A Novel Approach for Churn Prediction Using Deep Learning

Abinash Mishra
Research scholar
Department of Computer Applications
National Institute of Technology
Tiruchirappalli, Tamil Nadu, India
405117002@nitt.edu

U. Srinivasulu Reddy
Assistant Professor
Department of Computer Applications
National Institute of Technology
Tiruchirappalli, Tamil Nadu, India
usreddy@nitt.edu

***Abstract*—Customer retention in telecommunication is one of the prime issue in customer relationship management (CRM). The primary focus of CRM is on existing customer as it is difficult to acquire new customers. The main goal of churn prediction is to classify customers into churner & non-churner. Towards this, deep learning as they are equipped with large increasing data sizes and uncover hidden pattern insights, detects pattern, underlying risks and alert the Telecom Industry about customer behaviour with a better accuracy as compared to the traditional machine learning methods. In this paper, Deep learning by Convolutional Neural Network (CNN) is implemented for churn prediction and it showed good performance in terms of accuracy. The experimental results shows that the predictive model for churn prediction out performs with an accuracy of 86.85%, error rate of 13.15%, precision 91.08, recall 93.08%, F-score 92.06%.**

Keywords— Customer Relationship Management; Churn Prediction; Telecom Industries; Subscribers; Deep learning; Convolutional Neural Network (CNN); Convolution;

I. INTRODUCTION

Now a days smart phones play a significant role in human life and due to this many telecom companies are coming with various type of value added subscriptions. Hence, customer retaining with same service provider became questionable. In telecom perspective churn is defined as customers leaving the services from one service provider and get the services from other, due to dissatisfaction in services or getting better services. In telecom company customers are grouped into prepaid and post-paid. Prepaid customers has to pay some amount before utilising the service provide by the company but in case of post-paid customer they need to pay the amount after utilizing the services. ***Natwar Modani et al.*** [1] explained how the prepaid and post-paid customers are being churned in telecom industry. The authors analysed that, post-paid customers are more likely to churn using key performance indicators based on call detail records.

Various telecom companies are coming with advanced tactics in order to predict the churned customer in early stage. Traditionally, various types of machine learning approaches like Decision tree, Random Forest, and Bagging etc., were applied to predict churned customer. ***Huang et al.*** [2] explained the churn prediction problem using the concept of big data in terms of 3V's i.e. volume, velocity and variety. They showed that variety parameter plays an important role in teleco churn prediction model. They integrate the idea business support system (BSS) and operations support system (OSS) to enrich the variety parameter in order to improve the performance. Advancement in business analytics, deep learning is creating a boom in telecom industry to predict the churn customer.

This section describes how efficiently Deep Learning approach can be utilized for the churn prediction process in the telecom industry with a better accuracy and less processing time. Deep Learning is a class of neural network based on the machine learning algorithm. The basic principle behind the designing of deep neural network for churn prediction model is to transform raw inputs i.e., the predictor attributes through multiple layers of neural network and pass the information from input layer to hidden layer by using sigmoid function. Sigmoid functions calculate the weights based on the input parameter and binds the values between 0 and 1. As the churn prediction model deals with the class labels 0 and 1 i.e., 0 for non-churn customers and 1 for churn customers. The churn prediction model shown schematically in Fig.1.

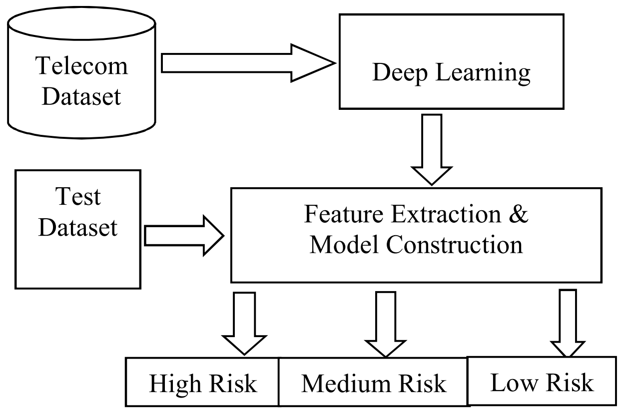

Figure 1: CNN Model for Churn Prediction

Deep neural network helps to deal with massive volume of telecom data in a more realistic and accurate manner because of layering architecture. Layering architecture of churn prediction model leads to hierarchical learning by forming number of hidden layers. Among the hidden layers information associated with each internal node is exchanged in terms of weights and finally the model is trained with the features which have more information for prediction of the class label in efficient way. Then the model is ready to predict for the test instances. The

978-1-5090-6621-6/17/$31.00 ©2017 IEEE

following section briefs the work carried out for churn prediction.

