2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS) | 2-4 September 2021 | Ernakulam

2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS) | 978-1-7281-7136-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/ACCESS51619.2021.9563347

LSTM Model to Predict Customer Churn in Banking Sector with SMOTE Data Preprocessing

Jesmi Latheef
Dept of Computer Science and Engineering
Rajiv Gandhi Institute of Engineering and Technology
Kottayam, Kerala, India
jesmilatheef@gmail.com

Vineetha S
Dept of Computer Science and Engineering
Rajiv Gandhi Institute of Engineering and Technology
Kottayam, Kerala, India
svineetha@rit.ac.in

***Abstract*—In any organization, the technique used to acquire new users is by Customer Relationship Management systems. In order to achieve more profitability with increase in customer retention is by maintaining a healthy association with them. Customer Churn is also known as Customer Loyalty or Retention. The inspiration behind churn forecast is to categorize and discover clients into churner & non-churner. A churned client implies there is a greater chance of the client is around to take off from the organization. A novel software can be utilized to discover the clients who will donate increased benefits for the organization. Moreover churn forecast can maintain a strategic distance from the misfortune of income by holding the existing clients. A few procedures are accessible for churn prediction with ensemble and hybrid models. This paper points to anticipate client Churn in banking sector with LSTM model and the data is preprocessed using SMOTE technique to overcome imbalanced information. The work is an extension to predicting customer loyalty in banking sector using Mixed Ensemble and Hybrid model. This paper proposes an accurate way to predict customer churn using LSTM model and the data is preprocessed using SMOTE technique. In this way the framework is more valuable for organizations to discover the clients with more chances to become churn. The results of the evaluation indicated that this is to be the case, the proposed systems for churn prediction performs with an accuracy of 88% and which is much better than the system without SMOTE technique.**

Keywords—Customer Relationship Management, Customer Churn, Synthetic Minority Oversampling Technique, LSTM

I. Introduction

In competitive markets, customers are the valuable factors. For business development the organizations primarily accept the pickups that acquired from the successful profitable clients. So CRM is nothing but managing the services to customers and it incorporates the concepts like client procurement, upkeep and fulfilment. As from the past works on this, it is evident that securing modern client makes an additional cost that of keeping up the old clients. Customer retention is a straight forward impediment for corporates and it can be explained as the customers are going to exit because of they were moved to some other better services by leaving the existing ones. By prior to know about who will become churn or non-churn, would provide an excellent perception with reference to the returns so as to maintain and develop their cycle of users. In order to make strong relationship between customer and the organization, they should able to keep the factors under their control and it will helps to forecast the behavior of each customer and which in turn minimize the churn rate. Besides, identifying and dividing customers into churn and non-churn will leads to a binary classification problem of forecasting. Fig.1 shows the customer churn cycle of any organization. These steps are followed to retain the existing customers and thereby reduce the expenses of acquiring new customers.

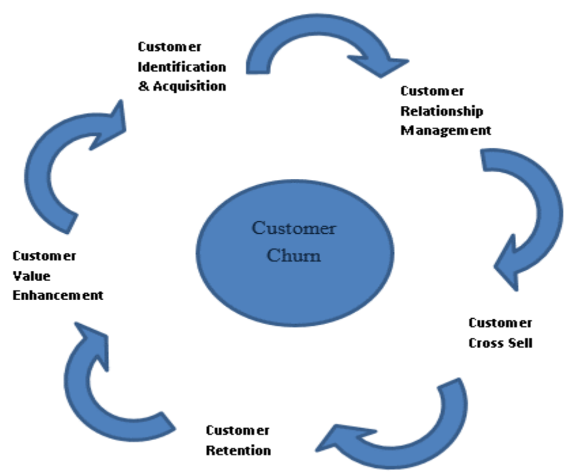

Fig. 1. Customer Churn Cycle

Thus the organization is able to maintain revenue and it leads to high profitability. Hence the customer churn forecasting framework can identify that which user have more propensity to takeoff the benefits by analyzing different factors of the input data. From the study, it is clear that the system can be implemented with various machine learning techniques like unsupervised, semi-supervised, and supervised. Deep learning models can too be each viable for this work [1], [2]. In this paper, LSTM model with SMOTE data pre-processing was utilized for a more precise forecast of client churn on bank data set. The work was a pure extension of the model that collaborates multi-layer perceptron to an ensemble and a model which was made by cumulating basic algorithms. SMOTE was used to overcome unequal distribution of class over the data set. Therefore, in this paper to forecast customer attrition LSTM method was used with SMOTE data pre-processing.

