

PREDICTION OF CUSTOMER CHURN IN TELECOM INDUSTRY USING DEEP LEARNING TECHNIQUES

¹S.Kavitha, ²R.Seetha, ³S.Sathyavathi

¹Assistant Professor II, Department of Information Technology, Kumaraguru College of Technology, Coimbatore, India,

² PG Student, M.Tech (Data Science), Kumaraguru College of Technology, Coimbatore, India,

³ Assistant Professor, SRG, Department of Information Technology, Kumaraguru College of Technology, Coimbatore, India.

ABSTRACT: Churn is the process, where a customer moves away or breaks off a subscription with his provider. Customer churn prediction performs a primary role in telecommunication industry due to the smart phone dominated era. It is very necessary to interpret and forecast the behaviour of customers. Achieving the new customer is more costly than engaging the old one. In previous study, machine learning techniques were used to forecast the client's churn which comparatively produced less accuracy. The latest technologies in neural networks are CNN (Convolutional Neural Network) and RNN (Recurrent Neural Network) that commit the result to the problem by predicting the customer.

IndexTerms - churn prediction, telecommunication, CNN, RNN.

I. INTRODUCTION

Switching over of client from one service to another competitive service is denoted as churn. Customers are crucial assets for any company's product and services. The main attempt is towards examining and inspecting why purchaser is quitting a particular product or service and choosing a competitor. The rate of churn is calculated using percentage of clients who leaves a product or service in a particular span of time.

In current scenario, the application of churn occurs in many sectors and expert services such as retail, investment, telecommunication and healthcare which depend on churn prediction. Among these industries telecommunication is one of the major applications of churn since this is the smartphone dominated era. In telecom sector churn is represented as count of purchasers who withdraw the existing subscription and shift over to some other network that suits their plans or method of subscribers either post paid or prepaid.

Prediction of customer churn is a broadly named anomaly to maintain its commodity customers. It is widely known that the rate of attaining unique customers is considerably higher than keeping the old ones. Analysis and prediction of churn is an important task, keeping an industry aspect in mind. Customer churn management system is divided as churning and confinement of customers. Churn process management is a branch of customer relationship management.

Deep learning is a division of machine learning algorithm which consist of numerous layers to continuously get higher level features from raw input. Deep learning algorithm such as convolutional neural network and recurrent neural network has been proposed for predicting the churn customers. Existing literatures evaluate variant aspect such as churn rate, performance of prediction and client retention using Machine learning algorithm. Analysts have suggested different results for maintaining clients by applying different techniques.

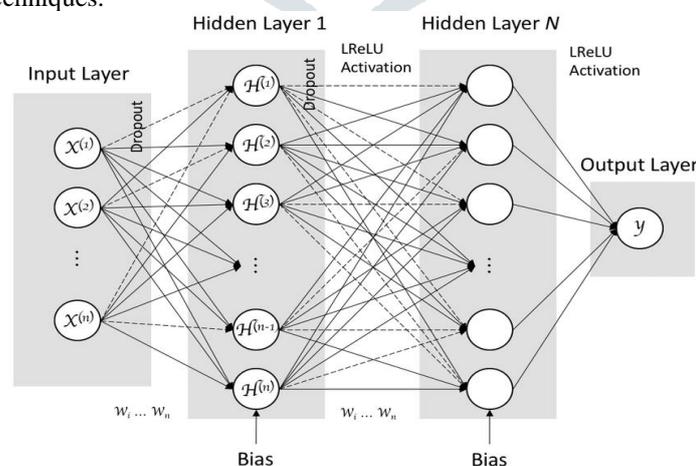


Figure 1 Deep Learning Architecture

The remaining part of the paper is formulated as follows: part 2 describes literature survey. The part 3 describes workflow of the system. Then the part 4 describes the implementation details. The part 5 describes the discussion and result. Finally, part 6 describes the conclusion.

