by combining them. For instance, Boosting, Bagging, Random Forest, RUSBoost [18],etc.

Thus the motivation of this study is to present an accurate model to predict customer churn as well as handle class imbalance, which can help the company identify the customers prone to leave, leading to not only cost reduction but also saving time and energy.

III. METHODOLOGY

A. Data Description

The data used in this study, including 1,254,194 transactions and the demographic data of 5552 users of an Iranian payment application in a period of 13 months, from March 2019 to March 2020, have been provided to the researchers as two separate Excel files. In this study, churner refers to the users who do not have any transactions in the next 12 months. Regarding Fig. 1, showing the distribution of the target variable, 38.3% of the users have left this application. This means that the data set is imbalanced.

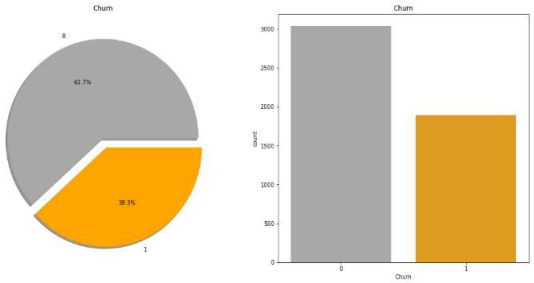

Fig. 1. Distribution of the target variable

B. Data Preparation

In order to prepare a suitable data set for data mining, the following steps need to be done:

1) Data Cleaning: Many unrelated and conflicting transactions have been deleted. Besides, the gender variable has been removed from the data set because of missing values of about 50%.

2) Derived Variables: At this stage, several new informative variables are created based on the transactional data using methods such as the RFM model explaining customers' behavior well. TABLE I shows all variables with their explanations, and the new ones are marked with a star.

3) Data Transformation: Data mining requires all the variables to be numerical for prediction models to be

implemented. Here, the categorical variables are converted to numerical ones by assigning unique numbers to each category.

4) *K-Fold Cross Validation:* It is one of the most popular methods in data mining model validation. First, it divides the dataset D into K equal and different parts: D_1, D2, ..., D_k (The number k is considered to be 10 here). The k model is trained and tested so that each time k-1 part is used as training data and 1 part as test data. In this method, the final accuracy of the classification was equal to the mean of k replication. Obviously, the larger k, the more reliable the result.

TABLE I. FINAL VARIABLES

No.	variables	*Explanation*
1	*User_ID*	-
2	*Phone_Operator*	Mobile operator
3	*Gender*	-
4	*Subscription_Lenght*	The time since the day of registration in the application
5	*Tenure**	The time since the first transaction was made
6	*Num_Account**	Number of bank accounts
7	*Num_Transaction_Type**	Number of types of transactions
8	*RFM_Level**	User-level based on RFM model
9	*AVG_Days_F**	Average days between failed transactions
10	*AVG_Days_S**	Average days between successful transactions
11	*Recency_S**	The number of days since the last successful transaction
12	*Frequency_S**	Total number of successful transactions
13	*MonetaryValue_S**	Amount paid for all successful transactions
14	*Failure_rate**	-
15	*Card_Transfer**	Card to card ratio
16	*BillPay**	The ratio of bill payment transactions
17	*GetInfo**	The ratio of information capture transactions
18	*PhoneCharge**	The ratio of transactions of the latest internet package or mobile phone charge
19	*Other**	The ratio of other transactions
20	*Churn*	Churn or non-churn (target variable)

Before Training and testing the models, the dataset was divided into train and test data with a ratio of 70 to 30.

C. Data Mining Models

In the modeling stage, a suitable supervised machine learning algorithm is needed to predict the churners accurately. In this study, state-of-the-art ensemble learning algorithms have been used to achieve an accurate model that can handle the class imbalance problem. Ensemble learning is a machine learning method seeking better performance by combining the predictions of multiple weak classifiers. The reason for this choice is the results of previous research[5].