The upcoming sections of this paper is arranged as follows. In Section 2, a brief description of the works related to procedures and algorithms used in forecasting of clients who will leave. Section 3 describes about the steps in model creation, and Section 4 reveals the execution and performance analysis. Finally, conclusion is come up with Section 5.

II. Related Works

Different data analytical strategies have been broadly utilized for assessing the likelihood of client loyalty. To the greatest extent, the related works are discussed about various machine learning algorithms such as deep learning algorithms, a summary of some of them are outlined here,

A. Data Mining Strategies

So as to create compelling and exact customer-churn prediction model, numerous data extraction and pattern analyzing methods have been utilized. Data mining is utilized for covered up, substantial, and possibly valuable pattern and knowledge discovery. Although, it is all around observing already obscure connections between the information. Also knowledge extraction can be utilized for forecast which includes utilizing a few factors or fields within the database to foresee obscure or future values of other factors of interest. For that some pattern extracting and analyzing strategies include classification, clustering, regression, and so on [3]-[9].

B. Ensemble Methods

Machine learning outcomes can be increased by combining a few models. Compared to single model performance, this approach permits the generation of higher predictive performance. So as to reduce the fluctuations and noises, less sensitivity towards single data, or to gain better predictions, ensemble methods were used. Ensemble methods are act as a wrapper that combining a few machine learning techniques in to one and execute. [10].

C. Artificial Neural Network

Artificial Neural Networks are made from neuron architecture and it will improve the predictive precision and interpretability of churn detection in financial institutions by imitating human neurons. In [11] churn and non-churn clients were anticipated using ANN and distinctive performance measurements are utilized for finding the adequacy and viability of the framework. MLP is an architecture of the artificial neural network that that comprises of different layers where each layer is completely associated with the following layer in a feed-forward manner. Inputs or independent variables are represented by the first layer and outputs or target variables of the system are represented by the final layer respectively. Weights are used to indicate the connection between nodes in each layers. Complexity of the network can be determined by controlling the number of hidden layers in the system [12].

D. Hybrid Models

A hybrid model can be constructed using two or more fundamental models. In order to check the predictive performance, there are several attempts on churn prediction using this hybrid model was made. In [13], it uses a combination of bagging, boosting and LOLIMOT algorithms and is named as Ordered Weighted Averaging (OWA) technique. *Elham Jamalian et. al* predicts customers churn more accurately, using data fusion and feature extraction techniques [14]. Two algorithms, LOLIMOT and C5.0, were trained with different size of features then the outputs of the individual classifiers were combined with weighted voting.

III. System Architecture

In this section, the system architecture of proposed model to churn prediction using LSTM Model with SMOTE is discussed. Fig.2 indicates the complete flow of stages of the system. To make the system begin by inputting some bank churn data and then it can preprocess the data to improve the performance results. There are a number of steps are involved between the beginning and end of the system. Those steps are explained below

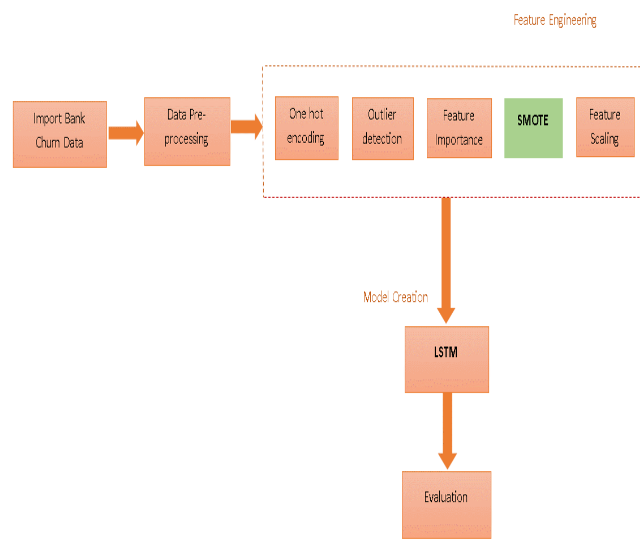

Fig 2. Overview of Different Stages

A. Removing Data

The first step is the preprocessing step. This step is to make the data well cleaned and properly arranged. This can be done through the discarding of fields with zero value, discarding of unrelated features and also discover numerical and graphical abstract of the data set. Recognize and discard those invalid attributes which may cause consequences to the procedure was performed by this step. For that find a zero variance was formed by unique value and thus the single valued attributes or properties could not be used as input for machine learning algorithms. It is important to remove those noise values and whose nearness will influence the ultimate result unfavorably.

