International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 02, 2021

Improved Churn Prediction Model In Banking Industry And Comparison Of Deep Learning Algorithms

Mr. M. John Britto[1], Dr. R. Gobinath[2]

[1]*Research Scholar, Department of Computer Science, VISTAS, Chennai.*
[2]*Associate Professor, Department of Computer Science, VISTAS, Chennai.*

Email: [1]britto.chennai@gmail.com, [2]drgobinathramar.scs@velsuniv.ac.in

Abstract: Increasing the revenue and profitability is the top most priority of a business. One of the major factors affecting the profit of a business is Customer Churn. Early prediction of customer churn and customer retention plays a vital issue in Customer relationship management. Retaining a customer is cost effective than attracting a new customer. This paper demonstrates a frame work for predicting customer churn in banking industry using the transactional data. It also compares with various other models. It uses the Behavioral aspects of the customer through the transactions made by them. It has been implemented through attention based Hybrid GRU BiLSTM model.

Key words: Customer Churn, Imbalanced Data, GRU, BiLSTM.

1. INTRODUCTION

Customer churn is a critical issue faced by almost every industry. Retail banking also suffers from customer churn. Around 2 % to 10% of customers leave the bank without any prior communication. It reflects in profit of an organization. Since customer retention is comparatively less expensive than acquiring new customers to any business. Therefore, accurate and early prediction of customer churn is critical in minimizing the cost of a bank's overall retention marketing strategy.

The deep Learning models performs better in prediction mechanism. Some techniques like LSTM which does not suffer from vanishing gradient problem are used for time series data prediction(Venkatesh and Jeyakarthic, 2020). The BiLSTM Technique allows the network to handle the sequence time step information in both forward and backward direction. In spite of that mechanism it needs attention for performing better churn prediction. It is achieved by attention layers based on the recent activity trends from RFM features.

Customers' behavioral patterns from their transactions can predict better results than using the demographic values of a customer(Kaya *et al.*, 2018). Instead of using customers transaction details which are time series data to cluster the individuals, these values are created as features and are passed to the model. For each customer Recency, frequency, and monetary features are extracted and allow the model to learn and identify the patterns.

Here, this framework follows data cleaning and reduction, Feature extraction, Handling imbalanced dataset, and classification.

2. RELATED WORKS

The experiments showed that neural networks typically outperform logistic regression and decision trees in churn prediction.(Belém, 2018). For Bank customer churn prediction it is necessary to handle transaction of the customer. It is time series based sequence data(hegde and Mundada, 2019). It needs methods enhanced methods for data reduction since millions of records involved in transaction.

A Hybrid classification approach was implemented through RNN and LSTM for churn prediction. Data reduction and data imbalance was not better focused in this model. Effective preprocessing would fetch better recall value for the imbalanced dataset(Simion-Constantinescu *et al.*, 2018)

High level of irrelevant data in any dataset will surely affect the classifiers that would result in bad or wrong prediction. It might also consume more computational time and storage space. It was tested on different datasets and results shows that in many cases noise removal improves classification performance (Ougiaroglou and Evangelidis, 2016).
A Dynamic Classification for optimize customer churn prediction was implemented with under sampling and SMOTE to handle imbalance dataset. But time series based sequence prediction algorithm would fetch better results (Leung and Chung, 2020).

Deep and shallow model for Insurance churn prediction was implemented using generalized linear model for shallow part and feed forward neural network for deep model. It has used demographic data alone. Time series data was ignore. Implementing with time series prediction could be focused for better results (Zhang *et al.*, 2017).

A hybrid BiGRU and BiLSTM model was developed to handle time series data. A sigmoid function was used for final classification. It handles high dimensional time sequence data and BiGRU was used for its fast and easier computation (Munawar *et al.*, 2021).

