2022 International Conference on Information and Communication Technology for Development for Africa (ICT4DA) | 978-1-6654-5587-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICT4DA56482.2022.9971224

Customer Churn Prediction Using Machine Learning: Commercial Bank of Ethiopia

Muhamed Hassen Seid
Cyber Security Research Division, Data Science
Information Network Security Administration (INSA)
Addis Ababa, Ethiopia
muhamed.hassens@aau.edu.et

Michael Melese Woldeyohannis
School of Information Science
Addis Ababa University
Addis Ababa, Ethiopia
michael.melese@aau.edu.et

***Abstract*—Identification of churned customers is critical to the operation and growth of any business. Identification of churned customers can assist businesses in understanding the reasons for churn and planning market strategies to boost business growth. The purpose of this study is to design and develop a machine learning model that can accurately predict churned customers from the total customers of the Commercial Bank of Ethiopia (CBE) in order to retain existing customers. A total of 204,161 datasets with eleven attributes were used for this study. For this study, the overall accuracy of the model was used as the evaluation metric to determine the best classifier. Accordingly, supervised machine learning methods such as Logistic Regression, Random Forest, Support Vector Machine, K-Nearest Neighbor, and Deep Neural Network were used to predict customer churn in a Commercial Bank of Ethiopia (CBE) context. Based on previous literature, these classifier algorithms for customer churn prediction have been widely used. In this study, feature importance and a correlation matrix were used to select features. Furthermore, the SMOTE technique is used to balance the data, and the results for the chosen algorithm were evaluated and compared. Among the various experiments performed, a Deep Neural Network (DNN) outperformed with an accuracy of 79.32%, precision of 85.08%, and recall of 78.19%.**

***Index Terms*—Churn Prediction, Customer Churn, Machine Learning, Deep Learning**

I. Introduction

In various markets around the world, an increasing number of customers are switching their registered services between competing companies such as telecommunications, banking, and insurance, and customers who stop using the company's products or services are referred to as customer churn [1] [2]. These businesses have realized that they should focus their marketing efforts on customer retention rather than customer acquisition. The cost of acquiring new customers is much higher than the cost of customer retention in a competitor service. Applying retention strategies becomes even more important in the case of mature businesses, where the customer base has peaked and retaining customers is critical [3]. For a variety of reasons, customers may wish to end their relationship with a company.

In today's competitive world, customer churn is one of the primary concerns of banking industries, particularly when customers complete transactions and leave a company [4]. This could result in a loss of income for the company and make it difficult to retain customers. As a result, it is critical to identify customers who may leave the bank in the near future.

Banks benefit greatly in terms of return on investment (ROI) and the growth of customer portfolios over time. To distinguish potential churn risks, the traditional approach to churn investigation employs rules based on actual practices [5]. Rules-based approaches are rigid and lose many customers who churn, as well as generate false positives that result in costly incentives for customers who may not be at risk. As a result, rather than a rule-based technique like Artificial intelligence (AI), an automatic way to churn customers is required.

AI is a discipline that studies and develops theories, methods, technologies, and applications by simulating the extension and expansion of human intelligence [6]. AI is an excellent solution for predicting customer churn because the problem involves complex data over time and interactions between various customer behaviors that can be difficult for humans to distinguish [5]. To determine risk, AI can examine a wide range of data, including new data sources, as well as somewhat complex interactions among behaviors, and compare it to individual history. It can also be utilized to suggest the best offer that most probably retains a valuable customer.

Machine learning is a sub-field of AI that allows machines to learn from training data [4]. Over the last decade, various machine learning techniques have been used to predict churn. Deep learning is a sub-field of machine learning algorithms that includes many hidden layers. Deep learning algorithms train each layer of neurons on the previous layer's data features. Machine Learning has become one of the backbones of information technology over the last two decades, and with it, a rather central, albeit often hidden, part of our lives. With the availability of ever-increasing amounts of data, there is reason to believe that smart data analysis will become even more absolutely indispensable as a necessary component of technological advancement [7].

