IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Predicting Churn: How Multilayer Perceptron Method Can Help with Customer Retention in Telecom Industry

To cite this article: NNA Sjarif *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **864** 012076

IOP Conf. Series: Materials Science and Engineering **864** (2020) 012076 doi:10.1088/1757-899X/864/1/012076

Predicting Churn: How Multilayer Perceptron Method Can Help with Customer Retention in Telecom Industry

NNA Sjarif[1,*], NF Azmi[1] HM Sarkan[1] SM Sam[1] and MZ Osman[2]

[1]Advanced Informatics Department, Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur Malaysia
[2]Soft Computing Research & Intelligent System Group (SPINT), Faculty of Systems Computer & Software Engineering, Universiti Malaysia Pahang, Lebuhraya Tun Razak, 26300 Gambang, Pahang, Malaysia

E-mail: nilamnur@utm.my

Abstract. Customer churn prediction has been used widely in various kind of domain especially subscription-basis industries. With the rapid growth of telecommunication industry over the last decade, this industry not only focuses on providing numerous products, but also satisfying the customers as it is one of the key solutions to remain competitive. This research proposed MultiLayer Perceptron Method for churn prediction. The evaluation is compared with three classifiers which includes are Support Vector Machine, Naïve Bayes and Decision Tree in term of several aspects. In preprocessing phase, we employed Principal Component Analysis and normalization to find the correlation among all the variables. For the post-processing, InfoGainAttribute is used to identify the highest factor attribute that leads to customer retention. It is found that MultiLayer Perceptron outperforms other classifiers and international plan plays important role to retain customer from leaving organization.

1. Introduction

As there are many service providers in telecommunication industry, customers are able to choose from a vast range of companies and able to switch their rights from one company to another that satisfy their requirements and preferences. Due to the number of competitors, the customers demand products and services that are perfectly tailored to suit their budget. Therefore, many telecom providers take necessary steps to deploy retention strategies as to prevent the customers from switching services to another telecom provider. In order to sustain the customers' proclivity or inclination towards the company, they will have to study the customers' behavior and in turn provide the best services in respect of the customers' preferences. This is called the 'customer churn prediction'. It is pivotal to implement the churn prediction in their approach to forecast high risk customers. In order to predict the customer churn, machine learning algorithms are used to model the data. Machine learning is a part of artificial intelligence that provide the ability to allow computer learns the algorithm automatically without human involvement. Most of the algorithm used for the telecom churn prediction are basically statistical modelling algorithm such as k-Nearest Neighbors, Support Vector Machine (SVM), decision tree, MultiLayer Perceptron Neural Network (MLP-ANN) and Naïve Bayes.

In this paper, there are several sections that have their own purposes. For section II, we have conducted some previous study on churn prediction in terms of different techniques and the results. In

IOP Conf. Series: Materials Science and Engineering **864** (2020) 012076 doi:10.1088/1757-899X/864/1/012076

section III, we briefly explain our proposed method. As for the results and evaluations, we will discuss elaborately in section IV. Section V of this paper focusing on the conclusion..

2. Literature Review

In telecommunication industry, churn is one of the key solutions in order to remain competitive. Churn is used to create a model that accurately encompasses the customer-survival and customer-hazard functions to acquire insights on churn rate [1]. According to Iyakutti et al, [2] the general definition of churn is the action of a customer terminating the customer's service due to dissatisfaction of the service offered or other companies providing better offers within the customer's budget. However, Ahmed [1] said that the customer's dissatisfaction is the main cause rather than the latter. In other words, churn prediction is a process of identifying the existing customer who are likely to discard the services in the near future. It will be a significant impact to the company's revenue if it loses customers.

There are various techniques in churn prediction to study customers' behaviour. Recently, a number of machine learning algorithms have been used for an accurate prediction in various fields. In the banking sector, from the results of using Random Forest algorithm, an ensemble of Decision Trees has been reported to manage highly accurate prediction. However, the results in telecommunication sector are not as good as in the banking industry [3].This is because the Random Forest algorithm works by generating trees on bootstrap samples which may not work well because of the imbalance of class distribution of telecommunication datasets. Ning Lu et al, [4] have suggested a model that use a boosting technique for churn prediction. This methodology uses logistic regression as the foundation followed by boosting technique to improve model prediction. However, there a few sets of limitations for both techniques used which is non-consideration of class rarity and incompetence to conclude the reasons for churn respectively. In order to enhance the customer churn and insurance fraud detection, Ravi et al, [5] suggested One Class SVM (OCSVM) based under-sampling technique. Sampling the data through OCSVM and applying machine-learning algorithm for classification resulted in a more accurate prediction of the Decision Trees than other classification algorithms with reduced system complexity.A feature selection technique propose by Sjarif [6] using the Pearson Correlation Function with k-Nearest Neighbor. The result shows that the KNN algorithm outperforms the others with the accuracy for training is 80.45% testing 97.78%.

