Prediction of Telecom Churn using Comparative Analysis of Three Classifiers of Artificial Neural Network

Youngkeun Choi, Jae Won Choi

Abstract*: *The purpose of this study is to evaluate existing individual neural network-based classifiers to compare performance measurements to improve the accuracy of deviance predictions. The data sets used in this white paper are related to communication deviance and are available to IBM Watson Analytics in the IBM community. This study uses three classifiers from ANN and a split validation operator from one data set to predict the departure of communications services. Apply different classification techniques to different classifiers to achieve the following accuracy with 75.63% for deep running, 77.63% for perceptron, and 77.95% for autoMLP. With a limited set of features, including the information of customer, this study compares ANN's classifiers to derive the best performance model. In particular, the study shows that telecom service companies with practical implications to manage potential departures and improve revenue.

Keywords: Artificial neural network, Telecom service, Churn; Deep learng, Perceptron, AutoMLP.

I. INTRODUCTION

Customers who do not renew their contract or find other service options are said to leave one airline to another. The confusion resulted in significant economic and financial losses. Deviation from the telecommunications industry is defined as the number of customers moving from one organization to another. Customer churn predictions are widely used to keep valuable customers at a cost five to seven times lower than new customers. In general, the desire to use better services leads to customer departures; if you can lower the customer departure rate to 5 to 10%, the growth rate of the company will increase to 30 to 85%. Customer secession and retention are part of the customer secession management system. Other researchers have used a variety of technologies, including data mining technology, machine learning algorithms, and statistical models, to implement customer departure management in the service industry.

Data mining is a process that applies to data with previously unknown information and is useful for some decision making. More precisely, you can extract related patterns and run large datasets in the database to design a variety of strategies for profitable business. Associative rules, SVM (support vector machines), artificial neural networks (ANNs), and DT (determining trees) are commonly used as data mining techniques that present many departure prediction models. Machine learning comes from the study and composition of algorithms in the calculation of artificial intelligence (AI) learning and pattern recognition research; machine learning algorithms have been used by many researchers to develop prediction systems. Algorithms are divided into two main categories: supervisory and non-knowledge learning algorithms; classification algorithms and regression algorithms are also included in the learning algorithm category. Classification models create models that apply to training past customer data and categorizing invisible patterns. Clustering algorithms, on the other hand, focus on similar functions that group data within a cluster and then classify invisible data into one of the related clusters. Other studies evaluate various factors such as failure, predictive performance and customer retention. Key drivers for predicting customer behavior. Many researchers have proposed a variety of solutions to retain customers by applying different technologies [1]. The researchers use algorithms, hybrid oversampling and oversampling, SMOTE oversampling technology, ESN (Ecostate Networks), clustering technology, artificial neural network, decision tree, and random forest algorithms to derive failure prediction results. PSO, mRMR, GA, and many technologies have been derived. s has been used. Ensemble classifier [1]. Early predictions of customer churn increase the likelihood of customer retention. This study estimates the deviation by evaluating different versions of individual classifiers based on the NeuralNetwork. These classifiers include the NN (Nural Network), AutoMLP (Multi-Layer Perceptron), and Deep Learning (DL). The results are calculated and evaluated using statistical measurements such as kappa, absolute error, relative error, and classification error as well as various performance measurements such as accuracy, precision, frequency, and f- measurement. The rest of the paper is as follows: Chapter 2 explains the literature review; the proposed research model is presented in Section 3, and Section 4 consists of results, evaluations and discussions. Finally, Section 5 concludes research work that suggests future research direction.

Revised Manuscript Received on July 20, 2020.