II. LITERATURE SURVEY

[1] **IRFAN ULLAH**. et al.,(2019)in their research paper investigated churn prediction in present market of telecom industry. CRM is a significant issue for churn prediction to retain the customer and provide offers or services to maintain that group of customers. In [1] churn prediction performed a better by using the machine-learning algorithm such as Random Forest, which produced the accuracy of 0.88. Clustering techniques has been done by identifying the main factors from the dataset, which are used in predicting the churn customers.

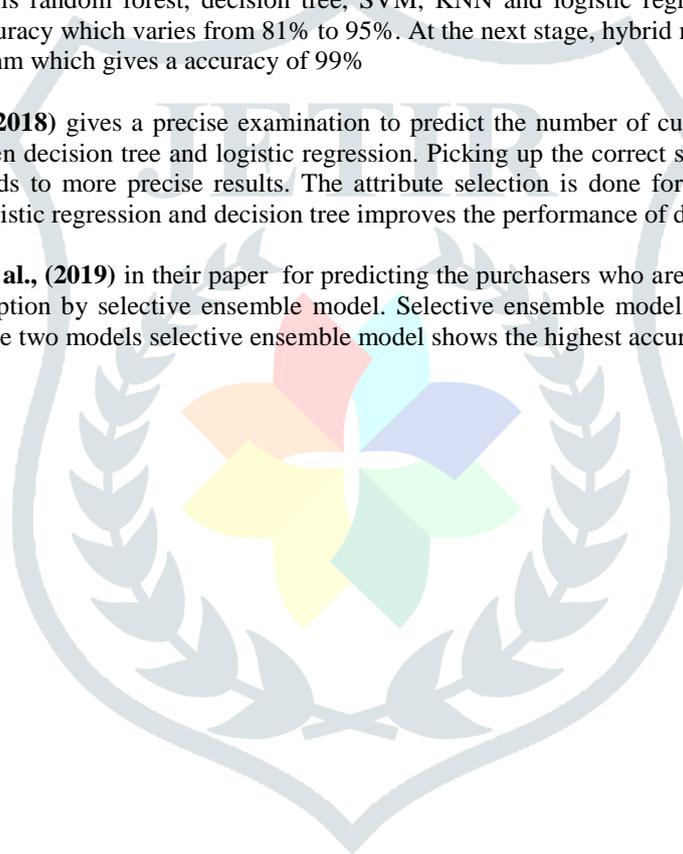
[3] **Abhishek Gaur (2018)** in his paper suggested that the telecom industry must know the reason for churn to retain the customer which can be extracted from telecom data. Four machine learning techniques has been used to analyse the customer churn prediction such as Logistic regression, SVM, Random forest and Gradient boosted tree. Among the four algorithms logistic regression gives a better accuracy 0.82 and SVM is underperforming model which gives 0.79.

[4] **Ammar A. Q. Ahmed (2018)** in his paper describes how to handle the customers by analysing and forecasting the churn in telecom sector. [4] Presents a hybridized firefly algorithm for churn prediction. A correlation has been done between the normal firefly algorithm and hybridized firefly algorithm in which hybridized firefly algorithm exceeds the normal firefly algorithm. Analysis was done on the basis of ROC, PR, F-Measure, accuracy and time.

[5] **Fahd Idrissi Khamlichi. et al., (2019)** Proposed a simulation method on public dataset using the most suitable classification method which is random forest, decision tree, SVM, KNN and logistic regression. First by evaluating each single algorithm gives an accuracy which varies from 81% to 95%. At the next stage, hybrid model is made by merging of two algorithms to a single algorithm which gives a accuracy of 99%

[6] **Atul Sunil Choudhari (2018)** gives a precise examination to predict the number of customers churning out the service using the interrelation between decision tree and logistic regression. Picking up the correct set of properties and settling edge qualities correctly which leads to more precise results. The attribute selection is done for the raise in performance of the decision tree. Both hybrid logistic regression and decision tree improves the performance of decision tree.

