Customer Churn Prediction in the Iranian Banking Sector

Seyed Jamal Haddadi [a, *], Mohammad Ostad Mohammadi [b], Mojtaba Bahrami [c], Elham Khoeini [d], Mehdi Beygi [d], Mehrdad Haddad Khoshkar [d]

a- Ph.D. at the Department of Electrical and Automation Systems, Federal University of Santa Catarina, Florianopolis, Brazil
b- Ph.D. at the Department of Computer Science, Sharif University, Tehran, Iran
c- Master of Science at the Department of Computer Science, Sharif University, Tehran, Iran
d- the Department of Information Technology and Infrastructure Governance, Pasargad Bank, Tehran, Iran
[*] *Corresponding Author Email: Seyed.haddadi@posgrad.ufsc.br*

2022 International Conference on Applied Artificial Intelligence (ICAPAI) | 978-1-6654-6781-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAPAI55158.2022.9801574

***Abstract*—in the financial system such as the banking sector, customers are valuable, and losing them is very expensive as customer churn is a major challenge facing banks. In this paper, we present a time series Deep Neural Networks (DNNs)-based approach for customer retention, in which a dataset has been collected from retail banking customers in the Republic Islamic of Iran. The dataset consists of real daily transactional data of about 50,000 customers in Pasargad bank in the months of November and December 2021. The goal of this study is to perform a highly churned customer predictor, attempting to observe the customer information in 30 days and predict the customer behavior in the next 30 days. Also, unlike other research in this field where the labels of customers are already determined, we present a new definition of the churned banking customer to label the data. Then, the data is cleaned, preprocessed, and prepared to import to a Bi-LSTM neural network. The proposed model has shown a significant superiority over Traditional Machine Learning Techniques. This paper can guide researchers in the field of banking and artificial intelligence, providing business knowledge to managers in the banking sector to reduce the risk of losing their customers.**

Keywords; Bi-LSTM; Deep Neural Network; Banking System; Churn Customer Prediction

I. Introduction

Nowadays, by growing at an ever-increasing amount of data, Machine Learning (ML) - a subfield of Artificial Intelligence (AI) - has gained much attention among both researchers and the industrial sector, to enhance the performance of marketing strategies [1-2]. Customer churn prediction [3], customer retention [4], credit card fraud detection [5], stock price prediction [6], and credit scoring [7] are some of the marketing tasks in the area of ML.

In this way, customer loyalty is one of the most crucial criteria in business and marketing. In fact, in most organizations, customer satisfaction is directly related to their loyalty, and these items are highly dependent on the quality of service (QoS) that organizations can provide them [8]. To evaluate the QoS, Customer Management System (CRM) can be used as a strategy of how to interact with customers and use their experience to increase customer retention for more profitability. Also, in CRM, some ML models enable to predict the customers who have decided to leave the organization and their reasons.

For this reason, customer retention plays a key role in financial groups, business, and marketing which its term means the current customer remains a customer in the future and does not reach out to competitors. The significance of customer retention becomes more when an organization knows that the cost of replacing a previous customer with a new one may be at least 5 to 20 times more depending on the business domain [9], [10]. In AI and ML, customer retention and churn prediction have made significant progress in many business domains such as telecommunications [11-12], banking [13-14], retail [15], and cloud services [16-17].

Many works are applied to address the problem of customer churn prediction using classification Traditional Machine Learning Techniques (TMLTs). In this way, in [18], the authors provided a comparison of customer churn prediction using a model including Decision Tree (DT), Random Forest, Gradient Boosted Machine Tree (GBM), and Extreme Gradient Boosting (XGBoost). The results acquired from their experiments proved that the XGBOOST had the highest accuracy with 89%. Logistic regression (LR) is another technique in TMLTs that has been widely used to examine the churn probability due to the leverage of the sigmoid function [19-20]. For instance, according to [19] which has implemented the customer attrition using two types of logistic regression, the outcome of their work indicated 95% accuracy. As well as these, Support Vector Machine (SVM) as a supervised ML technique has been widely used in churn prediction to be reached the high predictive performance. One limitation in the use of SVM may be lack of robustness in confronting the presence of a few outliers, leading to underperforming this method.