The conventional method utilizing an LSTM-AE was employed sequence Auto encoder for time series data set. It had used LSTM cells for execution of the encoder and decoder for learning temporal dependencies from one sequence to another from electricity consumption data set. It was developed for predicting electricity consumption.(Khan *et al.*, 2020)

3. DATA SET DESCRIPTION

The Czech Financial dataset has 4500 Customer data over 14 topics such as account balance, credit card information, demographic data, Full date and time of transaction and so on. It is dated from January 2015 to December 2018 around 1 million transactions.

Unlike other Datasets, the target variable is not present and so it is computed based on time-stamp differences. Customers with last three month of inactivity or with no balance amount are labeled as churned. The transactional data such as ATM transactions, withdrawal, deposit, internet transactions, Time are present in the dataset. The behavioral features are extracted from the dataset for last one year of transactions from that dataset for calculating recent month trend from each bin.

4. PROPOSED SYSTEM MODEL

The proposed solution for Bank Customer Churn Prediction typically consists of 5 main steps.

• In step (1), the data is preprocessed; missing values and outliers are filled and removed, respectively. Usually, missing values and irrelevant values degrade the model's performance due to ambiguous data. Here the credit transactions are taken into account for extracting customers' behavioral pattern. The debit transactions are ignored. Data reduction is done through reduction through exclusively homogeneous clustering(John Britto and Gobinath, 2020). It is a down sampling technique for imbalance dataset. It reduces exclusively the majority class that is non churn data.

• In step (2), the cleaned and reduced data is then passed to the next step in which SMOTE technique is implemented to handle the imbalance dataset. In real world, Churned customers' samples are rare. Training of the model on such imbalanced dataset leads to biasness towards a majority class; therefore, data balancing is a necessary requirement.

In step (3), based on the time series data the behavioral features are computed from the customer's transaction dataset.

•In step (4), the stratified available data is then passed to the next phase for classification purpose.

• In step (5), a hybrid Attention based GRU and bi-directional LSTM model is developed and time series data is passed.

The following figure 1 depicts the proposed framework

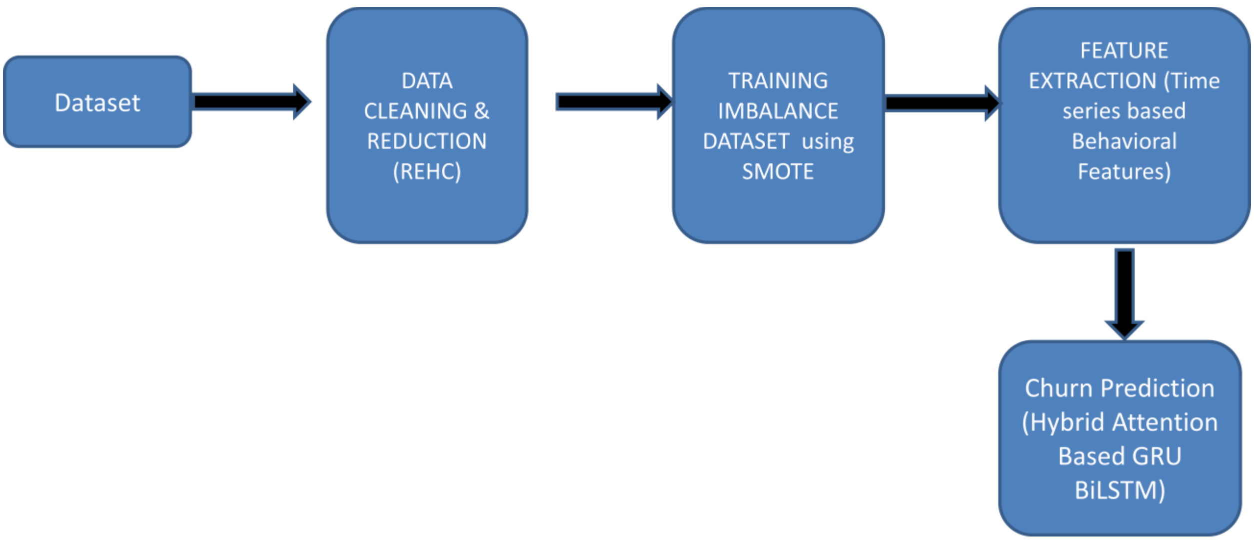

Figure 1 Proposed Bank Customer Churn Prediction Model

4.1 Data Preprocessing

Data preprocessing technique is used to transform the raw data into usable format. It involves data cleaning and data reduction. Here Reduction through exclusively homogeneous clustering technique is implemented(John Britto and Gobinath, 2020). REHC executes k-means clustering using initial means and builds n clusters. It repeats the routine exclusively for majority homogeneous clusters until not churned customer data are clustered.