[7] **Ahmad Hammoudeh. et al., (2019)** in their paper for predicting the purchasers who are leaving the existing subscription and move to another subscription by selective ensemble model. Selective ensemble model is compared with the averaging ensemble model. Among these two models selective ensemble model shows the highest accuracy.



III. SYSTEM MODEL

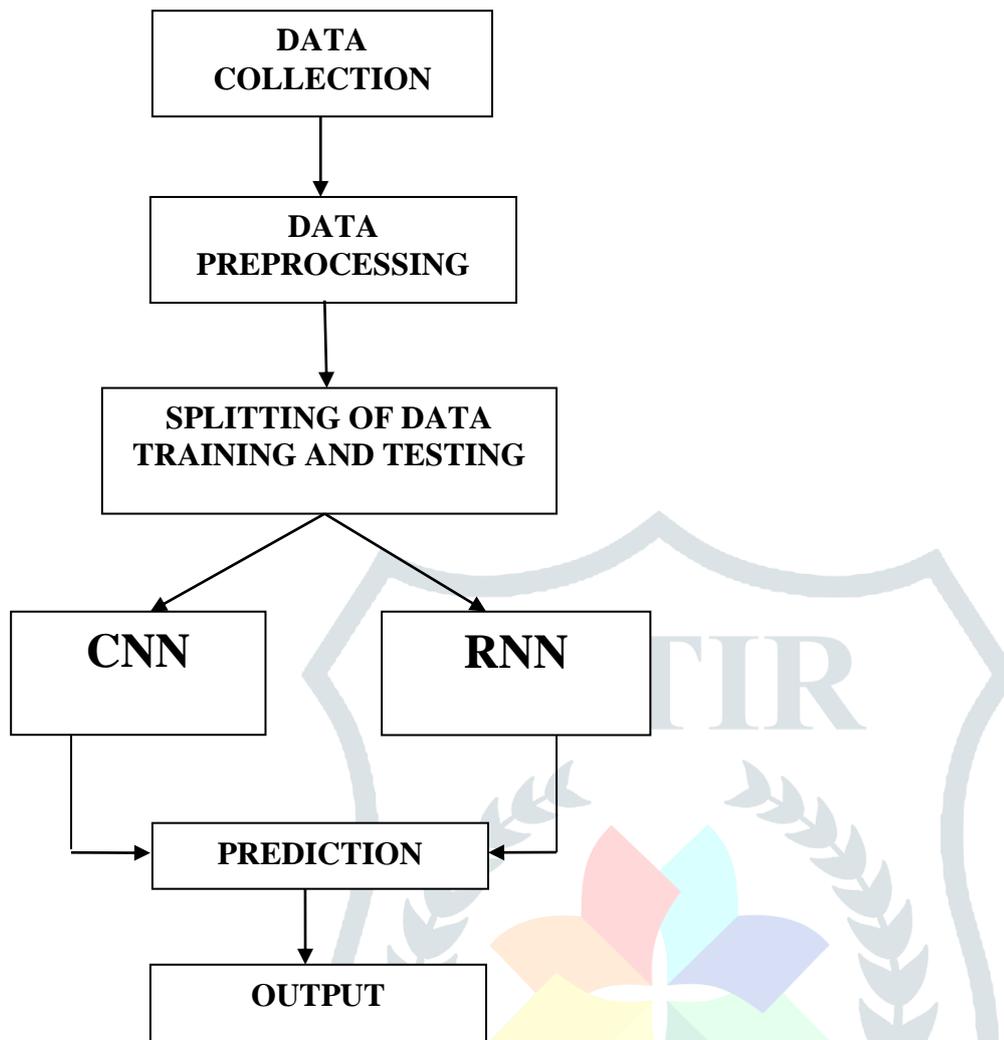


Figure 2: work flow diagram

The above diagram represents the work flow process of proposed system. The first step involves assortment of data. The collected data must be preprocessed using the python libraries. Then the data is splitted into training and testing. Deep learning techniques such as CNN and RNN are used for training the machine. Finally, with the help of testing data the comparison of two algorithms will be made to find which algorithm is better for prediction.