An approach of neural network basis was suggested by Anuj Sharma et al, [7] for customer churn prediction. The data are modelled into nodes and employed using Clementine 12.0 and the over-training data problem is resolved in this approach by randomly selecting training data for network training. The evaluations prove high rate prediction of 92%. However, this model is performed for data reduction only without implementing reduction of dimensionality and eventually increase the model's complexity. In another study, Pınar Kisioglu et al, [8] suggested Bayesian-Belief-Networks (BBN) for customer churn. The authors use the Chi-squared Automatic Interaction Detector (CHAID) algorithm for discretizing the continued data variables and analyze the CDR and other customer services using the casual map as the base of BBN. This method however does not reflect the relationship between the variables. Ionut Brandusoiu et al, [9] performs machine algorithm such as Multi-Layer Perceptron (MLP), support vector machines (SVM) and Bayesian networks (BN) for customer churn prediction on the telecom data. The approach started with the pre-processing of dataset using Principle Component Analysis (PCA) then tailed by the classification of machine learning. It is found that SVM has high rate prediction accuracy than both of MLP and BN. Another implementation of machine learning algorithms namely decision trees and logistic regression was carried out by Preeti K. Dalvi et al, [10]. The suggested approach is based on performing a combination and comparative analysis of data mining techniques. In order to determine the degree of which feature affects the decision, a machine learning algorithm called logistic regression is being used. Consequently, the decision tree is employed to deliver graphical overview of the data. The evaluation results prove that the prediction accuracy is improved, and the time taken for churn prediction is reduced. However, the classification is limited to a few classes only..

3. Methodology

In this section, we briefly explain the steps of the churn prediction methodologies. Then, in Section 4 we evaluate all the churn prediction models by using prediction accuracy, precision and False Positive Rate. Finally, Section 5 concludes this research work.

3.1. Data processing

This process includes preprocessing, feature selection and post processing. For pre-processing step, this study will use two types of filter which are Principal Component Analysis (PCA) and Normalizer for the data. PCA is a technique used to describe the correlation among all the continuous independent variables using a linear combination. PCA will generate the same number of principal components as there are continuous variables in the dataset. Usually, the variety of variables of dataset can be explained by smaller number of components, meaning that the same information resides in these components as in the entire dataset. Meanwhile, Data normalization is a technique to scale the variables to have values between 0 and 1. Undergo this process, it will help to eliminate the units if measurement for data and enable the study to easily compare the data from different places. Fundamentally, normalization of the data does not mean normalizing the data, the process means normalizing residuals of the data by transforming it into value 0 to 1. Therefore, data normalization is a process of normalize the residuals by using transformation method. Next, Feature selection is used to select the important variables which tally with the purpose study. Based on the telecommunication churn dataset, there are 21 features with 3333 constraints. From 21 features available in the dataset, the study only used 15 features that is believed to correspond with customer churn prediction. For Post Processing, Information Gain Attribute evaluation is used to identify which attributes that most likely influence on customer retention

3.2. Model Construction

Artificial Neural Networks (ANN) is roused by the natural sensory systems. Numerous undertakings that people perform normally quick, for example, recognition turns out to be an exceptionally confounded assignment for a computer when conventional programming techniques are utilized. In order to predict customer churn, Neural Network technique is used to create an internal and complex structure of rule to classify different outputs. By using a learning process, basically neural networks are useful for data recognition or data classification. Fundamentally, the learning process take the adjustment of synaptic connection or the weights between neurons into accounts to map a set of input-nodes into a set of output-nodes. The application of ANN is broadly used in big data industry as it could solve a lot of data mining problems. The reason is ANN can adapt to unknown situations, robustness, fault tolerance, autonomous leaning and generalization. In this paper, Multilayer Perceptron (MLP) which is one of the class of feedforward in ANN is used to train the churn data. MLP-ANN has been known to classify the model by using back-propagation learning algorithm [11]. Apart of that, 10-folds Cross-validation will be used in this study. 10-folds Cross-validation. This paper will compare the model using three types of evaluation which are Accuracy, Mean Absolute Error (MAE) and Precision. For the first comparison, the study compares the accuracy of the model as it shows the quality of the predictive model and obvious criterion for prediction. Accuracy is measured by the ratio of correct predictions to the total number of cases evaluated. Thus, the higher the accurate the result. The second comparison is based on the Mean Absolute Error (MAE). Basically, MAE is a measure of prediction accuracy of forecasting method which calculate the magnitude of the difference between the inferred values of a quantity with the actual one and represent the value in percentage. The last evaluation is precision which calculated the ratio of correct predictions over the total number of positive cases..

4. Results and Discussion

Three methods will be compared with the propose method. The methods include Naïve Bayes (NB), Decision Three (DT) and Support Vector Machine (SVM). Table 1 and Table 2 show the comparison of the classifiers used. The analyzed techniques are compared and summarized in term of score accuracy, precision and mean absolute error.