In this study, the ensemble models are created based on Decision Tree, Logistic Regression, Naive Bayes, Support Vector Machine, and K-Nearest Neighbors using Bagging, an ensemble learning method. Besides, some widely-Known ensemble algorithms like Random Forest, XGBoost, and LightGBM are used to compare their results with the former ones. Furthermore, to achieve accurate models and prevent overfitting, their hyperparameters have been tuned using Cross-Validation. Moreover, the performance of these models has been compared with the base models before and after tuning, and the most accurate one was selected to predict the customer churn.

D. Evaluation Metrics

There are various metrics to evaluate the effectiveness of different algorithms. The following are some examples used in this study[19]:

- *Confusion Matrix*: To better illustrate the performance of the models, the Confusion Matrix, shown in TABLE II, is used.

TABLE II. CONFUSION MATRIX

Confusion Matrix		Actual	
		YES	***NO***
Predicted	***YES***	TP (True Positive)	FP (False Positive)
	NO	FN (False Negative)	TN (True Negative)

Each element of the matrix is as follows:

- *TP (True Positive):* The number of churners that the algorithm has correctly predicted.
- *FN (False Negative):* The number of churners that the algorithm has incorrectly predicted in the non-churner category.
- *TN (True Negative):* The number of non-churners that the algorithm correctly predicted.
- *FP (False Positive):* The number of non-churners that the algorithm incorrectly predicted in the churner category.

- *Accuracy:* This measure shows the number of correct predictions and calculated using Eq. (1). This criterion is very sensitive to data distribution and does not perform well in class imbalance problems.

$$Accuracy = \frac{TP+TN}{P+N} \quad (1)$$

- *Precision:* The number of positive observations that were found to be positive. This criterion is calculated using Eq. (2).

$$Presicion = \frac{TP}{TP+FP} \quad (2)$$

- *Recall:* The proportion of positive observations that were correctly predicted as positive cases. This criterion is calculated using Eq. (3).

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

- *F-Measure:* It is the harmonic average of precision and recall. This criterion is calculated using Eq. (4).

$$F1 - Score = \frac{2.Recall\ .Precision}{Recall+Precision} \quad (4)$$

- *Receiver Operating Characteristics Curve(ROC curve):* This graph shows the ability to evaluate a binary classification system with a changeable threshold. The ROC curve is created by plotting the true positive rate (TPR), to the false positive rate (FPR).

$$TPR = \frac{TP}{P} \quad (5)$$

$$FPR = \frac{FP}{N} \quad (6)$$

$$AUC = \int_0^1 TPR(t)\ dFPR(t) \quad (7)$$

IV. Experimental Results

In this section, the predictive performance of the CCP models Should be evaluated by several metrics, which was mentioned in the previous section.

First and foremost, all models were implemented in their default version on training data, and the results are presented in TABLE III. Concerning their performance, XGBoost outperforms every single algorithm in all metrics except precision, where the bagging version of SVM has a better performance.

TABLE III. The Results of Applying CCP Models Before Tuning

Method	Version	Accuracy	Precision	Recall	F-measure	AUC
Decision Tree	Simple	77.320	70.684	69.817	70.207	75.932
	Bagging	81.351	77.234	72.997	74.963	89.405
Naive Bayes	Simple	78.684	69.786	78.288	73.746	85.628
	Bagging	78.713	69.993	77.910	73.688	85.967
Logistic Regression	Simple	81.757	81.407	68.003	73.936	90.674
	Bagging	81.931	81.818	68.079	74.130	90.621
KNN	Simple	79.378	75.061	69.289	71.936	85.534
	Bagging	79.379	75.518	68.532	71.748	86.487
SVM	Simple	82.859	82.264	70.572	75.800	90.997
	Bagging	82.888	**82.537**	70.269	75.735	90.969
Random Forest	-	83.150	78.300	77.686	77.894	91.044
XGBoost	-	**83.411**	77.589	**79.959**	**78.659**	**91.519**
LightGBM	-	81.729	75.960	76.629	76.191	90.514

After tuning models' hyperparameters by cross-validation, they are trained by 10-fold cross-validation on 70% of the dataset, and the results are shown in TABLE IV. Obviously, all tuned models performed better than their default versions, and the bagging version of Decision Tree beats other algorithms in both f-measure and AUC (79.40% and 91.77%), two important metrics in class imbalance problems. So, the Bagging of the Decision Tree was chosen as the best option for prediction.