Retaining existing customers, on the other hand, is critical for businesses looking to expand without relying too heavily on the significantly higher cost of acquiring new customers. To reduce churn, marketing, sales, and customer retention departments must collaborate to ensure customer satisfaction, provide incentives, and present offers at the right time. Customer retention is an important component of banking strategy in today's increasingly competitive environment. Banks around the world are struggling to maintain their competitiveness, which is limiting their traditional revenue streams. Customer retention measures a company's success not only in acquiring new customers, but also in satisfying existing customers. It also increases return on investment, increases loyalty, and attracts new customers. Because of the aforementioned issues, accurate and timely identification of those customers is critical in reducing the value of a company's overall retention marketing strategy. As a result, machine learning algorithms are required to analyze datasets and then draw inferences or make predictions based on identified patterns.

Many researchers stated that various industries operate in a highly competitive environment [8]. Customer churn is now the most common concern of companies operating in industries with low switching costs [9]. Organizations suffer greatly from churn and the cost of acquiring new customers is significantly higher than the cost of maintaining and upgrading existing customer relationships [10]. For example, the more customers you lose, the more money you must spend to recoup your losses by acquiring new ones. Churning of existing customers most likely results in the loss of businesses and, as a result, a decrease in profit [11]. Loss of customers not only results in a missed opportunity due to reduced markets, but it also increases the need for attracting new customers, which is five to six times more expensive than customer retention [12].

Furthermore, it was discovered that churn of good customers has irreversible consequences for a well-known company [13]. Among all industries affected by this issue, Commercial Bank of Ethiopia (CBE) is frequently included on the customer churn list, and a variety of customers are leaving CBE for unknown reasons [14]. CBE, on the other hand, has a large customer base that is still growing as a result of the bank's various promotional strategies to attract new customers [15]. If this trend continues, it will have a negative impact on CBE's progress, even though the number of new subscribers appears to be increasing.

As the number of customers recorded in a bank's database grows, banks face difficulties in understanding their customers' behaviors and desires, as well as challenges in effectively implementing customer relationship management processes [13]. The massive amount of data makes it difficult for organizations to manually investigate and retrieve useful information as needed by decision-makers. However, using machine learning techniques, we can gain valuable insights that will help you better understand our customers' needs and inform overall decision-making.

Because the size of knowledge grows and the number of dimensions increases, the manual process of data analysis becomes tedious; thus, the process of data analysis must be computerized [16]. Whereas machine learning helps to develop smart market decisions, run accurate campaigns, make predictions, and more from massive amounts of data, we can analyze customer behaviors and their insights with the help of machine learning. Because of the large amount of data [17], this leads to great success and data-driven business. The development of various machine learning algorithms, as well as the availability of a large amount of customer-related data within the banking industry, created a better opportunity to support the CRM Process, particularly for customer retention. However, only a few banks are taking advantage of technology by segmenting customers based on homogeneity in behavior, predicting potential customers and credit defaults, and predicting churning customers based on their behavior [10].

To address the aforementioned gaps, this study focused on designing and developing a predictive model using machine learning techniques; and it also helps to know the churners before they churn.

II. Related Works

A number of researchers attempted to study customer churn in today's competitive world [18]–[24]. The majority of the studies focused on using machine learning algorithms to analyze customer data and predict customer churn.

Among these research studies, customer churn prediction via credit card using specific attributes based on economic data gathered from a real Chinese bank [18]. They design mis-classification cost estimation by taking the two types of error and the economic sense into account, in addition to the accuracy of analytic results. The findings were more suitable for evaluating credit card churn prediction using logistic regression and decision tree, as demonstrated by experienced and powerful classification algorithms.