IV. IMPLEMENTATION DETAILS

The telecom dataset are collected for predicting the customers churn. Then preprocessing step is done by removing the missing values. The data are splitted into training and testing data. Two algorithms have been used in the proposed work for predicting the customers are CNN and RNN. Finally, the comparison of two algorithms will be made to find which algorithm is better for prediction.

a. DATA COLLECTION

The telecom customers churn prediction data was collected from kaggle website. The dataset comprises of 7043 records and 21 attributes. It is IBM sample data set which includes the client who quit in the previous month which describes the churn column. The attributes include customer id, gender, payment method, monthly charges, churn, etc

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Customer	gender	SeniorCiti	Partner	Dependent	tenure	PhoneSeri	MultipleLine	InternetSeri	OnlineSec	OnlineBac	DeviceProte	TechSupport	Streaming	Streaming	Contract	Pagerless	PaymentMethi
2	7590-VHVEG	Female	0	Yes	No	1	No	No phone si	DSL	No	Yes	No	No	No	No	Month-to-rc	Yes	Electronic che
3	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes	No	No	No	One year	No	Mailed check
4	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No	No	No	No	Month-to-rc	Yes	Mailed check
5	7795-CFOCW	Male	0	No	No	45	No	No phone si	DSL	Yes	No	Yes	Yes	No	No	One year	No	Bank transfer (
6	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	No	No	No	Month-to-rc	Yes	Electronic che
7	9305-CDSVC	Female	0	No	No	8	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	Month-to-rc	Yes	Electronic che
8	1452-KIDVK	Male	0	No	Yes	22	Yes	Yes	Fiber optic	No	Yes	No	No	Yes	No	Month-to-rc	Yes	Credit card (au
9	6713-OKOMI	Female	0	No	No	10	No	No phone si	DSL	Yes	No	No	No	No	No	Month-to-rc	No	Mailed check
10	7892-POOKF	Female	0	Yes	No	28	Yes	Yes	Fiber optic	No	No	Yes	Yes	Yes	Yes	Month-to-rc	Yes	Electronic che
11	6388-TABSU	Male	0	No	Yes	62	Yes	No	DSL	Yes	Yes	No	No	No	No	One year	No	Bank transfer (
12	9763-GRSVC	Male	0	Yes	Yes	13	Yes	No	DSL	Yes	No	No	No	No	No	Month-to-rc	Yes	Mailed check

Figure 3: Dataset

b. PREPROCESSING

After collecting the data preprocessing step is done by checking the missing values. The missing values should be dropped out. This dataset does not have any missing values. The next step is to convert the string values into the integer values. The output column churn has to be dropped for predicting whether the customer has been churned out or not.

```

Customer gender SeniorCitizen ... MonthlyCharges TotalCharges Churn
0 7590-VHVEG Female 0 ... 29.85 29.85 No
1 5575-GNVDE Male 0 ... 56.95 1889.50 No
2 3668-QPYBK Male 0 ... 53.85 108.15 Yes
3 7795-CFOCW Male 0 ... 42.30 1840.75 No
4 9237-HQITU Female 0 ... 70.70 151.65 Yes
...
7038 6840-RESVB Male 0 ... 84.80 1990.50 No
7039 2234-XADUH Female 0 ... 103.20 7362.90 No
7040 4801-JZAZL Female 0 ... 29.60 346.45 No
7041 8361-LTMKD Male 1 ... 74.40 306.60 Yes
7042 3186-AJIEK Male 0 ... 105.65 6844.50 No

[7043 rows x 21 columns]
>>> |
    
```