In the following, the ensemble model of gradient boosting decision tree (GBDT) is explained in [21], in which the proposed algorithm is evaluated using two types of features such as dynamic features and static features. From the results of evaluation on a large-scale customer data set, the authors concluded that the static and dynamic features are complementary, and by using them as a combination, the accuracy can be reached to 84%.

Although these techniques in some types have reached a successful performance, there are some constraints in applying the TMLTs such as decision tree and logistic regression that prevent an appropriate accuracy. Especially when the volume of data is huge and there are complex nonlinear connections between various attributes [22].

The use of Artificial Neural Networks (ANNs) is another method employed in customer retention a churn customer prediction. ANNs are ML methods inspired by the biological neural networks in the human brain [8]. Very recently, the work of [23] investigated the comparison between decision tree and neural network model, and the results indicated that the mode containing the neural network model has achieved higher accuracy than a single decision tree. In this way, the authors in [24] implemented a Convolutional Neural Network (CNN) for churn prediction in the telecommunication industry and their results have shown that the CNN has gained an accuracy of 86.85%, error rate, and F-score of 92.06% which are the appropriate performances. However, it is not compared with other NNs and TMLTs. Although the work of [25] presented an NN-based method to predict the churn of customers and has obtained an accuracy of over 92%, the reduction of feature dimensionality is not performed.

In this work, huge data on Iranian banking customers is extracted. Since the data is quite raw, hence, it is required to be properly preprocessed. To this end, a major part of this process is dedicated to preprocessing of data which its task is to prepare the data for analysis and enter to an AI prediction algorithm. Then, the proposed algorithm is compared to other classical classification techniques.

The main objectives for the work presented in this paper are highlighted threefold:

- Introduction of a novel dataset for the Iranian customer churn prediction in the banking sector.
- Introduction of new criteria to detect the churned customers.
- Experimental deep learning-based customer churn prediction.

The remainder of this paper is organized as follows. In Section II, the related work and background of classification techniques in the banking sector are explored. Section III presented the data structure and how to determine the churned customers. The methodology employed in this article such as how to preprocess the data and modeling are provided in section IV. In section V, to evaluate the proposed algorithm, some experiments and discussions are performed. Finally, section VI has explained the conclusion and future work of this research.

II. Related Work

In this section, an overview of related works on customer churn, in the banking sector using two major methods including TMLTs and deep learning methods is explained.

A. Customer Churn in Banking Sector

In the banking sector, the customers are one of the most valuable assets to finance and have a positive impact on the revenue, hence, attempting to retain them is one of the vital approaches in the bank's policies. Also, since banks, as the service providers have to compete against others for the attention of customers [26] and this competition in this industry, is growing, banks have to implement policies to retain customers.

B. TMLTs for Customer Churn in the Banking Sector

To predict the customer churn, the work of [13] has provided an improved SVM model for the Chinese banks, leveraging a random sampling method to cope with data imbalance. In [14], the decision tree method has been exerted to predict the customer churn in an electronic bank. Though in terms of the feature selection method, their work is compared in two manners such as backward elimination and forward selection, their results are not compared to other TMLTs. According to [31], a hybrid method is proposed to perform rules extraction from CRM to predict customer churn in bank credit cards, employing the SVM-recursive feature elimination (SVM-RFE) to reduce the number of features. Then, the model is obtained via SVM, and finally, rules are generated from the Naive Bayes Tree (NB-Tree). Their empirical results indicated that the hybrid proposed method outperformed other TMLTs, however, no NN-based approach is in their comparison. To extract the probability of customer churn in a banking sector, [31] compared the different TMLTs such as Logistic regression (LR), decision tree (DT), K-nearest neighbor (KNN), random forest (RF), etc. Using time-series predictors is another technique that has been applied to predict the customer churn in a retail bank in Florida, the USA by [32]. Their research used trend modeling to analyze the change of customer behavior over time and obtain the trend factor as an input to a supervised model. The results of their study proved that this dynamic classification method increased the accuracy and recall of the model overall.