Moreover, another attempt has been made to address the banking sector's turnover problem through client retention using predictive data mining techniques [19]. The study used classification and regression trees to improve overall classification rates by incorporating customer behavior, customer perceptions, customer demographics, and macro-environmental factors. Similarly, a dataset collected from a Nigerian bank was used to cluster using K-means, followed by JRip to generate a prediction rule using the WEKA tool for knowledge analysis [20].

Furthermore, a quarterly churn prediction that simulates a real-world scenario was made using an unbalanced 7,190 (6,512 non-churners and 678 churners) customers drawn at random from the banking industry [21]. For this, a robust and sequential based learning Adaptive Boost (AdaBoost) algorithm is used to provide a better separation of a high-risk customer from highly skewed class distribution.

In addition to this, a customer churn prediction on the Telecommunication industry were made to minimize the structural risk minimization using different machine learning algorithm [22]. The experiment result shows that the support vector machine (SVM) outperformed with the best accuracy, hit, and covering rate over the other techniques on the same dataset.

Beside the telecommunication and banking industry, a financial industry also suffers from the customer churn which mainly with Spatio-temporal data [23]. In the study, Spatio-temporal and choice features were more superior to the demographic features in financial churn prediction. It is also revealed that young people more probably to leave the bank.

Furthermore, the research attempt made on the customer churn prediction for Commercial Bank of Ethiopia were made using 13,172 customers with 9 attributes and their corresponding 628,634 transactions with 10 attributes are collected from the bank [24]. From this a total of 6045 instances and 18 attributes is prepared. The results of the study show that the J48 modeling technique is the best model with a performance of 94.8% followed by bagging (93.9%) and Logistic Regression (76.6%).

A. Summary of Related Works

The topic of customer churn prediction has discussed various customer churn algorithms, which are relevant to the current work. Different methods have been proposed to classify multiple customer churn problems and each approach has numerous advantages and limitations. While the performance of existing approaches has been improved substantially, there is still abundant room for further progress on the current topic.

Some of the researchers have not handled the class imbalance problem. Additionally, most of the above-listed research was done in developed countries and they have used archived data, especially taken from Kaggle and also which cannot easily describe the real-world problem. Most of the above-listed studies have used a small number of datasets, which may not describe the overall problem and it may not develop and select a better model. some studies incorporated strongly related attributes in different columns (redundant attribute), not used normalization, the models were vulnerable to over-fitting, and the studies have not used neural networks.

To address the limitations and research gaps presented in this section, this research used 204,161 instances with 11 attributes, it has used new features such as atm user or not, mobile banking user or not, and so on. This study has focused on covering the data preparation steps, and handling class imbalance using SMOTE technique, and also this study has applied five machine learning algorithms, which are Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbour (KNN), and Deep neural network (DNN) algorithms are used in customer churn prediction are applied and the algorithm with the best predictive performance is to be selected for model performance optimization. Though the aforementioned researches have a substantial contribution to showing the directions in conducting this research, this area (customer churn prediction) is not been studied much locally. So, the study fills the knowledge gap as to how churners can be predicted from existing historical data.

III. Methodology

In attempt to design and develop a machine learning model for customer churn prediction, a number of methodologies followed. From these, Section III-A presents the data collection detail followed by the data pre-processing including feature selection and extraction in Section III-B, then system design and architecture III-C.

A. Data Collection

Data is required for machine learning tasks, so it must be collected [25]. This study focused on the CBE; thus, during the data collection phase of the research methodology, a total of 204,161 instances dataset were collected from potential data sources at CBE using a simple random sampling method. The collected data has 11 attributes. These are; ATM card status which is whether the customer used ATM or not; Mobile banking status is whether the customer used mobile banking or not; Book type is different book account types of CBE; Age is denoted the age of the customers of CBE; Gender is which describe the customer sex; Industry name is the type of organization he/she work; Number of credit transaction: is the number of credit transactions for consecutive three months; Number of debit transaction is the number of debit transactions for consecutive three months; Tenure is the length of time a customer has a customer for CBE; Working balance is current balance in birr on the account; and Churn is the target variable which has a Boolean variable that denotes whether a customer was churned or non-churned. Table I presents the distribution of the data collected from CBE.