Figure 4: string dataset

```

Customer gender SeniorCitizen ... MonthlyCharges TotalCharges Churn
0 7590-VHVEG 1 0 ... 29.85 29.85 1
1 5575-GNVDE 0 0 ... 56.95 1889.50 1
2 3668-QPYBK 0 0 ... 53.85 108.15 0
3 7795-CFOCW 0 0 ... 42.30 1840.75 1
4 9237-HQITU 1 0 ... 70.70 151.65 0
...
7038 6840-RESVB 0 0 ... 84.80 1990.50 1
7039 2234-XADUH 1 0 ... 103.20 7362.90 1
7040 4801-JZAZL 1 0 ... 29.60 346.45 1
7041 8361-LTMKD 0 1 ... 74.40 306.60 0
7042 3186-AJIEK 0 0 ... 105.65 6844.50 1

[7043 rows x 21 columns]
>>> |
    
```

Figure 5 : converted to binary

c. TRAINING AND TESTING DATA

The data should be divided into training and testing data. Training data use 70% and the testing data uses remaining 30%. The machine is prepared using the training dataset. Subsequent to training the machine will anticipate the yield of the using the testing dataset.

d. ALGORITHM IMPLEMENTATION

1. CNN (CONVOLUTIONAL NEURAL NETWORK)

CNN - class of deep neural structure that comes under feed forward neural network. The application of CNN has been primarily used in classification and recognition analysis with the help of image data. Additionally, CNN is additionally used as one dimensional data which is proposed in this paper.

ALGORITHM:

1. The first step is to initialize the weight of every layer in CNN.
2. The proposed model is created by using the one dimensional convolutional layer.
3. ReLU activation has been used in the convolution layer to which converts the negative value to zero.
4. Softmax function is used in the dense layer to classify the output based on the class labels.
5. Epoch is the time taken to train the dataset. By increasing the number of epoch accuracy will also be increased.

2. RNN (RECURRENT NEURAL NETWORK)

RNN - class of deep neural network. the input is fed to the layer by taking the output from the existing layer. it works same as the convolutional neural network but it uses the sigmoid function instead of relu function which is a s shaped curve that limits the output between 0 to 1. Then the softmax function is used to predict the class variable.

V. DISCUSSION AND RESULT

In this paper, different deep learning techniques such CNN and RNN has been used to predict the number of customers. The loss and accuracy of train and test data for two algorithms has been calculated. By increasing the number of epoch the accuracy of the algorithm increases. Among these two algorithms CNN shows the better accuracy with 80% while RNN shows 74% accuracy.