C. ANNs for Customer Churn in the Banking Sector

In [41], different Neural Network methods including Radial Basis Function (RBFNN), Generalized Regression (GRNN), and Multilayer Perceptron (MLPNN) are used in a case study of Iranian banking. The results indicated that the MLPNN has had higher precision than other models. More clearly, MLP is a subset of DNN, so MLPs are always feed-forward with one hidden layer, and DNNs have two or more two hidden layers. Since deep learning has good features and representation of input data, it is widely used for predictive and analytical tools in different application domains such as customer churn prediction. For this reason, by diving into DNNs in the field of the banking sector, many various works have been carried out. For instance, authors in [42] have used the transactional data of a banking sector and implemented an attention-based hybrid Gated Recurrent Unit (GRU) Bidirectional Long Short-Term Memory (Bi-LSTM), in which their results showed that the GRU Bi-LSTM has outperformed other DNN-based methods. Another study is explored in [43], where authors have presented a novel deep ensemble classifier, in which it integrated the individual classifiers such as Convolutional neural networks (CNNs) to show the superiority of the proposed method, compared to traditional classification methods in prediction of churn customer in a retail financial institution in Canada. The aim of [44] is to forecast the risk and behavior in financial retail, in which to predict the heterogeneous patterns such as trader behavior, a DNN-based algorithm and TMLTs were developed. The results of the research have implied that the feature learning capability of deep learning caused it has better performance than TMLTs and rule-based benchmarks.

From the reviewed papers, we identified that the DNN-based algorithm can be considered as an inseparable part of the problem solution of customer churn prediction.

III. Data Structure

In this section, we aim to prepare the data to two purposes including behavior customer analysis and customer churn prediction. The data considered in this study is collected from 5 million customers in Pasargad bank in IRAN, containing transactional data. The more quality of data, the better results in analysis and prediction. To this end, the data structure required for this project, how to detect the churned or not churned customers, as well as the feature vector used by each customer are described.

A. Customer Churn Criterion

Initially, it is necessary to determine what features are involved in customer churn detection. In addition to biographical information, the transactional information of each customer is extracted and then aggregated. Indeed, we selected 45 features for each customer among them there are some aggregated transactional data including average residual amount, summation of the debtor and creditor amount, and debtor and creditor count that are employed to label the data. The transactional data used at this step is from November 1st until December 31st, 2021, and the labels are obtained from the timeline of these two months based on some defined banking rules. Note that the reason to use this timeline and comparison of two consecutive months is based on the policy and requirements of the marketing office in the banking sector to predict the customer situation next month.

The relation to detect the churned customer is formulated as equations (1-3) below:

$$\{c\_avg_{monthly} \leq \alpha * p\_avg_{monthly} \quad (1)$$

$$\begin{aligned}&\{c\_count_{creditor} \leq \alpha * p\_count_{creditor}\\ &\text{OR} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (2)\\ &\{(c\_count_{creditor} - c\_count_{debtor})\\ &\qquad \leq \alpha * (p\_count_{creditor} - p\_count_{debtor})\end{aligned}$$

$$\begin{aligned}&\{\left(c_{sum\_Amount_{creditor}} - c_{sum\_Amount_{debtor}}\right)\\ &\qquad \leq \alpha * (p_{sum\_Amount_{creditor}}\\ &\qquad - p_{sum\_Amount_{debtor}})\\ &\text{OR} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (3)\\ &\{(c\_sum\_Amount_{creditor}) \leq \alpha * p\_sum\_Amount_{creditor}\end{aligned}$$

where prefix c and p are the abbreviation of current and previous respectively. Also, $avg_{monthly}$ denotes the minimum daily average balance in a month. In addition, $sumAmount_{\ debtor}$ and $sumAmount_{\ creditor}$ are the total amounts of withdrawal transactions and total deposit transaction amounts, respectively. Besides, $count_{debtor}$ indicates number of withdrawal transactions and $count_{creditor}$ is the number of deposit transactions.