TABLE IV. The Results of Applying CCP Models After Tuning

Method	Version	Accuracy	Precision	Recall	F-measure	AUC
Decision Tree	Simple	82.686	73.928	**84.943**	78.961	83.114
	Bagging	83.817	77.627	81.471	**79.400**	**91.765**
Naive Bayes	Simple	80.481	75.157	73.371	74.171	87.051
	Bagging	80.539	74.963	73.976	74.387	87.126
Logistic Regression	Simple	82.076	81.381	69.137	74.587	90.621
	Bagging	82.105	**81.518**	69.061	74.612	90.645
KNN	Simple	81.177	78.530	70.117	73.971	88.000
	Bagging	81.264	78.756	70.118	74.034	88.088
SVM	Simple	82.888	79.432	74.810	76.926	90.620
	Bagging	83.410	79.160	77.228	78.052	90.630
Random Forest	-	83.236	78.311	78.066	78.083	91.389
XGBoost	-	83.556	78.017	79.651	78.757	86.061
LightGBM	-	**83.846**	78.849	79.351	78.976	91.445

On the other hand, regarding TABLE III and IV, ensemble models perform significantly better than the simple ones, it means that ensemble methods can better handle the class imbalance problem since they can improve the results of the simple models especially in f-measure.

In the prediction phase, the remaining 30% of the dataset was predicted by some high-performance models, shown in TABLE V. This table demonstrates the supremacy of the tuned version of bagging with Decision Tree as the base classifier. It has the best performance in three main metrics, accuracy, f-measure and AUC (84.64%, 80.45%, and 90.50%).

Method	Version	Accuracy	Precision	Recall	F-measure	AUC
Decision Tree	Simple	83.288	74.024	**86.949**	79.968	83.979
	Bagging	**84.641**	78.620	82.363	**80.448**	**91.504**
Random Forest		84.100	**79.964**	78.131	79.037	91.258
XGBoost		84.438	79.203	80.600	79.895	86.836
LightGBM		83.627	79.492	77.249	78.354	91.273

TABLE V. The Prediction Results

Fig 2. represents the confusion matrix of the proposed model and Fig 3. Compares the ROC curves of the proposed and other models. As it is shown in Fig 3. Decision Tree and XGboost perform weak concerning their ROC curve, and the ROC curve of the bagging version of Decision Tree, Random Forest and LightGBM are almost similar.

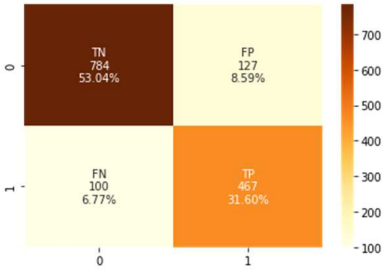

Fig. 2. Confusion Matrix of the Proposed Model.

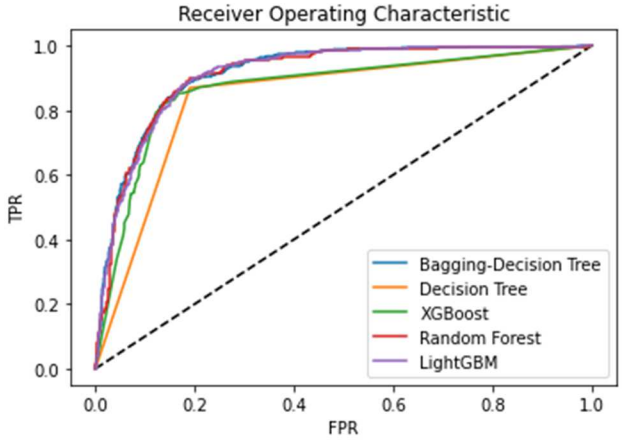

Fig. 3. ROC Curve of the Proposed and Other Used Models.

V. Conclusion

As mentioned in the previous section, due to stiff competition between businesses, studying and analyzing customer behavior is necessary to achieve success and profitability in any company. Sometimes it is necessary to predict these behaviors, resulting in taking appropriate action. One of the most critical issues in this area is predicting customer churn. Customer churn prediction is done to identify customers who are prone to leave the organization, and this provides the proper insight to apply customer retention strategies.