The following tables 1 and 2 depicts the data set before and after reduction respectively

Table 1 Before REHC reduction

account_id	trans_id	type	operation	balance	bank	account	Year	month	k_symbol	fulldate withtime	fulltime	Churn Status
A00000001	T00000001	Credit	Credit in Cash	1000		NA	2015	3		2015-03-24T10:21:45	10:21:45	0
A00000001	T00000005	Credit	Collection from Another Bank	4679	JPMorgan Chase	41403269	2015	4		2015-04-13T08:54:57	8:54:57	0
A00000001	T00000199	Credit	Credit in Cash	17279		NA	2015	4		2015-04-23T10:54:46	10:54:46	0
A00000001	T03530438	Credit		17298.2		NA	2015	4	Interest Credited	2015-04-30T11:01:59	11:01:59	0
A00000001	T00000006	Credit	Collection from Another Bank	20977.2	JPMorgan Chase	41403269	2015	5		2015-05-13T10:33:00	10:33:00	0
A00000001	T00000200	Credit	Credit in Cash	23077.2		NA	2015	5		2015-05-23T15:35:53	15:35:53	1
A00000001	T03530439	Credi		23156.2		NA	2015	5	Interest	2015-05-	11:28:34	1

International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 02, 2021

		t							Credited	31T11:28:34		
A00000001	T00000007	Credit	Collection from Another Bank	26835.2	JPMorgan Chase	41403269	2015	6		2015-06-13T13:11:12	13:11:12	1
A00000001	T00000201	Debit	Cash Withdrawal	26635.2		NA	2015	6	Interest Credited	2015-06-22T08:46:02	8:46:02	1

Table 2 Data set after reduction using REHC

account_id	trans_id	operation	type	amount	balance	year	month	day	fulldate	fulltime	fulldatewithtime	Churn Status
A00000024	T00006052	Cash Withdrawal	Debit	1320	13531.7	2018	1	14	1/14/2018	10:46:47	2018-01-14T10:46:47	0
A00000024	T00006053	Cash Withdrawal	Debit	1440	22713.7	2018	1	6	6/1/2018	10:27:24	2018-01-06T10:27:24	0
A00000024	T00006067	Cash Withdrawal	Debit	2760	28248.4	2018	3	6	6/3/2018	14:18:53	2018-03-06T14:18:53	0
A00000024	T00006118	Cash Withdrawal	Debit	21500	19729.2	2018	6	28	6/28/2018	13:31:02	2018-06-28T13:31:02	0
A00000024	T00006119	Cash Withdrawal	Debit	26800	12404.5	2018	10	26	10/26/2018	12:38:13	2018-10-26T12:38:13	0
A00000033	T00009370	Credit Card Withdrawal	Debit	2200	111976	2018	12	17	12/17/2018	11:39:28	2018-12-17T11:39:28	0
A00000033	T00009401	Credit Card Withdrawal	Debit	1700	93657.3	2018	2	25	2/25/2018	13:36:14	2018-02-25T13:36:14	0
A00000033	T00009402	Credit Card Withd	Debit	3900	78717.9	2018	3	4	4/3/2018	14:53:00	2018-03-04T14:53:00	0

		rawal										
A0000 0033	T0000 9529	Credit Card Withd rawal	De bit	800 0	756 60.3	20 18	2	1	1/2/20 18	14:4 1:03	2018-02-01T14:4 1:03	0

4.2 Balancing Dataset using SMOTE

The balanced dataset is essential to train ML and deep learning models to prevent biasness towards majority class, which leads to misclassification. Here, we implemented SMOTE technique to handle imbalance data which synthesis the data samples and oversamples the minority class. SMOTE has proved successful in variety of applications(Fernández *et al.*, 2018). The following figure depicts the data imbalance.