In the formulations above, α is an adjustable parameter that can be considered as a threshold that is manually regulated to 0.3. If all the three statements above be simultaneously met, the customer is entered into the churned customer category. Consequently, while in all statements, the unequal direction is changed and the α to be eliminated, the formulation representing the not churned customers will be appeared.

Finally, the label obtained in the previous step is appended to the biographical information of customers in which 1 denotes the churned customer and 0 indicates not churned. By applying equations 1 to 3, on all real customers, 23155 customers are churned and 81253 customers are not to be churned.

B. Dataset

Since the type of the problem under this study is supervised, hence, the aim is to estimate the output from algorithm input. In each machine learning supervised algorithm, 3 types of datasets are required consisting of training data, validation data, and test data. To teach the algorithm used which in this project is a neural network, the training data is utilized. Then, validation data is employed to evaluate the network performance in every step of training. Finally, test data is used to validate the progress of the algorithm's training.

Figure 1 showed the dataset partitioning in which the portion of training, validation, and test datasets are 80%, 10%, and 10%, respectively.

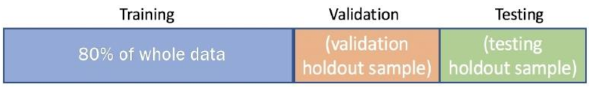

Figure 1. dataset segmentation in this study

Table I showed the detailed statistical information of dataset used in this study.

TABLE I. Statistical Information of Dataset

Dataset	Churned	Not Churned
Train	21696	21802
Validation	2472	2361
Test	2425	2408

C. Feature Vector

As it is explained earlier, the information extracted from customers corresponding to November 2021 is used to train the network. To this end, for each customer, a vector is constituted for every day of the month, in which the transactional data of customers are aggregated. This vector contains 330 types of categorical data, representing various states for each transaction. Taking into account the sum of transaction amount and number of transactions, each state is stored into a dimensional 660 matrix. Consequently, for a period of 30 days for a customer, a 660×30 dimensional matrix is made that the customer transactional information is summarized, and this data is leveraged to teach the network.

IV. METHODOLOGY

This section explains the adopted experimental methodology, containing main four steps such as Exploratory Data Analysis (EDA), pre-processing and Bi-LSTM.

A. EDA

The first time, Exploratory data analysis (EDA) was introduced by Tukey [45] who his aim was to complement formal confirmatory data analysis (CDA) – a field of data analysis that is mostly concerned with statistical hypothesis testing, confidence intervals, estimation [46]. One of the methods in EDA is to employ data visualization to perceive the initial insight from the dataset.

As it can be explicitly observed in Figure 2, the summation of transactions count in every day of the month reached a dip at the weekend to about 3 billion transactions. Also, the graph experienced a peak at 5 billion transactions at the end of the month which is a usual behavior from the side of customers.

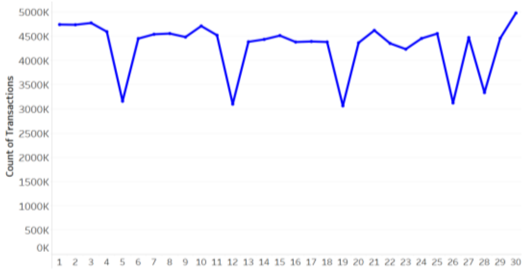

Figure 2. The count of transactions over 30 days

In this regard, from Figure 3, it can be inferred that the amounts of transactions are in accordance with the count of transactions in Figure 2 such that the minimum amount summation of transactions have occurred at the end of every week to approximately 15000 billons Rial.

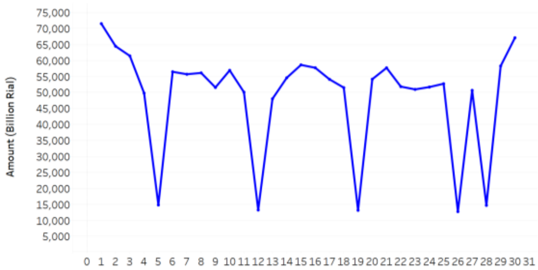

Figure 3. the amount summation of transactions over 30 days

B. Data Preprocessing

Data preprocessing plays a key role in machine learning problems, which is used to effectively reduce forecasting errors and improve accuracy. To this end, data cleansing – such as handling missing values - data normalization, discretization, transformation, and feature extraction are the proper preprocessing steps, leading to better performance. In this way, feature selection can be considered as a reduction process of the number of input variables to improve the final prediction accuracy, which in this study, ExtraTreesClassifier [48] - is an ensemble learning technique to seek the most important features.