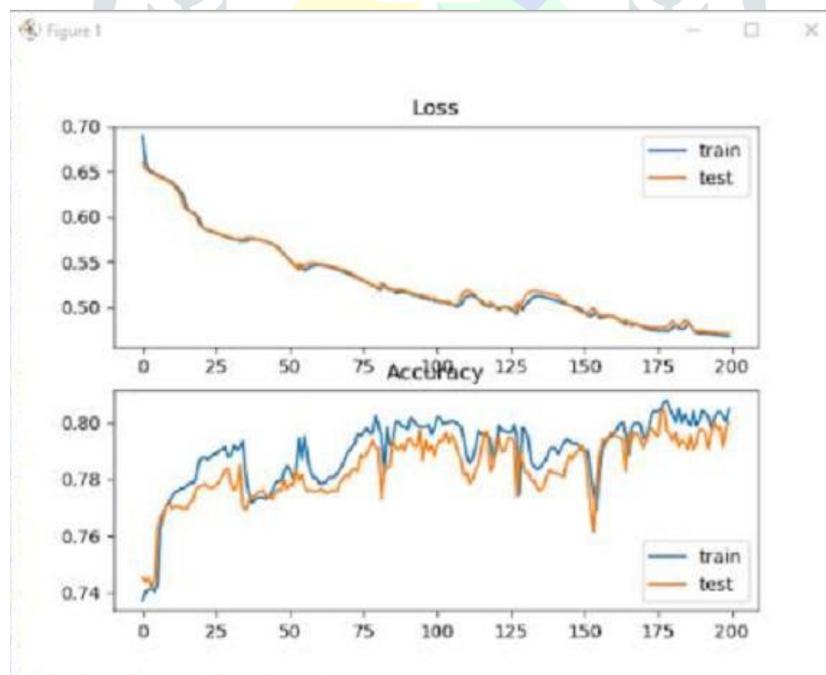


Figure 6: CNN result

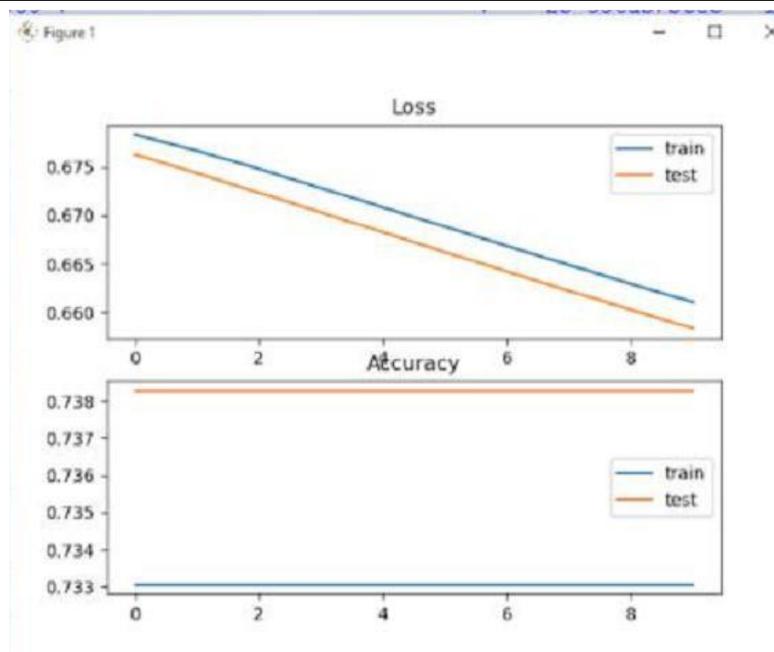


Figure 7: RNN result

VI. CONCLUSION

In this technological world, it is very important to stay competitive with the other partners. By analyzing and identifying the churn correctly prevents the loss for the company. In previous studies machine learning techniques has been used but it face the problem due to imbalanced data. In this paper, deep learning technique rectifies it by removing the missing values and redundant data. Since CNN is commonly used for classification of images, but this paper has conferred the suitability of CNN on one dimensional data for churn prediction. It shows the accuracy of 80% for CNN algorithm and 74% for RNN algorithm. It has been shown that the CNN algorithm comparatively shows the better result on detecting the customers who are about to likely move the competitor. This technique helps the customers to retain in their existing particular subscription.

REFERENCE

- [1] Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access*, 7, 60134-60149.
- [2] Saghir, M., Bibi, Z., Bashir, S., & Khan, F. H. (2019, January). Churn Prediction using Neural Network based Individual and Ensemble Models. In *2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST)* (pp. 634-639). IEEE.
- [3] Gaur, A., & Dubey, R. (2018, December). Predicting Customer Churn Prediction In Telecom Sector Using Various Machine Learning Techniques. In *2018 International Conference on Advanced Computation and Telecommunication (ICACAT)* (pp. 1-5). IEEE.
- [4] Ahmed, A. A., & Maheswari, D. (2017). Churn prediction on huge telecom data using hybrid firefly particle swarm optimization algorithm based classification. *IOSR Journal of Computer Engineering*, 19 (4), 30-39.
- [5] Khamlichi, F. I., Zaim, D., & Khalifa, K. (2019, October). A new model based on global hybridization of machine learning techniques for "customer churn prediction". In *2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS)* (pp. 1-4). IEEE.
- [6] Choudhari, A. S., & Potey, M. (2018, August). Predictive to Prescriptive Analysis for Customer Churn in Telecom Industry Using Hybrid Data Mining Techniques. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)* (pp. 1-6). IEEE.
- [7] Hammoudeh, A., Fraihat, M., & Almomani, M. (2019, April). Selective Ensemble Model for Telecom Churn Prediction. In *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)* (pp. 485-487). IEEE.