***** Correspondence Author

Youngkeun Choi*, Division of Business Administration, Sangmyung University, Seoul, Korea. Email: penking1@smu.ac.kr

Jae Won Choi, Department of Computer Science, University of Texas at Dallas,Richardson, TX, USA, Email: jxc190057@utdallas.edu

II. RELATED STUDY

Customers have smarter and more accessible options. Prices are a must for customers to consider. TW Telecom reported a $2.4 million loss of revenue and a 6.6% decline in telecommunications services due to contract renewals and loss of revenue from rising prices. AT&T, well known as a wireless operator, reported a serious crisis in its fight against the spectrum against rival T-Mobile. AT&T had to lose frequency allocations and could not provide or provide limited services to customers. If customers want to find good network services from their competitors, they will terminate immediately: AT&T rose slightly from 1.32% in the fourth quarter of 2010 to 1.36% in the third quarter. Even Twitter considers the impact of deviations on services: In less than a month, 60 percent of Twitter accounts were idle. Twitter said it is looking for ways to improve relevance and user experience by adding new services and features. Many researchers have proposed a variety of techniques to predict customer departures. To predict customer departure, a six-step data mining model was introduced that compared existing technologies such as SVM, DT, RA, NN and FL (Fuzzy Logic). Shaaban et al. [2] concluded that using a deep learning algorithm, better results could be obtained. Idris et al. [3] proposed a genetic programming (GP) algorithm and a deviation prediction model based on adaboost.GP programming worked efficiently for data retrieval, and AdaBoost was used to identify various elements of customer departure behavior in a repetitive approach. The data set was balanced using a Particle Set Optimization (PSO) undersampling method. This model worked very efficiently to solve many complex problems, improving performance by implementing effective functional selection techniques using machine learning or deep learning. Anjum et al. [4] proposed a new departure model to improve customer departure forecasts from increased recalls; the model used ensemble techniques to predict departures. The various combinations of algorithms such as C5, QUEST, CHAID, CRT, logistic regression analysis were evaluated; the best result was 93.4% of the combinations of C5 and QUEST. Data sets for experimentation and evaluation are not mentioned. Integrating customer data records can improve the accuracy of these algorithms. In order to analyze the deviance prediction model, 10 techniques were compared and studied. Ensemble-based technology (Random Forest and Adobeost) outperforms other algorithms with an accuracy of 96% in a data set containing 3333 records. The SVM and MLP (Multi-Layer Perception) show the highest performance for a given data set; however, the results were calculated using a small data set. Accuracy can be enhanced by using hybrid or the deep learning model. Verbeke et al. [5] used miners + and ALBA to propose rule-based classification techniques that exceed exit estimates These techniques focused primarily on predictive accuracy, comprehension and legitimacy. Mitkees et al. [6] discussed several reasons why customers leave the company to define behavioral attributes to predict churn. This study applied data mining technology, machine learning, pattern recognition, supporting vector machines, statistics, k-means, and DB scan algorithm to realize the actual behavior of customers. The widespread use of memory for small amounts of data is one of the main drawbacks of this study. Three main factors were discussed: predictive performance, starting and holding capacity used in the revenue model. Analysis of potential conservation models is provided for maximum benefit. Research has shown that profits and holdings have a monotonous relationship with predictions. The company can benefit from improved maintenance. Boosting was used to increase accuracy. Semrl & Matei [7] conducted a series of experiments to increase customer retention. The The Big ML and Azure ML platforms were utilized to predict departures, focusing on user services based on demographic and behavioral data.

The SVM and ESN training algorithms were used to predict customer churn using carrier data. Accuracy, good generalization, and resistance to impacting issues have helped maintain existing customers. Clustering algorithms include semantic-based subtraction clustering methods (SDSCM). Based on SCM (Subtractive Clustering Method) and AFS (self-evident fuzzy set). New algorithms increase accuracy and reduce risk. The proposed SDSCM provides efficient results compared to the k- average cluster algorithm. In the proposed SDSCM, after selecting a property, you determined the adjacent radius and finally calculated the number of clusters. Although many techniques have been presented in recent literature, the reviews provided in this section clearly illustrate the pros and cons of each method. This study aims to improve customer exit prediction performance..

III. MATHODOLOGY

A. Dataset

The dataset used in this paper was related to the disruption of telecommunications services and is available from IBM Watson Analytics in the IBM community. Each row represents a customer, and each column contains the attributes of the customer described in the metadata column.

The data set contains information about:

• Customers who left last month – this heat is called nagging.