Generally speaking, initially, data is daily extracted from Data Ware House (DWH) for a period of three months. Next, data cleaning and feature selection are performed. Then data is aggregated based on the summation of some features such as transaction count (debtor and creditor) transaction amount (debtor and creditor). Finally, the data is formed into a time series data frame for each customer. It means that for every customer, there are 30 rows of records containing transactional behaviors.

C. Bidirectional Long Short Term Memory (Bi-LSTM)

LSTM is the most well-known type of Recurrent Neural network (RNNs), consisting of LSTM cells where each word corresponds to a cell. Also, except that instead of each self-connected hidden unit, there is a kind of memory unit. Even though the RNNs are designed to deal with variable lengths sequences, they suffer from some problems the main one is vanishing or exploding gradient. Indeed, these problem cases to prevent the learning of long-term dependencies. In this process, while calculating the error gradients, the domination of the multiplicative term increases over time, leading to the gradient becoming smaller and smaller as far as the weights of the model may converge to zero and no training is exerted [47]. This problem can be handled using LSTM, congaing memory units.

In a traditional LSTM network, one limitation is to consider the past sequence in calculating the output of a hidden unit. To cope with this limitation, a bidirectional structure of LSTM can be employed. The basic idea of Bi-LSTM is that each training sequence is connected to two separated hidden units and in two ways backward and toward. In fact, in Bi-LSTM, we can preserve information from both the past and future.

Figure 4 indicated the LSTM Neural Network architecture designed in this study.

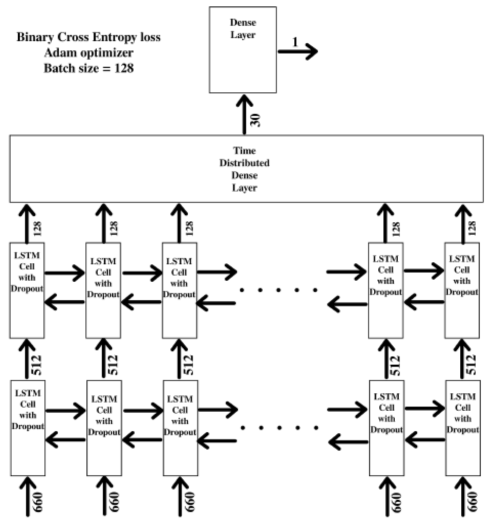

Figure 4. LSTM Neural Network architecture designed in this study

V. EXPERIMENT RESULTS AND DISCUSSION

In the experiments, apart from the proposed model, the other four machine learning classification models were also performed. The subsequent sections explain the evaluation metrics, results acquired, and discussion. The experiment has been executed on a powerful Graphic Processor Unit (GPU) Server utilizing 750 CPU Cores 2.70 GHz with 800 GB RAM, running Ubuntu 20.04 LTS.

The experimental test for this study was commenced using the designed Bi-LSTM and hyperparameters tuned mentioned in section V, and then its result is compared to some classification machine learning methods.

Based on the results mentioned in Table II, the performance of the proposed model showed the highest accuracy (84%) among others. From the given table, it can be inferred that the SVM and Logistic Regression with around 78% % and 74% accuracy, respectively, outperformed Naïve Bayes and Decision Tree. On the other hand, the Decision Tree has the lowest accuracy, compared to all the methods used in this study.