The users' features have been obtained from the demographic and transactional data using different methods like RFM model, and the number of these features has reached 20.

Prior to classification, various methods are used to preprocess, clean, and standardize the data. The class imbalance problem has been observed by data visualization. In order to handle this problem, Ensemble learning algorithms have been used, and their performance was compared with simple models. Due to the results mentioned above the bagging version of Decision Tree overperformed. The accuracy, f-measure and AUC reached about 84.64%, 80.45% and 91.50%, respectively, which were the highest.

Overall, the ensemble models outperfomed the simple version of the algorithms and can handle the class imbalance much better.

References

1. De Caigny, A., K. Coussement, and K.W. De Bock, *A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees.* European Journal of Operational Research, 2018. **269**(2): p. 760-772.
2. Al_Janabi, S. and F. Razaq. *Intelligent Big Data Analysis to Design Smart Predictor for Customer Churn in Telecommunication Industry*. in *International Conference on Big Data and Smart Digital Environment*. 2018. Springer.
3. Han, J., M. Kamber, and J. Pei, *Data mining concepts and techniques third edition.* The Morgan Kaufmann Series in Data Management Systems, 2011. **5**(4): p. 83-124.
4. De Bock, K.W. and D.J.E.S.W.A. Van den Poel, *Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models.* 2012. **39**(8): p. 6816-6826.
5. Zhu, B., B. Baesens, and S.K. vanden Broucke, *An empirical comparison of techniques for the class imbalance problem in churn prediction.* Information sciences, 2017. **408**: p. 84-99.
6. Khodabandehlou, S. and M.Z. Rahman, *Comparison of supervised machine learning techniques for customer churn prediction based on analysis of customer behavior.* Journal of Systems and Information Technology, 2017.
7. Gordini, N. and V. Veglio, *Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry.* Industrial Marketing Management, 2017. **62**: p. 100-107.
8. Wang, Q.-F., M. Xu, and A. Hussain, *Large-scale ensemble model for customer churn prediction in search ads.* Cognitive Computation, 2019. **11**(2): p. 262-270.
9. Zhu, B., Y. Pan, and Z. Gao. *Application of Active Learning for Churn Prediction with Class Imbalance*. in *Proceedings of the 2018 International Conference on Machine Learning Technologies*. 2018.
10. Idris, A., A. Iftikhar, and Z.J.C.C. ur Rehman, *Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling.* 2017: p. 1-15.
11. Liu, M., et al., *Three categories customer churn prediction based on the adjusted real adaboost.* 2011. **40**(10): p. 1548-1562.
12. Lalwani, P., et al., *Customer churn prediction system: a machine learning approach.* Computing, 2021: p. 1-24.
13. Keramati, A., et al., *Improved churn prediction in telecommunication industry using data mining techniques.* Applied Soft Computing, 2014. **24**: p. 994-1012.
14. Vafeiadis, T., et al., *A comparison of machine learning techniques for customer churn prediction.* Simulation Modelling Practice and Theory, 2015. **55**: p. 1-9.
15. Naser, A. and E. Al-Shamery, *Predicting Customer Churn in Telecom Sector based on Penalization Techniques and Ensemble Machine Learning*. 2019. 199-207.
16. Krawczyk, B., *Learning from imbalanced data: open challenges and future directions.* Progress in Artificial Intelligence, 2016. **5**(4): p. 221-232.
17. Burez, J. and D. Van den Poel, *Handling class imbalance in customer churn prediction.* Expert Systems with Applications, 2009. **36**(3): p. 4626-4636.
18. Dwiyanti, E. and A. Ardiyanti. *Handling imbalanced data in churn prediction using rusboost and feature selection (case study: Pt. telekomunikasi indonesia regional 7)*. in *International Conference on Soft Computing and Data Mining*. 2016. Springer.
19. He, H. and E.A. Garcia, *Learning from imbalanced data.* IEEE Transactions on knowledge and data engineering, 2009. **21**(9): p. 1263-1284.