TABLE I
TARGET VARIABLE COUNT

Dataset	Instance	Percentage Share
Churn (1)	44,591	21.8%
Non-Churn (0)	159,570	78.2%
Total	204,161	

As depicted in Table I, The 'Churn' is the target variable with values as '1' for those who have Churn their customers with CBE and '0' for those who are still the customers of CBE. In this study, a total of 204,161 instances of customers were used, with 21.8% (44,591) of the total datasets being Churn and 78.2% (159,570) being Non-Churn. A class imbalance ratio of around 39:11 is discovered, implying that for every 39 non-churned clients, 11 will be churned.

B. Data Pre-processing

One of the most critical phases of machine learning is data pre-processing [26], [27]. This signifies that throughout the data preparation phase, all operations to turn the raw data into the final dataset were completed. Several data pre-processing processes were performed based on the preceding phase's data knowledge to create the final dataset for the experiment.

The data pre-processing phase included handling missing values, which are used in Python for imputing missing values in the dataset by mode values for categorical variables including Gender, Industry name, and Mobile banking status. Like categorical data, the continuous data with missing values were imputed by mean values. The mean was a reasonable estimate for randomly selected observations from a normal distribution. Such as age, number of credit transaction, number of debit transaction, working balance, and tenure were handled by mean values. This is followed by encoding, encoding means the process of converting categorical data into numeric data. In this study a total of 7 attributes were categorical values or nominal values, these values were transformed into numerical using sklearn's LabelEncoder function. After encoding the dataset normalization technique is applied, normalizing the dataset by utilizing the Sklearn MixMaxScaler() function between 0 and 1; feature extraction, standardizing the data, data splitting, and finally the SMOTE approach are used to balance the unbalanced data.

C. System Design and Architecture

Design Science Research (DSR) is a problem-solving paradigm that aims to strengthen human knowledge via the creation of innovative artifacts. The approach that strives to strengthen scientific and technology knowledge bases through the development of new artifacts that solve issues and improve the environment in which they are instantiated [28].

While it is widely understood that design science research results in design objects, the mechanisms by which these artifacts are created vary amongst investigations. This study describes the architecture and proof-of-concept implementation of models for churn prediction based on different machine learning algorithms, with CBE customer churning as the background scenario. The use of machine learning methods offers the promise of detecting new patterns of churn in the given training datasets.

The machine learning architecture defines the various ways involved in the machine learning cycle and involves the major steps being carried out in the transformation of raw data into training data sets capable of enabling the decision-making of a system. Fig. 1 presents the general architecture of the customer churn prediction.

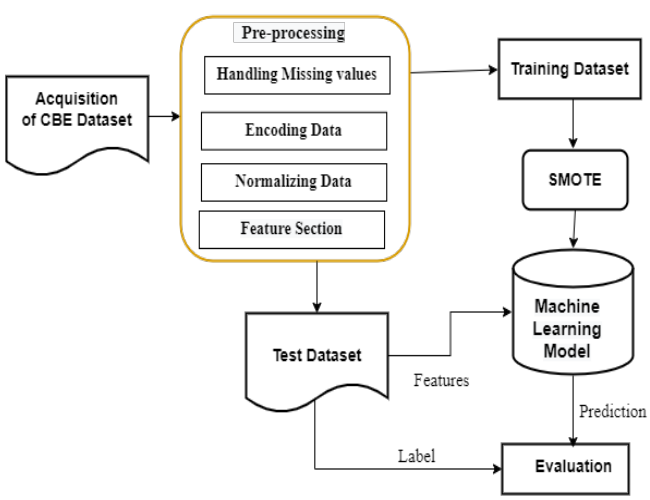


Fig. 1. Architecture of Customer Churn Prediction

As depicted in Fig. 1, the architecture of the customer churn prediction starts with data collection followed by basics of data pre-processing including handling missing value, data encoding and normalisation before the feature extraction phases. Once the feature extracted, the data is divided in to training and testing with 80% and 20% respectively. Since the training data is imbalanced, Synthetic Minority over sampling technique (SMOTE) applied to balance the imbalanced data. Then, the training data is used for creating a machine learning model to be evaluated using the 20% of test data split and finally, evaluate the performance of the model using the reference label.