TABLE II. PERFORMANCE COMPARISON FOR THE PROPOSED METHOD, SVM, NAÏVE BAYES, LOGISTIC REGRSSION AND DECISION TREES

Model Name	Accuracy	F1-Score	Recall	Precision
Decision Trees	0.64	0.65	0.66	0.61
Naïve Bayes	0.66	0.62	0.64	0.63
Logistic Regression	0.74	0.69	0.71	0.70
SVM	0.78	0.75	0.73	0.71
Our Model	**0.84**	**0.83**	**0.82**	**0.82**

Since in this project, our focus is on the proposed neural network, it is crucial to know the effect of the number of epochs on the accuracy and validation loss. We have used an approach called early stopping which it aims to avoid overfitting during training. Although the number of epochs hyperparameter for this model was configured to 20, due to leveraging early stopping, training was stopped at epoch 14.

Figures 5-a and 5-b demonstrate the trend of training accuracy and validation loss while the number of epochs is increasing.

In Figure 5-a, by increasing the number of epochs, the train accuracy and validation accuracy are raised. In contrast, in Figure 5-b, the validation loss and train loss are reduced by growing the number of epochs.

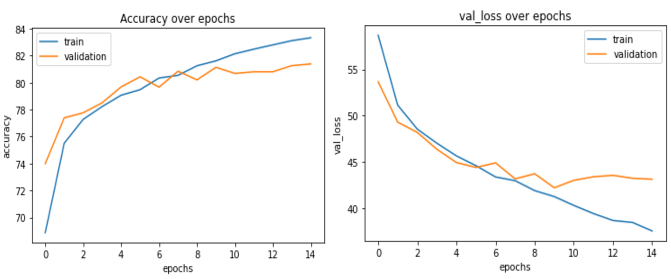

Figure 5. a) Trend of accuracy over number of epochs, b) Trend of validation loss over number of epochs

VI. CONCLUSION AND FUTURE WORK

In this study, we presented a new definition for churned banking customers for a period of one month. Unlike most churn customer predictions, our strategy is to perform customer churn prediction with the approach of a time series algorithm can be an appropriate methodology to predict the customers tending to leave the bank. The reason is because of that the daily transactional behavior of customers is trained and it is possible to predict the churned customers at a certain period in the future (next 30 days). First, the daily raw transactional is extracted from DWH for about 50,000 banking customers. Then, the data is cleaned, aggregated, and preprocessed to be inserted into the algorithm. In addition, we leveraged a 5 layers DNN-based method called Bi-LSTM which is a type of RNN. The performance of the proposed algorithm proved that it outperformed other classic classifiers such as Logistic Regression, Naïve Bayes, SVM, and Decision Trees. Two of the challenges in implementation were a large amount of dataset and its high dimensionality, where TMLTs are sensitive to them. Also, thanks to taking advantage of an existing powerful server, our algorithm can be regulated to predict the customers for any specified period in the future.

Finally, there are still some works to be performed for the future, such as using hybrid neural networks in form of the time series algorithm, and designing an auto-self tuner DNN-based to automatically tune the hyperparameters with any size of banking transactional data.

ACKNOWLEDGMENT

This investigation is established at the department of Information and Technology (IT) and software office in Pasargad bank - a leading bank in the field of smart banking - in the Republic Islamic of Iran.

REFERENCES

[1] Hsu, Tien-Yu. "Machine learning applied to stock index performance enhancement." *Journal of Banking and Financial Technology* (2021): 1-13.

[2] Henrique, Bruno Miranda, Vinicius Amorim Sobreiro, and Herbert Kimura. "Literature review: Machine learning techniques applied to financial market prediction." *Expert Systems with Applications* 124 (2019): 226-251.

[3] Vafeiadis, Thanasis, et al. "A comparison of machine learning techniques for customer churn prediction." *Simulation Modelling Practice and Theory* 55 (2015): 1-9.

[4] Simanjuntak, Megawati, et al. "Enhancing customer retention using customer relationship management approach in car loan bussiness." *Cogent Business & Management* 7.1 (2020): 1738200.

[5] Awoyemi, John O., Adebayo O. Adetunmbi, and Samuel A. Oluwadare. "Credit card fraud detection using machine learning techniques: A comparative analysis." *2017 International Conference on Computing Networking and Informatics (ICCNI)*. IEEE, 2017.