• Services subscribed by each customer, such as telephone, multi-line, Internet, online security, online backup, device protection, technical support, streaming TV and video

• Customer account information – customer, contract, payment method, paperless billing, monthly billing and total billing period

• Customer Population-Gender, age group, partners and dependents.

B. Analysis Method

Applying preprocessing technology to data sets provides the best performance results. Pre-treatment technology handles missing values, removes duplicate values, sets roles to assign attributes, and removes outliers. Pre-process and convert to balanced data sets prior to use in your experiment.

The study uses the RapidMiner tool to perform three classifiers on ANN to predict the variation in communication services.

First, deep learning is a algorithm of machine learning based on the AI domain. Basically, it's an new version of ANN. This model contains a larger structure where several nodes are connected together and operate in the same way as the neurons in the human brain. The important purpose of training the classifier is to minimize the number of errors. Zero_one loss was applied to some invisible instances to reduce errors.

$$f : R^D \rightarrow \{0, \ldots . L\} \tag{1}$$

Zero_one loss will the written as

$$l_{0,1} = \sum_{i=0}^{|D|} If\left(x^{(i) \neq y^{(i)}}\right) \tag{2}$$

Here D is the training dataset where

$$\mathbf{D} \cap \mathbf{D}_{\text{train}=\varphi} \tag{3}$$

Due to the reservation of its optimization for large nodes it is very expensive so instead of it we mostly prefer to use Negative log-likelihood loss function as

$$NLL(\theta, D) = -\sum_{i=0}^{|D|} \log P\left(Y = \frac{y^{(i)}}{x^{(i)}}, \theta\right) \tag{4}$$

ANN works well for the DL classifier.

Second, it is the ANN type developed by Frank Rosenblatt in 1957. Perceptron is a linear classification model and is the simplest form of anterior neural network. In addition to all biological similarities, the single-layer Perceptron is a linear classification model that is efficiently trained with simple update rules. The weight vector increase or decrease perforator for all misclassified data points trains a linear classification model called a single receptor looking for separate hyperplanes (if any). This operator cannot handle multi-category attributes. Third, Auto MLP is a simple algorithm that adjusts the learning speed and size of ANN in the learning process. The AutoMLP algorithm combines the concept of genetic algorithm and probability redundancy optimization. The AutoMLP algorithm maintains a small network ensemble trained for different learning speeds and different numbers of hidden units. The error rate is determined according to the verification set according to the fixed number of Epics, and the number of hidden units and the learning rate are changed to be different to replace the poorest with the best network copy. The number of hidden units and the learning speed are derived based on the success rate and probability distribution derived from the size.

IV. RESULTS

We have been using the RapidMiner tool to calculate the results. The results were calculated for the entire RapidMiner version and worked with over 10,000 records for training and model evaluation. The classifier specified in Section 3 was applied to a data set with 19 attributes and 7,044 records. Data mining techniques can be used to predict or classify behavior by finding interesting patterns or relationships in data and fitting models based on available data. If the training data set and the test data set are separated for machine learning, the test data set must meet the following requirements: First, training data sets and test data sets must be written in the same format; second, test data sets should not be included in training data sets. Third, training data sets and test data sets must be consistent, but it is difficult to generate test data sets that meet these requirements. In data mining, various verification frameworks using one data set have been developed to solve this problem. This study supports the use of the Split Validation operator provided by RapidMiner. The operator divides the input data set into an education data set and a test data set to support performance evaluation. In this study, the relative division is selected among the subdivision parameters of the operator and 70% of the input data is used as educational data. The performance results of the data set are presented in Table 1. By applying classification techniques to different data sets, the following results are obtained with DL 75.63%, Perceptron 77.63%, and AutoMLP 77.95%.