IV. Experiment and Result Discussion

A series of experiments were conducted in an attempt to evaluate and select machine learning models for customer churn prediction. Section IV-A presents the experimental setup and the different hyper-parameters used for each experiment. While the experiment Section IV-B presents the detailed machine learning algorithm along with the discussion for customer churn prediction.

A. Experimental Setup

Various methods and techniques are utilized to implement the customer churn prediction model. For the purpose of implementing the suggested system, all tests have been performed on a laptop set up as follows: 8 GB of RAM, an Intel(R) Core i7 processor clocked at 1.80 GHz, and Windows 10 as the operating system. The algorithms have been implemented using software tools, Python is a dynamically typed programming language that is high-level, interpreted, and general-purpose programming language. The language also support and sklearn, TensorFlow, and Keras. TensorFlow is a feature-rich, extensible ecosystem of tools, libraries, and community resources that allows academics to push the frontiers of machine learning and developers to swiftly build and deploy ML-powered apps [29]. Similarly, Keras uses Tensorflow as a backend and follows best practices for decreasing cognitive burden [30]. It also offers consistent APIs, which decrease the amount of user activity required for ordinary applications.

B. Experiment

In this study, five distinct supervised machine learning algorithms were used to develop a model for the given dataset. These algorithms are Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Deep Neural Network (DNN). The part that follows discusses the experimental explanation of each experiment as well as their setting using the experimental setting discussed in IV-A. Beside this, all the experiment conducted under the 80% of the total data for training while the remaining 20% for testing purpose.

The first experiment is carried out using logistic regression with a 'lbfgs' parameter solver, which does not have any essential hyper-parameters to modify. Then, on average, the Logistic Regression classification performance indicators are 63.64% accuracy, 81.65% precision, 56.97% recall, and 66.71% f1 score. The outcome might be explained by the target label having no linear association with the characteristics. Because logistic regression (or linear regression for regression issues) has a linear decision surface, it cannot predict targets with high accuracy (even on training data), making it unsuitable for complicated data.

The skewness of the dataset for the problem might explain why the logistic regression performed poorly.

Like the logistic regression, the second experiment is conducted using Random Forest (RF) approach is based on ensemble learning, in which numerous machine learning models are combined to create superior predictions on a dataset. The Random Forest was employed in this investigation. The parameters n estimators is 100, criteria is "gini," and with a maximum depth of five. The Random Forest classification performance measures are then produced, with an average of 78.52%, 84.85% precision, 77.06% recall, and 80.10% f1 score.

Similarly, the third experiment conducted using support vector classifier. The classifier model has employed 'linear' kernel parameters to reduce running time (efficiency). The kernel function is the most significant parameter in the Support Vector Machine model. The kernel function was set to linear in this investigation, and it divides the class linearly using a single line. It comes in handy when the dataset is vast. The linear kernel's key benefit was its speed of processing. The classification performance of the Support Vector Machine reveals 61.75%, 81.36% precision, 54.42% recall, and 64.88% f1 score.