[6] Mehtab, Sidra, Jaydip Sen, and Abhishek Dutta. "Stock price prediction using machine learning and LSTM-based deep learning models." *Symposium on Machine Learning and Metaheuristics Algorithms, and Applications*. Springer, Singapore, 2020.

[7] Dumitrescu, Elena, et al. "Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects." *European Journal of Operational Research* (2021).

[8] Sabbeh, Sahar F. "Machine-learning techniques for customer retention: A comparative study." *International Journal of Advanced Computer Science and Applications* 9.2 (2018).

[9] The Chartered Institute of Marketing, Cost of customer acquisition versus customer retention (2010).

[10] Colin Shaw, CEO, Beyond Philosophy, 15 Statistics That Should Change The Business World – But Haven't, Featured in: Customer Experience, June 4, 2013.

[11] Rahul J. Jadhav, Usharani T. Pawar, "Churn Prediction in Telecommunication Using Data Mining Technology", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No.2 , February 2011

[12] Adnan Amin, Sajid Anwar, Awais Adnan, Muhammad Nawaz, Khalid Alawfi, Amir Hussain ,Kaizhu Huang, "Customer churn prediction in the telecommunication sector using a rough set approach", Neurocomputing Volume 237, Pages 242-254,2017.

[13] Ben lanHe , Yong Shi, Qian Wan, XiZhao, "Prediction of Customer Attrition of Commercial Banks based on SVM Model",2nd International Conference on Information Technology and Quantitative Management, ITQM, Procedia Computer Science Volume 31, Pages 423-430, 2014.

[14] Alisa Bilal Zorić, "PREDICTING CUSTOMER CHURN IN BANKING INDUSTRY USING NEURAL NETWORKS", nterdisciplinary Description of Complex Systems 14(2), page:116-124, 2016.

[15] M. Clemente, V. Giner-Bosch, and S. San Matías, "Assessing classification methods for churn prediction by composite indicators", Dept. of Applied Statistics, OR & Quality,Universitat Politècnica de València, Camino de Vera s/n, 46022 Spain, 2010.

[16] Anthony E. R. Sukow, Rebecca Grant, "Forecasting and the Role of Churn in Software-as-a-Service Business Models", iBusiness, Vol. 5 No. 1A, 2013, pp. 49-57.

[17] Yizhe Ge ; Shan He ; Jingyue Xiong ; Donald E. Brown, "Customer churn analysis for a software-as-a-service company", In the proceedings of Systems and Information Engineering Design Symposium (SIEDS), 2017.

[18] Ahmad, Abdelrahim Kasem, Assef Jafar, and Kadan Aljoumaa. "Customer churn prediction in telecom using machine learning in big data platform." *Journal of Big Data* 6.1 (2019): 1-24.

[19] Oghojafor, B. E. A., et al. "Modelling telecom customer attrition using logistic regression." *African journal of marketing management* 4.3 (2012): 110-117.

[20] Andrew h. Karp, using logistic regression to predict customer retention, 1998.

[21] Wang, Qiu-Feng, Mirror Xu, and Amir Hussain. "Large-scale ensemble model for customer churn prediction in search ads." *Cognitive Computation* 11.2 (2019): 262-270.

[22] Ullah, Irfan, et al. "A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector." *IEEE Access* 7 (2019): 60134-60149.

[23] Hu, Xin, et al. "Research on a customer churn combination prediction model based on decision tree and neural network." *2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*. IEEE, 2020.

[24] A. Mishra and U. S. Reddy, ''A novel approach for churn prediction using deep learning,'' in Proc. IEEE Int. Conf. Comput. Intell. Comput. Res., Dec. 2017, pp. 1–4.

[25] Anuj Sharma, DrPanigrahi, and Prabin Kumar. "A neural network based approach for predicting customer churn in cellular network services." arXiv preprint arXiv:1309.3945, 2013.

[26] Hudaib, Amjad, et al. "Hybrid data mining models for predicting customer churn." *International Journal of Communications, Network and System Sciences* 8.05 (2015): 91.