B. Feature Engineering

It is the vital step in any kind of predictive problem, thereby it will extend the effectiveness of the proposed model. Based on the necessary conditions of the desired problem, the dataset must be prepared to make it adaptable with the machine learning algorithms that may going to use. One hot encoding creates a binary column for each category. Extreme deviations in any category can be detected and removed in outlier detection. It can be visualized using box plot diagrams. Next step was finding the important features in the data set, for that feature importance property of any tree based model can be used.

The most important step in feature engineering was SMOTE based data processing. SMOTE was used because of the unequal distribution of class in the dataset. That is, the quantity of majority class based data points is very large compared to the minority class. So this imbalanced data can lead to degrade the performance of the model. In this work synthetic samples are generated for the minority class (SMOTE). In the next stage, MinMaxScalar is used for scaling the dataset.

C. *LSTM-SMOTE model*

This is the modelling stage, thus the input data to the model was well suited and well arranged for attaining maximum performance gain. This paper constitute a Long Short Term Memory model to find the customer retention in banking sector. The model is trained with SMOTE processed data points and will give a better progress in terms of performance.

As in the previous work [15], by combining several models to form an ensemble and it will beat the results of single machine learning model. Along with that the provided data was trained against a hybrid model, which cumulates the output of specific models. Here, the deep neural network related processes are used and is specific to LSTM (Long Short Term Memory). This strategy gives the superior prescient execution compared to all the outfit models like ensemble and all other single models. Also in [16], a KNN-SMOTE-LSTM model is used to detect anomalies in wireless network data characterized by unbalanced data distribution. This technique can succeed in dealing with the imperfection of deranged data classification by conventional procedures. It applies the SMOTE algorithm to amplify the data to solve the overfitting caused by unbalanced data and then KNN-LSTM algorithms are used to remove the noised samples. This will viably move forward the execution and exactness of classification. SMOTE can also be used in credit risk evaluation models to improve the efficiency of the model. In [17], in order to remove the unbalancing of data due to time variation can be reduced before building the model. Ensemble model of deep learning algorithms are used to train the model.

LSTM is an artificial Recurrent Neural Network (RNN) architecture. All the other architecture has feed forward connection and LSTM has feedback connection. In the proposed system, the data set is of two dimension and it should be reshaped to form the proper input to the model. By this way, the model is created and data with synthetic samples were given as input.

D. *Evaluation of Models*

In order to measure the performance of models, some evaluation strategies are used. In case of problems with machines that trained using labelled data, such as classification and regression, efficiency of the system is evaluated based on results of the corresponding metrics used. Among them confusion matrix and ROC curve based evaluations are the most common methods. The curves generated by AUC-ROC will help us to visualize how well the model is performing. For binary classification problems, it will give an estimate about the performance of how good a classifier discriminate among classes. Also confusion matrix is a summarized table with actual and predicted values of a classification problem. By observing the table one can get adequate information about the execution of the system.

IV. Execution, Evaluation and Results

The framework is executed in Colab stage given by Google using Python3 technology. The expected system is made by choosing a dataset that ought to be justifiable and simple to be intelligible, that composes all the relevant details about customers of banking sector based on country wise. Also it should be an adaptable data set for churn forecasting of financial institution. So bringing in a bank churn demonstrating dataset from Kaggle with 14 features and 10,000 records. The system begins with make the data to be compatible to process. Within the pre-processing step, invalid values and unessential highlights were evacuated. Irrelevant features can be removed based on checking unique characters of the corresponding category. In this work, three unique characters were checked and if the feature has greater than 3 unique characters, it is not relevant for the model and it was removed. Here, to find the correlation of features, Correlation Matrix is used. After understanding the trends and distribution of data, the data was prepared for modelling, this was done by feature engineering.