Table 1. Principle Component Analysis + Classifiers

Classifier	Accuracy	Mean Absolute Error	Precision
SVM	91%	0.0876	0.913
MLP-ANN	91%	0.0950	0.910
Naïve Bayes	86%	0.2069	0.829
Decision Tree	87%	0.1759	0.852

Table 2. Normalize + Classifiers

Classifier	Accuracy	Mean Absolute Error	Precision
SVM	86%	0.1449	0.855
MLP-ANN	90%	0.0963	0.903
Naïve Bayes	88%	0.1844	0.865
Decision Tree	91%	0.1319	0.906

Based on Table 1, it is found that all classifiers generate high prediction accuracy which are over than 80%. The churn prediction using MLP-ANN has the highest accuracy with 91% while the Naïve Bayes generates 86% as the lowest prediction accuracy. In term of the time taken for the execution, the performance of the classifiers is similar with exception of the MLP-ANN, which is several times slower than the rest, because of the layers of nodes nature. As for the mean absolute error, the lowest value means that the classifier is better due to less error in the models. The mean absolute error obtained with SVM and MLP-ANN are 0.0744 and 0.0552 respectively, which are better than other classifiers even other classifiers also has small value of mean absolute error. MLP-ANN once again has the highest precision value while Naïve Bayes produces the lowest value. In other hands, Table 2 summarizes the results' classifiers with the use of Normalize in pre-processing phase. The result is not significantly different from Table 1. However, the highest accuracy prediction generated by Decision Tree and followed by MLP-ANN. As for mean absolute error value, MLP claimed the lowest value with 0.0963 and other classifiers are in range from 0.1319 to 0.1844. We concluded that both MLP – ANN and Decision perform better than other classifiers with the use of Normalize in terms of accuracy, mean absolute error and precision value. In short, these classifiers produce good results in many aspects that have been discussed previously. Among all the classifiers, MLP-ANN develops better results than other classifiers for both PCA and Normalize pre-processing techniques. However, the longer time for execution and consuming more memories becomes the limitation of this classifier. Compared to MLP-ANN, the time needed for the construction of these classifiers is faster even the results is not good as MLP-ANN

5. Conclusions

In this paper, we have compared three classifiers that can predict whether the customer will terminate the service and choose another organization or not. The comparison of several classifiers will help us to accurately predict customer churn as well as addressing the main factor that leads to customer retention. Initially, the PCA is performed in pre-processing phases. Then, we analyzing all the classifiers in several aspects. Based on the results and evaluation, we identified that the MLP-ANN

IOP Conf. Series: Materials Science and Engineering **864** (2020) 012076 doi:10.1088/1757-899X/864/1/012076

classifiers outperforms the others. Also, it is proven that it is vital for the organization to run more promotion for any day plan services.

Acknowledgement
The authors would like to thank Ministry of Higher Education (MOHE) and Universiti Teknologi Malaysia (UTM) for their educational and financial support. This work is conducted at Razak Faculty of

References
[1] A Ahmed and D M. Linen 2017. A review and analysis of churn prediction methods for customer retention in telecom industries. *4th International Conference on Advanced Computing and Communication Systems IEEE*. pp 1-7.
[2] K Iyakutti and V. Umayaparvathi 2016. Attribute Selection and Customer Churn Prediction in Telecom Industry *Int. Conf. Data Min. Adv. Comput.* pp 84-90
[4] N Lu, H Lin, J Lu, and G Zhang 2014. A Customer Churn Prediction Model in Telecom Industry Using Boosting. *IEEE Transactions on Industrial Informatics*, **10(2)**, pp 1659–1665.
[5] GG Sundarkumar, V Ravi,and V Siddeshwar 2015. One-class support vector machine based undersampling: Application to churn prediction and insurance fraud detection. *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*. pp 1–7.
[6] NNA Sjarif, MRM Yusof, HT Wong, S Ya'akob, R Ibrahim, and MZ Osman 2019. A Customer Churn Prediction using Pearson Correlation Function and K Nearest Neighbor Algorithm for Telecommunication Industry. *International Journal of Advances in Soft Computing & Its Applications* . **11(2)**. pp 46-59
[7] A Sharma, D Panigrahi, and P Kumar 2013. A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services. *Int. Journal of Comp. Application*. **27(11)** pp 26-31
[8] P. Kisioglu and Y. I. Topcu 2011 Expert Systems with Applications Applying Bayesian Belief Network approach to customer churn analysis : A case study on the telecom industry of Turkey *Expert Systems with Applications*. **38(6)**. pp 7151-7157
[9] I. Brandusoiu, G. Toderean, and H. Beleiu 2016. Methods for churn prediction in the pre-paid mobile telecommunications industry. *2016 International conference on communications (COMM).* pp 97–100
[10] P. K. Dalvi, S. K. Khandge, A. Deomore, A. Bankar, and V. A. Kanade 2016. Analysis of Customer Churn Prediction in Telecom Industry using Decision Trees and Logistic Regression. *In 2016 Symposium on Colossal Data Analysis and Networking (CDAN) IEEE* .pp. 1-4
[11] Y. Zhang, R. Liang, Y. Li, Y. Zheng, and M. Berry 2011. Behavior-based telecommunication churn prediction with neural network approach. *2011 International Symposium on Computer Science and Society IEEE.* pp 307-310