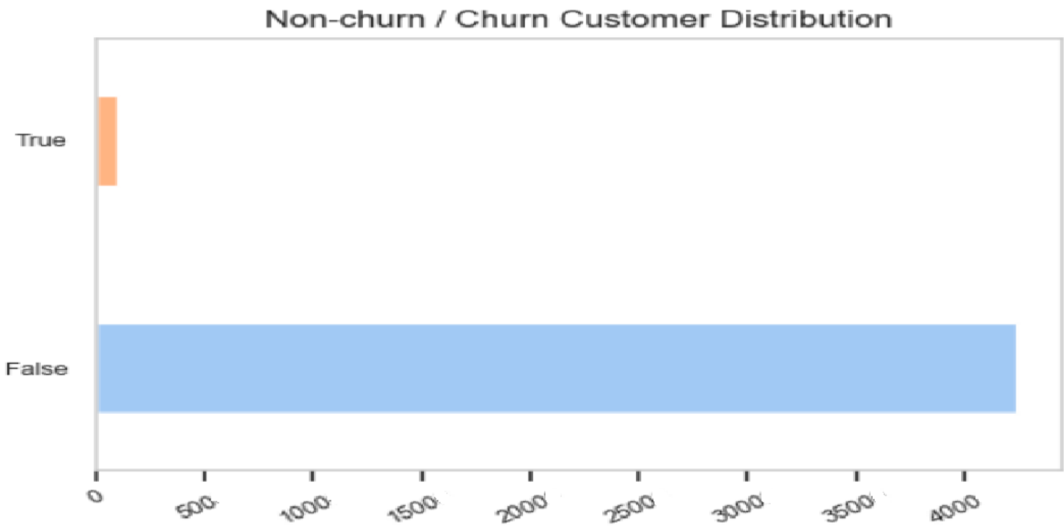

Figure 2 Distribution of Churned and not churned customers

4.3 Time series based Features

The three key features used to define and analyze the customer behavior through their transactions are Recency, frequency, and monetary value (RFM). Recency is the time interval since the last transaction made; frequency is the number of transactions made in a specified time window and monetary is the amount spent during a specified time. These sequence based time series data are computed quarter wise (3 months) for prediction.

4.4 Hybrid Attention based GRU and bi-directional LSTM

The Hybrid Attention based GRU BiLSTM is proposed for early banking churn prediction. Recent activity trend for each bin is used to attain new weights in attention layer. The behavioral features extracted from the transaction dataset have been given as input. Influencing recent activity among RFM activity is calculated from each bin separately. The Hybrid Attention based GRU BiLSTM is depicted in figure 3

The attention weights are transferred into the network at right part and produce a corresponding loss value based on the predicted error(Li *et al.*, 2019). The churning customer

International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 02, 2021

of a bank is not seasonal or periodic. It needs special attention based on recent activity trends of a customer. The recent month activity trend from each bin is computed and is given as attention weight.