In the upcoming step, imbalanced data, the numerical attributes that are diverse in terms of ranges were tended to. In order to overcome the challenges involved in handling imbalanced data, synthetic minority (SMOTE) samples of the data set was created. After completing the data preparation next is the modelling phase. In the proposed system, modelling includes creation of an RNN based LSTM model. In this previous work, a Mixed Ensemble model and a hybrid model was implemented. In order to understand the impact of SMOTE data processing in the proposed work, an LSTM model without SMOTE was implemented and compared the performance of them. Table I shows a comparison of each model accuracies that studied. From the table it is clear that the LSTM with SMOTE data processing has an unbeatable performance than others. The overall accuracy of the LSTM model with SMOTE data processing is 88%.

TABLE I. Comparison of Accuracy and AUC-ROC of different models with LSTM-SMOTE model

Model	***Accuracy***	***AUC-ROC***
Mixed Ensemble	86	0.87
Hybrid Model	87	0.87
LSTM	85	0.85
LSTM with SMOTE	**88**	**0.93**

Error matrix and Area under the curves were used to assess the system performance. Error matrix or Confusion matrix is an outline representation of expectation versus actual values of the prediction problem. Forecasting Rate or Accuracy of the model can be calculated based on the diagonal elements of the error matrix. The AUC - ROC curves are constructed based on two parameters called, True Positive Rate (TPR) and False Positive Rate (FPR). Fig. 3 shows the ROC curves of ordinary LSTM. Within the figure the zone beneath highlighted color bend appears the execution proficiency of the system. Fig. 4 illustrates the AUC-ROC curve of the LSTM model with SMOTE, it focuses out the degree of model execution.

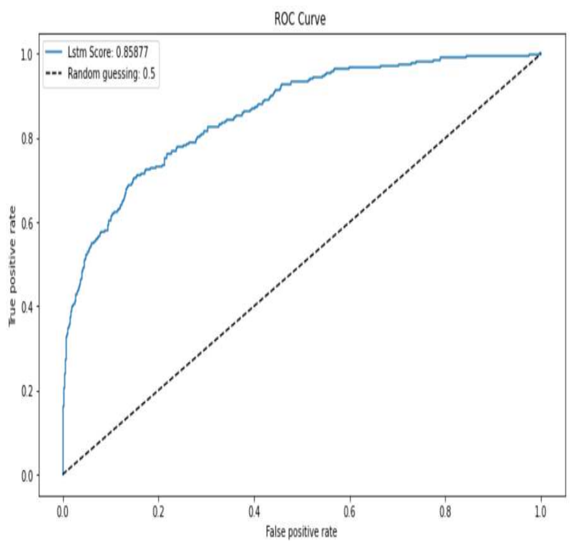

Fig. 3. ROC curve of LSTM model

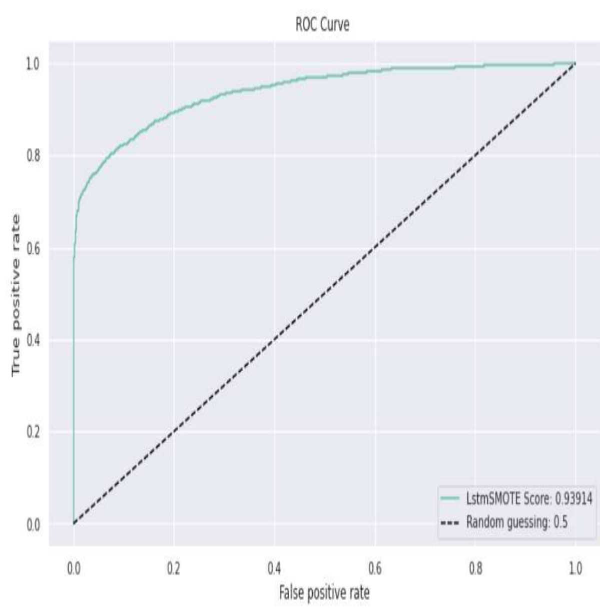

Fig. 4. ROC curve of LSTM with SMOTE

V. Conclusion and Future Work

In this system, the customer data of banking sector is used to predict whether the customer is going to leave the bank or not. For that LSTM model with SMOTE data pre-processing was used. In SMOTE technique, synthetic minority samples are generated for minor class of data. Thus it can overcome the issue of unequal distribution of data. Here, LSTM model is used based on RNN architecture. Deep learning techniques will give efficient performance results for these kind of classification problems. Subsequently the strategy of LSTM with SMOTE have more noteworthy execution than conventional LSTM models and all other models presented in early efforts. Also, the evaluation results obtained shows that the LSTM with SMOTE model outperforms other models in term of prediction accuracy. Hence, it is concluded that the LSTM model with SMOTE can perform in a way better than the standard models. In addition, this model performs more stable than the other models. Hence, an effective way to find customer retention and loyalty rate is by using LSTM model with SMOTE data processing technique.