[27] Tsai, Chih-Fong, and Yu-Hsin Lu. "Customer churn prediction by hybrid neural networks." *Expert Systems with Applications* 36.10 (2009): 12547-12553.

[28] Shirazi, Farid, and Mahbobeh Mohammadi. "A big data analytics model for customer churn prediction in the retiree segment." *International Journal of Information Management* 48 (2019): 238-253.

[29] Karvana, Ketut Gde Manik, et al. "Customer churn analysis and prediction using data mining models in banking industry." *2019 International Workshop on Big Data and Information Security (IWBIS)*. IEEE, 2019.

[30] Xia, Guoen, and Qingzhe He. "The Research of online shopping customer churn prediction based on integrated learning." *Proceedings of the 2018 International Conference on Mechanical, Electronic, Control and Automation Engineering (MECAE 2018), Qingdao, China*. 2018.

[31] Farquad, Mohammed Abdul Haque, Vadlamani Ravi, and S. Bapi Raju. "Churn prediction using comprehensible support vector machine: An analytical CRM application." *Applied Soft Computing* 19 (2014): 31-40.

[32] Leung, Hoiyin Christina, and Wingyan Chung. "A dynamic classification approach to churn prediction in banking industry." (2020).

[33] De Caigny, Arno, Kristof Coussement, and Koen W. De Bock. "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees." *European Journal of Operational Research* 269.2 (2018): 760-772.

[34] Olaniyi, Abdulsalam Sulaiman, et al. "Customer Churn Prediction in Banking Industry Using K-Means and Support Vector Machine Algorithms." *International Journal of Multidisciplinary Sciences and Advanced Technology* 1.1 (2020): 48-54.

[35] Keramati, Abbas, Hajar Ghaneei, and Seyed Mohammad Mirmohammadi. "Investigating factors affecting customer churn in electronic banking and developing solutions for retention." *International Journal of Electronic Banking* 2.3 (2020): 185-204.

[36] Anil Kumar, Dudyala, and Vadlamani Ravi. "Predicting credit card customer churn in banks using data mining." *International Journal of Data Analysis Techniques and Strategies* 1.1 (2008): 4-28.

[37] Mozer, Michael C., et al. " Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry." *IEEE Transactions on neural networks* 11.3 (2000): 690-696.

[38] Au, Wai-Ho, Keith CC Chan, and Xin Yao. "A novel evolutionary data mining algorithm with applications to churn prediction." *IEEE transactions on evolutionary computation* 7.6 (2003): 532-545.

[39] Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik. "A training algorithm for optimal margin classifiers." *Proceedings of the fifth annual workshop on Computational learning theory*. 1992.

[40] Shaaban, Essam, et al. "A proposed churn prediction model." *International Journal of Engineering Research and Applications* 2.4 (2012): 693-697.

[41] Gholamiangonabadi, Davoud, et al. "Customer Churn Prediction Using a New Criterion and Data Mining; A Case Study of Iranian Banking Industry." *Proceedings of the International Conference on Industrial Engineering and Operations Management*. 2019.

[42] Britto, Mr M. John, and R. Gobinath. "Improved Churn Prediction Model In Banking Industry And Comparison Of Deep Learning Algorithms.", International Journal of Aquatic Science, (2021).

[43] Chen, Yuzhou, et al. "Deep ensemble classifiers and peer effects analysis for churn forecasting in retail banking." *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, Cham, 2018.

[44] Kim, Alisa, et al. "Can deep learning predict risky retail investors? A case study in financial risk behavior forecasting." *European Journal of Operational Research* 283.1 (2020): 217-234.

[45] J. W. Tukey. Exploratory data analysis, vol. 2. Reading, Mass., 1977.

[46] Martinez, Wendy L., Angel R. Martinez, and Jeffrey L. Solka. *Exploratory data analysis with MATLAB®*. Chapman and Hall/CRC, 2017.

[47] Reimers, Nils, and Iryna Gurevych. "Optimal hyperparameters for deep lstm-networks for sequence labeling tasks." *arXiv preprint arXiv:1707.06799* (2017).

[48] http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html.