Like the other, the fourth experiment is conducted using K-nearest Neighbor algorithm. In the K-Nearest Neighbor model, the n-neighbors function is the most important parameter. In this paper for the KNN model (K=505), n-neighbors was selected based on the rule of tump [31], which is the square root of the training datasets (255,252 datasets used for training). The KNN algorithm was often used because of its simplicity. The main advantage of finding the optimal value of K was to indicate the number of nearest neighbors to include in the majority voting process. Accordingly, the experiment result K-nearest neighbors algorithm with 505 neighbors resulted a 73.94% accuracy, 83.38% precision, 71.11% recall and 76.09% f1 measure. The accuracy of the K-Nearest Neighbor may be decreased, when we use high number of datasets and for complex datasets, and also in the case of defining parameter k. In such cases, KNN cannot predict targets with good accuracy.

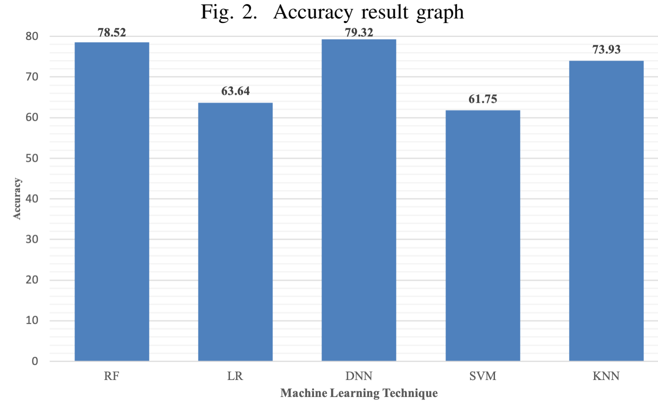
Unlike the classical and ensemble machine learning, A deep neural network takes the complex feature and process more intensive tasks by simultaneously executing many complex operation to predict the churn and non-churned customerof CBE in comparison with other traditional algorithms on the same dataset. As a rule of thumb, the DNN algorithm use number of hidden layers taking two-third of the total features [32]. So this study has used 7 hidden layers. Deep Neural Network has used parameters activation function is Relu, Number of neurons in each layer is 10, Batch size is 40, Number of hidden layers is 7, Epoch is 80, and optimizer is Adam. The accuracy of a Deep Neural Network decreases when the data is complex, and when the data become a small number of attributes because DNN depends on a lot of training data. The result obtained from the experiment shows that, a 79.32% accuracy registered with 85.08 precision, 78.19% recall and 80.79% F1 measure. Table II presents the detailed performance measure of the five machine learning algorithm in terms of accuracy, precision, recall and F1 score.

TABLE II
DETAILED EXPERIMENTAL RESULT RESULT

	Accuracy	Precision	Recall	F1 score
LR	63.64%	81.65%	56.97%	66.71%
RF	78.52%	84.85%	77.06%	80.10%
SVM	61.75%	81.36%	54.42%	64.88%
KNN	73.93%	83.38%	71.11%	76.09%
DNN	79.32%	85.08%	78.19%	80.79%

As shown in Table II, the results obtained from each experiment shows a maximum performance in a deep neural network of while the minimum is 63.64% using the logistic regression. Fig. 2, presents the performance comparison of the two classical, one ensemble and one deep learning techniques.

Fig. 2. Accuracy result graph



As shown in Fig.2 and Table II, results were obtained for five of the classifiers with an average accuracy of 79.32%, 78.52%, 73.93%, 63.64%, and 61.75% for DNN, RF, KNN, LR, and SVM respectively. The SVM's performance result is relatively poor when compared to the other experiment. This is due to the use of a linear classifier in the customer churn prediction, which is incapable of classifying non-linear separable parts. LR surpasses SVM in terms of performance by 1.87%. This is due to target overlaps in customer turnover. When compared to LR, KNN performs better since it relies on the nearest 505 elements to predict the category of customer turnover. Similarly, RF outperforms KNN due to its superior abilities in unbalanced data, less influence of outliers, and big data size.

DNN outperforms RF, KNN, LR, and SVM by 0.8%, 5.39%, 15.68%, and 17.57% respectively. Compared to results the DNN gives better results than other machine learning classifier for predicting customer churn for the CBE customer churn dataset. Thus, due to the fact, that when we have a large number of datasets and complex dataset, Deep Learning algorithms are more accurate than other classical machine learning algorithms.