II. RELATED WORK

V. Umayaparvathi and K. Iyakutti [3] developed three deep neural network architecture and build the churn prediction model for cell2cell dataset and crowd analytic dataset. Also they showed that how feature selection techniques made the process more difficult in building the churn prediction model using traditional machine learning approach.

Maryam M Najafabadi et al. [4] showed that how deep learning used to analyse massive volume of data, extracting valuable information from huge dataset. Also they detailed that deep learning could be applied to certain kind of problems like data tagging, classification and prediction, information extraction from raw data, and semantic indexing. They investigated some aspects of deep learning to tackle certain challenges in big data analytics including multidimensional data and streaming data.

Niall McLaughllin et al. [5] designed an android malware detection system based on the concept of deep convolutional neural network. It was an innovative application of deep learning in the field of malware analysis. The model was capable of simultaneously learning to perform feature extraction and malware detection from large number of malware sample and, finally concluded that the proposed model is computationally efficient than *n-gram* based malware detection model.

Fedrico Castanedo et al. [6] proposed deep neural network an excellent tool for prediction of customer churn in such a way that, it forms multiple hidden layer and it propagates weights from one layer to the next layer. Deep learning helps to extract the features which have maximum information automatically, as a result there was a significant improvement in the performance of the model in terms of accuracy. They used billions of call records from telecom business enterprises and applied a feed forward network in order to design the predictive model for churn prediction. It achieved 77.9% area under curve (AUC) on validation data. Also, they extended the idea for fraud detection in various domains like banking, life insurance and telecom.

Rong Zhang et al. [7] proposed a churn prediction model for an insurance company and their main focus is to improve the model accuracy to maximize the profit. The proposed model was based on the combination of shallow model and deep model known as Deep Shallow Model (DSM). The authors applied this model on real dataset of an insurance company and concluded that, DSM performs better than the Convolutional Neural Network model (CNN), Linear Discriminant Analysis (LDA), Long Short-Term Memory (LSTM) etc. using the accuracy, precision, recall, F-score, AUC metric.

Joerg Evermann et al. [8] designed a model based on Deep Learning with recurrent neural network which is used to predict the next event in a business problem. They applied on two real datasets and showed that the proposed model surpass the state-of-the art in prediction precision.

Afan Ali and Fan Yangyu [9] proposed the idea of Automatic Modulation Classification (AMC) using deep learning approach. They applied this concept for classification of digitally modulated signals with varying channel condition using 2-layered feed forward deep neural network. They extended this idea for churn prediction problem by treating as binary classification task. They designed a two-stage intelligent system to classify the churn customer and non-churn customer.

Carlos Affonso et al. [10] applied the concept of Deep Learning for biological image classification. They applied the convolution function in order to design the Convolutional Neural Network model (CNN) to deal with the classification problem. They showed that the CNN helped for better feature extraction which in turn showed a significant improvement in the performance.

III. WORKING METHODOLOGY

A. *Convolutional Neural Network (CNN)*

The general idea behind Convolutional Neural Network (CNN) has been explained in four steps i.e. first convolution, second non-linearity, third pooling and finally classification. Convolution helps to extract the important features from the customer data, and preserves the dependency between the class label and input features. Non-linearity explains the use of functions like sigmoid and rectified linear unit (ReLu) to connect the input features with hidden layers of the model. Then, pooling operation used for dimensionality reduction of the input feature space to train the model faster. Finally the classification was executed by taking the input from the hidden layer and can be shown through the output layer.