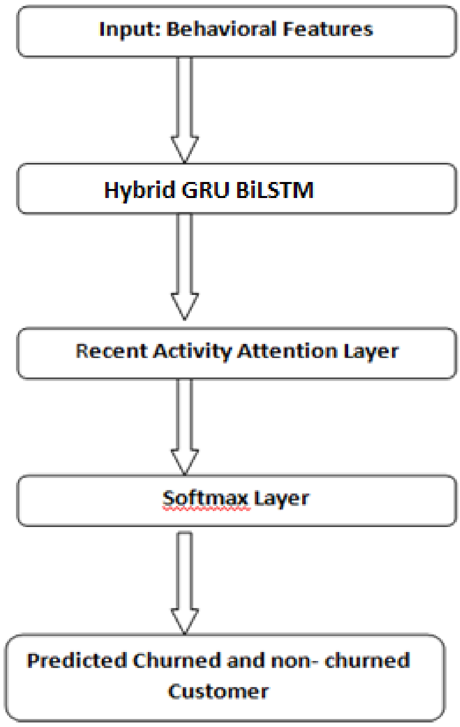

Figure 3 Hybrid attention based GRU BiLSTM

5. PERFORMANCE EVALUATION

The performance of the Hybrid attention based GRU BiLSTM model is evaluated using Precision, Recall, F1-score and PR – AUC.

5.1 Precision and Recall

Precision is the evaluation metric preferably used in the case of imbalanced data. Precision measures what proportion of positive class predictions are actually positive and is calculated as:

Precision = TP / (TP+FP)

Recall is known as the true positive rate (TPR, also called sensitivity), and it represents the proportion of the actual positive class which is predicted correctly. Recall is calculated as:
Recall = TP / (TP+FN)

5.2 F1- Score and PR-AUC

F1-score gives the weighted average value of both recall and precision.

F1−score = 2×(Precision×Recall) / (Precision+Recall)

Table 1 reveals the evaluation results.

Table 3 Evaluation results

	Precision	Recall	F1-Score	PR-AUC
RNN	98.29	90.87	17.02	13.69
LSTM	78.16	97.34	12.27	37.58
BiLSTM	84.32	95.56	14.4	34.11
Attention based GRU BILSTM	83.74	**99.08**	**13.21**	23.21

Area under Precision - Recall Curve: AU ROC is used in the case of imbalanced data set where positive class (like churn in our case) is rare, an alternative measure, which is the precision recall curve is used. Figure 3 shows the area under ROC curve.