V. Conclusion and Recommendation

In this paper, we have introduced a customer churn prediction model in CBE to retain the existing customer. A total of 204,161 customer instances were utilized in the paper's data collection from CBE using simple random sampling, with 21.8% (44,591) of the total datasets representing churn and 78.2% (159,570) non-churn. To make it easier for machine learning tasks, the data is further cleansed, pre-processed, balanced, and divided (into training and testing). In order to prepare the features for machine learning algorithms, this study also used feature engineering, efficient feature transformation, and a selection technique. Based on past work, five machine algorithms were selected. These are LR, RF, SVM, KNN, and DNN algorithms. Finally, the Deep Neural Network model accuracy has 79.32% and also it achieved best results in all measurements; the accuracy value of 78.52% for Random Forest algorithm comes in second place followed by KNN, LR, and SVM with 73.93%, 63.64%, and 61.75% accuracy, respectively. The result obtained from the experiment shows a promising that it is possible to churn the customer using selected attribute.

Furthermore, we are working towards developing a machine learning model by incorporating additional attribute for the customer churn prediction using different machine learning techniques.

References

[1] Hadden, John, Ashutosh Tiwari, Rajkumar Roy, and Dymitr Ruta. ”Computer assisted customer churn management: State-of-the-art and future trends.” Computers and Operations Research 34, no. 10 (2007): 2902-2917. : 10.1016/j.cor.2005.11.007.

[2] Kim, Moon-Koo, Myeong-Cheol Park, and Dong-Heon Jeong. ”The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services.” Telecommunications policy 28, no. 2 (2004): 145-159.

[3] Kim, Hee-Su, and Choong-Han Yoon. ”Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market.” Telecommunications policy 28, no. 9-10 (2004): 751-765.

[4] Umayaparvathi, V., and K. Iyakutti. ”Automated feature selection and churn prediction using deep learning models.” International Research Journal of Engineering and Technology (IRJET) 4, no. 3 (2017): 1846-1854.

[5] Ascarza, Eva, Scott A. Neslin, Oded Netzer, Zachery Anderson, Peter S. Fader, Sunil Gupta, Bruce GS Hardie et al. ”In pursuit of enhanced customer retention management: Review, key issues, and future directions.” Customer Needs and Solutions 5, no. 1 (2018): 65-81.

[6] BOUCHER, Philip Nicholas. ”Artificial intelligence: How does it work, why does it matter, and what can we do about it?.” (2020).

[7] Galbusera, Fabio, Gloria Casaroli, and Tito Bassani. ”Artificial intelligence and machine learning in spine research.” JOR spine 2, no. 1 (2019): e1044.

[8] Wang, Haifeng, Lei Wang, and Chunquan Liu. ”Employee competitive attitude and competitive behavior promote job-crafting and performance: A two-component dynamic model.” Frontiers in psychology (2018): 2223.

[9] Santacruz Nagua, Alfredo Fabian. ”Customer churn prediction by using support vector machine.” Bachelor's thesis, Universidad de Investigación de Tecnología Experimental Yachay, 2021.

[10] Lemmens, Aurélie, and Christophe Croux. ”Bagging and boosting classification trees to predict churn.” Journal of Marketing Research 43, no. 2 (2006): 276-286.

[11] Sharma, Anuj, Dr Panigrahi, and Prabin Kumar. ”A neural network based approach for predicting customer churn in cellular network services.” arXiv preprint arXiv:1309.3945 (2013).

[12] Bhambri, Vivek. ”Data mining as a tool to predict churn behavior of customers.” India: International Journal of Computer and Organization Trends–Volume2Issue3 (2012).

[13] Soeini, Reza Allahyari, and Keyvan Vahidy Rodpysh. ”Applying data mining to insurance customer churn management.” International Proceedings of Computer Science and Information Technology 30 (2012): 82-92.