As future, the work the forecasting of customers who will become churn may be extend to other range of algorithms such as Fuzzy, Genetic Algorithms, ANFIS, etc. Also innovations of experimental tuning can be introduced such as multiple objective optimization, explanatory variables selection, transformation process and also different evaluation criteria can be introduced to measure the performance of the model.

Acknowledgment

I hereby express my sincere thanks and gratitude to Computer Science and Engineering Department of Rajiv Gandhi Institute of Technology, Kottayam

References

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

[2] Afaq Alam Khan, Sanjay Jamwal, and M.M.Sepehri, "Applying data mining to customer churn prediction in an Internet Service Provider," International Journal of Computer Applications, vol. 9, No.7, 2010.

[3] P. K. Dalvi, S. K. Khandge, A. Deomore, A. Bankar and V. A. Kanade, " Analysis of customer churn prediction in telecom industry using decision trees and logistic regression, " Symposium on Colossal Data Analysis and Networking (CDAN), Indore, 2016.

[4] Kumar, A. Saran, and D. Chandrakala, " An Optimal Churn Prediction Model using Support Vector Machine with Adaboost," International Journal of Scientific Research in Computer Science, Engineering and Information Technology 2.1, pp. 225-230, 2017.

[5] Mahajan, Deepika, and Rakesh Gangwar, "Improved Customer Churn Behaviour by Using SVM", International Journal of Engineering and Technology 2395-0072, 2017.

[6] Kumar, Saran, S. Viswanandhne, and S. Balakrishnan, " Optimal Customer Churn Prediction System using Boosted Support Vector Machine", International Journal of Pure and Applied Mathematics 119.12 1217-1231, 2018.

[7] ApurvaSree, G., Ashika, S., Karthi, S., Sathesh, V., Shankar, M., Pamina, J. "Churn prediction in Telecom using classification algorithms", International Journal of Scientific Research and Engineering Development 5, pp. 19-28, 2019.

[8] Prajapati, Darshana, and Ratnesh Kumar Dubey. "Analysis of Customer Churn Prediction in Telecom Sector Using Random Forest."

[9] Jaisakthi, S. M., Gayathri, N., Uma, K., Vijayarajan, V, "Customer Churn Prediction Using Stochastic Gradient Boosting Technique", Journal of

Computational and Theoretical Nanoscience 15, no. 6-7 , 2410-2414, 2018.

[10] Jesmi Latheef, and Vinnetha S, “Exploring Data Visualization to Analyze and Predict Customer Churn in Banking Sector with Ensemble Learning”, IJIRASE, volume 4,2021,

[11] Amuda, Kamorudeen A., and Adesesan B. Adeyemo. ”Customers Churn Prediction in Financial Institution Using Artificial Neural Network.” arXiv preprint arXiv: 1912.11346, 2019.

[12] Khan, Yasser, et al.”Customers Churn Prediction using Artificial Neural Networks (ANN) in Telecom Industry.” Editorial Preface from the Desk of Managing Editor 10.9, 2019.

[13] Basiri, Javad, et al, "A hybrid approach to predict churn." , IEEE Asia-Pacific Services Computing Conference, 2010.

[14] Elham Jamalian et.al, “A Hybrid Data Mining method for Customer Churn Prediction”, Engineering, Technology & Applied Science Research, Vol. 8, No. 3, 2018, 2991-2997

[15] Jesmi Latheef, Vineetha S, “Predicting Customer Loyalty in Banking Sector with Mixed Ensemble Model and Hybrid Model”, Springer, 4th International Conference On Smart Computing And Informatics, 2020.

[16] Bao, F., Wu, Y., Li, Z., Li, Y., Liu, L., & Chen, G, “Effect Improved for High-Dimensional and Unbalanced Data Anomaly Detection Model Based on KNN-SMOTE-LSTM.”, 2020.

[17] Shen, F., Zhao, X., Kou, G., & Alsaadi, F. E, “A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique”, Applied Soft Computing, *98*, 106852, 2021