Table- I: Performance Measure of Dataset

	DL	Perceptron	AutoMLP
True Negative	365	60	267
False Negative	196	501	294
True Positive	1,233	1,517	1,380
False Negative	319	35	172
Accuracy	75.63%	77.63%	77.95%
AUC	0.816	0.785	0.799
Precision	53.36%	63.13%	60.82%
Recall	65.06%	10.70%	47.59%
F-measure	58.63%	18.29%	53.40%

V. CONCLUSION

Many studies have been reported on pots, but no one can make universal human tools to predict pots, or say they know all the reasons. Hon is so complex and associated with many factors that researchers tend to use fewer and ignore the influence of other factors. This paper aims to use quantitative methods to understand the factors that leaders should consider. Since customer manpower is often changed and customer manpower is constantly monitored, mobile carriers can experience problems and personal information damage. Some studies examined age, gender and geographic location; however, researchers still cannot express cultural and behavioral factors that could affect the departure. The methodology used in this white paper can be seen as a roadmap for applying the one-day procedure to follow the steps taken by the reader in this case study and to predict the departure of the communication service. This study proposed an ANN classifier that is best suited to data sets in a given communication industry. This study, along with IBM Watson Analytics, leads the IBM community in exploring the reasons for the departure of communication services using data sets. The results provide a comprehensive and in-depth understanding of the factors that determine the departure of the communication service. This white paper compares classifiers of ANNs based on a limited set of features, including customer information, to derive the best performance models. Get the best results in terms of accuracy with ANN technologies such as deep running, perceptron, and autoMLP. The same model can be applied to all data sets for departure predictions to obtain the best prediction results and to use various types of classifiers to secure loyal company customers before departure.

Indeed, the study provides stakeholders such as telecom service providers with insights to manage potential departures and improve revenue. In addition, this study can provide specific work guidelines to the carrier practitioners who are trying to prevent departure when quantifying actual decision making factors. Still, the study recognizes the study's important limitations: Economic modeling is used to explore data sets and identify the link between various factors and departures. However, social or psychological factors that manage customer departure cannot be considered, so it is important to conduct a gender-sexual study to explore the grounds for deviation. Future studies of this study may include studies on other functional selection systems, such as the importance of random forest characteristics.

REFERENCES

1. Z. Zhang, R. Wang, W. Zheng, S. Lan, D. Liang, and H. Jin, “Profit maximization analysis based on data mining and the exponential retention model assumption with respect to customer churn problems,” In Data Mining Workshop (ICDMW), 2015 IEEE International Conference on (pp. 1093-1097). IEEE.
2. E. Shaaban, Y. Helmy, A. Khedr, and M. Nasr, “A Proposed Churn Prediction Model,” *International Journal of Engineering Research and Applications*, vol. 2, issue. 4, 2012, pp. 693-697.
3. A. Idris, A.Iftikhar, and Z. Rehman, “Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling,” *Cluster Computing*, vol. 22, 2017, pp.7241-7255.
4. A. Anjum, A. Zeb, L. Afridi, P. Shah, et al., “Optimizing Coverage of Churn Prediction in Telecommunication Industry,” *International Journal of Advanced Computer Science and Applications*, vol. 8, issue 5, 2017, pp. 179-188.
5. W. Verbeke, D. Martens, C., Mues, and B. Baesens, “Building comprehensible customer churn prediction models with advanced rule induction techniques,” *Expert Systems with Applications*, vol. 38, issue 3, 2011, pp. 2354-2364.
6. I. M. Mitkees, S. M. Badr, and A. I. B. Elseddawy, “Customer churn prediction model using data mining techniques,” In Computer Engineering Conference (ICENCO), 2017 13th International (pp. 262-268). IEEE.
7. I. M. Mitkees, S. M. Badr, and A. I. B. Elseddawy, “Customer churn prediction model using data mining techniques,” In Computer Engineering Conference (ICENCO), 2017 13th International (pp. 262-268). IEEE.

AUTHORS PROFILE

Young Keun Choi Division of Business Administration, School of Business and Economics, Sangmyung University, Seoul, The Republic of Korea. His research areas are business analytics, data science, entrepreneurship, technology management, and etc.

J**ae Won Choi** Department of Computer Science, Erik Jonsson School of Engineering and Computer Science, University of Texas at Dallas, TX, USA. His research interests are data science, artificial intelligence, blockchain, game, and etc..