[14] MESGANAW, YALELET. ”ASSESSMENT OF SERVICE QUALITY AND CUSTOMER SATISFACTION: THE CASE OF COMMERCIAL BANK OF ETHIOPIA, ADDIS ABABA BRANCH.” PhD diss., St. Mary's University, 2016.

[15] Worku, Gardachew. ”Electronic-banking in Ethiopia-practices, opportunities and challenges.” The Journal of Internet Banking and Commerce 15, no. 2 (1970): 1-8.

[16] Beniwal, Sunita, and Jitender Arora. ”Classification and feature selection techniques in data mining.” International journal of engineering research and technology (ijert) 1, no. 6 (2012): 1-6.

[17] Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman. Mining of massive data sets. Cambridge university press, 2020.

[18] Nie, Guangli, Wei Rowe, Lingling Zhang, Yingjie Tian, and Yong Shi. ”Credit card churn forecasting by logistic regression and decision tree.” Expert Systems with Applications 38, no. 12 (2011): 15273-15285.

[19] Chitra, K., and B. Subashini. ”Customer retention in banking sector using predictive data mining technique.” In ICIT 2011 The 5th International Conference on Information Technology. 2011.

[20] Oyeniyi, A. O., A. B. Adeyemo, A. O. Oyeniyi, and A. B. Adeyemo. ”Customer churn analysis in banking sector using data mining techniques.” Afr J Comput ICT 8, no. 3 (2015): 165-174.

[21] Lu, Ning, Hua Lin, Jie Lu, and Guangquan Zhang. ”A customer churn prediction model in telecom industry using boosting.” IEEE Transactions on Industrial Informatics 10, no. 2 (2012): 1659-1665.

[22] Xia, Guo-en, and Wei-dong Jin. ”Model of customer churn prediction on support vector machine.” Systems Engineering-Theory and Practice 28, no. 1 (2008): 71-77.

[23] Kaya, Erdem, Xiaowen Dong, Yoshihiko Suhara, Selim Balcisoy, and Burcin Bozkaya. ”Behavioral attributes and financial churn prediction.” EPJ Data Science 7, no. 1 (2018): 41.

[24] "KASSAHUN, APPLICATION OF DATA MINING TECHNIQUES TO PREDICT CUSTOMERS'CHURN AT COMMERCIAL BANK of ETHIOPIA," no. September, 2013.

[25] Attenberg, Josh, Prem Melville, Foster Provost, and Maytal Saar-Tsechansky. ”Selective data acquisition for machine learning.” Cost-sensitive machine learning (2011): 101.

[26] Kotsiantis, Sotiris B., Dimitris Kanellopoulos, and Panagiotis E. Pintelas. ”Data preprocessing for supervised leaning.” International journal of computer science 1, no. 2 (2006): 111-117.

[27] Brownlee, Jason. ”Data preparation for machine learning.” (2022).

[28] Hevner, Alan R., Salvatore T. March, Jinsoo Park, and Sudha Ram. ”Design science in information systems research.” MIS quarterly (2004): 75-105.

[29] Bruneo, Dario, Salvatore Distefano, Maurizio Giacobbe, Antonino Longo Minnolo, Francesco Longo, Giovanni Merlino, Davide Mulfari et al. ”An iot service ecosystem for smart cities: The smartme project.” Internet of Things 5 (2019): 12-33.

[30] Grattarola, Daniele, and Cesare Alippi. ”Graph neural networks in TensorFlow and keras with spektral [application notes].” IEEE Computational Intelligence Magazine 16, no. 1 (2021): 99-106.

[31] Hart, Peter E., David G. Stork, and Richard O. Duda. Pattern classification. Hoboken: Wiley, 2000.

[32] Heaton, Jeff. ”The number of hidden layers.” Heaton Research Inc (2008): 157-158..