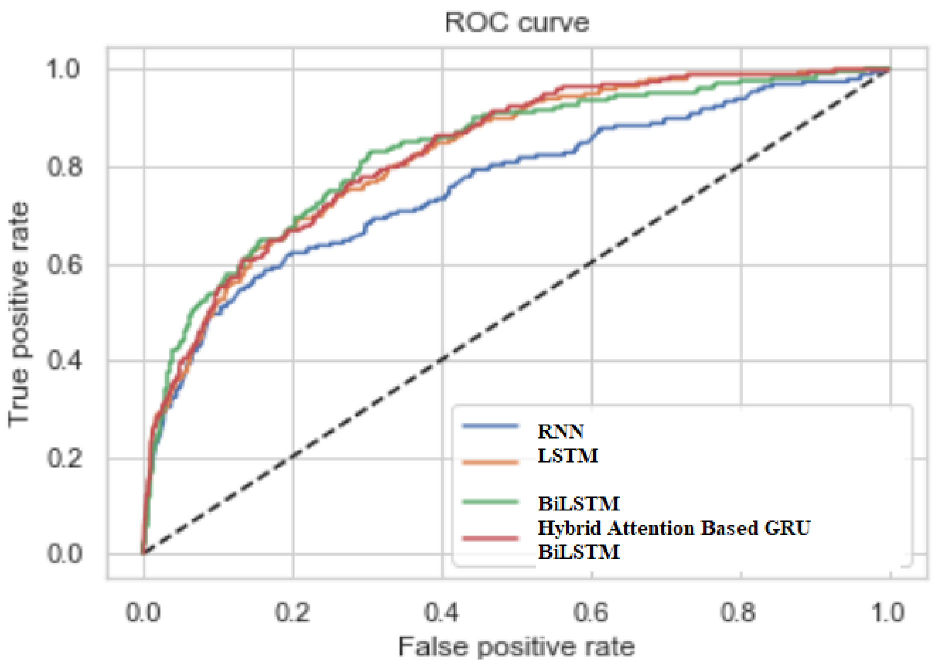

Figure 4 Area under the ROC Curve

6. CONCLUSION

In this paper, improved churn prediction model for banking industry is proposed and compared with various deep learning models. It takes time series customer transaction dataset for prediction. Down sampling and up sampling has been applied for raw data set using REHC and SMOTE respectively. Behavioral features has been computed and fed as input for the hybrid GRU BILSTM model. An Attention has computed and given to the model based on recent activity for each bin. Since, it is imbalanced data set the recall value is used for evaluation metrics. It predicts the churning customers (True Negative) and thus recall value is used. It predicts true negative (churning customers) with recall value of 99.08. Thus, a improved model for bank customer churn has been proposed and evaluation results shows better performance when compared with other deep learning models.

7. REFERENCES

[1] Belém, N. (2018) 'Gauging and Foreseeing Customer Churn in the Banking Industry : A Neural Network Approach'.

[2] Fernández, A. *et al.* (2018) 'SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary', *Journal of Artificial Intelligence Research*, 61, pp. 863–905. doi: 10.1613/jair.1.11192.

[3] hegde, S. and Mundada, M. R. (2019) 'Enhanced Deep Feed Forward Neural Network Model for the Customer Attrition Analysis in Banking Sector', *International Journal of Intelligent Systems and Applications*, 11(7), pp. 10–19. doi: 10.5815/ijisa.2019.07.02.

[4] John Britto, M. and Gobinath, R. (2020) 'REHC: Reduction through exclusive homogeneous clusters for imbalance dataset', *Materials Today: Proceedings*. Elsevier Ltd, (xxxx), pp. 1–6. doi: 10.1016/j.matpr.2020.11.093.

[5] Kaya, E. *et al.* (2018) 'Behavioral attributes and financial churn prediction', *EPJ Data Science*. The Author(s), 7(1). doi: 10.1140/epjds/s13688-018-0165-5.

[6] Khan, Z. A. *et al.* (2020) 'Towards efficient electricity forecasting in residential and commercial buildings: A novel hybrid CNN with a LSTM-AE based framework', *Sensors (Switzerland)*, 20(5), pp. 1–16. doi: 10.3390/s20051399.

[7] Leung, H. C. and Chung, W. (2020) 'A dynamic classification approach to churn prediction in banking industry', *26th Americas Conference on Information Systems, AMCIS 2020*, pp. 0–5.

[8] Li, Y. *et al.* (2019) 'EA-LSTM: Evolutionary attention-based LSTM for time series prediction', *Knowledge-Based Systems*. Elsevier B.V., 181, p. 104785. doi: 10.1016/j.knosys.2019.05.028.

[9] Munawar, S. *et al.* (2021) 'Electricity Theft Detection in Smart Meters using a Hybrid Bi-directional GRU Bi-directional LSTM Model', (April).

[10] Ougiaroglou, S. and Evangelidis, G. (2016) 'Efficient editing and data abstraction by finding homogeneous clusters', *Annals of Mathematics and Artificial Intelligence*. Annals of Mathematics and Artificial Intelligence, 76(3–4), pp. 327–349. doi: 10.1007/s10472-015-9472-8.

[11] Simion-Constantinescu, A. *et al.* (2018) 'Deep Neural Pipeline for Churn Prediction', *Proceedings - 17th RoEduNet IEEE International Conference: Networking in Education and Research, RoEduNet 2018*. doi: 10.1109/ROEDUNET.2018.8514153.

[12] Venkatesh, S. and Jeyakarthic, M. (2020) 'Adagrad Optimizer with Elephant Herding Optimization based Hyper Parameter Tuned Bidirectional LSTM for Customer Churn Prediction in IoT Enabled Cloud Environment', *Webology*, 17(2), pp. 631–651. doi: 10.14704/WEB/V17I2/WEB17057.

[13] Zhang, R. *et al.* (2017) 'Deep and Shallow Model for Insurance Churn Prediction Service', *Proceedings - 2017 IEEE 14th International Conference on Services Computing, SCC 2017*, pp. 346–353. doi: 10.1109/SCC.2017.51.