CNN based churn prediction model consists of one input layer, one hidden layer and one output layer. In the input layer it is required to send all the information of customer from training set. From the customer information the sigmoid function calculates the weight of each edge which is linked to the hidden layer. The generated weight helps to bind the input information with the class label. The weights of the edge was improved using equation (1) and equation (3). Then the improved weighted equation is applied in equation (7) in order to calculate the class label of the customer. So it is required to increase the number of hidden layers to improve the performance of the CNN based churn prediction model.

$$f(x;\theta,b) = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k + b \ldots\ldots\ldots (1)$$

$f(x;\theta,b)$ Depends on the value of $\theta_1, \theta_2 \ldots\ldots\ldots \theta_k$ and parameter b, represent the co-efficient, $x_1, x_2, x_3 \ldots\ldots\ldots x_k$ and represents the feature vector. Equation 1 can be represented in vector form as follows:

$$f(x;\theta,b) = \theta^T x + b \ldots\ldots\ldots (2)$$

Sigmoid function helps to map the predicted values in between *0* and *1* which is shown in equation i.e.

$$f(x;\theta,b) = g(\theta^T x + b), where$$
$$g(z) = \frac{1}{1+\exp(-z)} \ldots\ldots\ldots (3)$$

Using the equation churn prediction model can be designed mathematically as follows:

$$f\left(x^{(1)};\theta,b\right) \approx y^{(1)} \dots\dots\dots (4)$$

$$f\left(x^{(2)};\theta,b\right) \approx y^{(2)} \dots\dots\dots (5)$$

$$\dots\dots\dots$$

$$\dots\dots\dots$$

$$f\left(x^{(m)};\theta,b\right) \approx y^{(m)} \dots\dots\dots (6)$$

$f\left(x^{(1)};\theta,b\right) \approx y^{(1)}$ represents the class label for first customer, $f\left(x^{(2)};\theta,b\right) \approx y^{(2)}$ represents the class label for second customer, m^{th} customer class label is represented as $f\left(x^{(m)};\theta,b\right) \approx y^{(m)}$

In order to minimize the objective function the following mean square error method is used:

$$l(\theta,b) = \left(f\left(x^{(1)};\theta,b\right) - y^{(1)}\right)^2 + \left(f\left(x^{(2)};\theta,b\right) - y^{(2)}\right)^2 + \dots + \left(f\left(x^{(m)};\theta,b\right) - y^{(m)}\right)^2 = \sum_{i=1}^{m}\left(f\left(x^{(i)};\theta,b\right) - y^{(i)}\right)^2 \dots\dots\dots (7)$$

IV. PROPOSED WORK

The proposed work is explained in four sub-sections. Sub-section one describes about the dataset which is used for the churn prediction model in telecom industry. Then in the next subsection, it describes the proposed model and also working methodology. Then result and discussion part is described with the help of confusion matrices for CNN model of churn prediction. Finally, performance of the model is compared in the table with corresponding bar-plot in terms of accuracy, error rate, precision, recall, F-score.

A. Dataset Description

The dataset which is used in this research paper was collected from the following web link:

http://www.ics.uci.edu~mlearn/MLRepository.html.

A snapshot of the dataset is given in Fig. 2 which is used for designing churn prediction model. Each customer information stored in the telecom database consisting of various customer's details like area, vmail, vmail.msgs, day.mins, day.calls, day.charge respectively.

All the fields associated for a particular customer represents the predictor attributes, in turn the predictor attributes helps to find the target attributes i.e. the customer is a churned customer or a non-churned customer. So it is required to train the model with the predictors attributes which possess more information for the target attribute. From the whole dataset 70% (2334) instances of data is used for training, and remaining 30% (999) instances is used for testing the model.

churn	area	vmail	vmail.msgs	day.mins	day.calls	day.charge	eve.mins	eve.calls	eve.charge	night.mins	night.calls
1	415	0	25	265.1	110	45.07	197.4	99	16.78	244.7	91
1	415	0	26	161.6	123	27.47	195.5	103	16.62	254.4	103
1	415	1	0	243.4	114	41.38	121.2	110	10.3	162.6	104
1	408	1	0	299.4	71	50.9	61.9	88	5.26	196.9	89
1	415	1	0	166.7	113	28.34	148.3	122	12.61	186.9	121
1	510	1	0	223.4	98	37.98	220.6	101	18.75	203.9	118
1	510	0	24	218.2	88	37.09	348.5	108	29.62	212.6	118
1	415	1	0	157	79	26.69	103.1	94	8.76	211.8	96
1	408	1	0	184.5	97	31.37	351.6	80	29.89	215.8	90
1	415	0	37	258.6	84	43.96	222	111	18.87	326.4	97
0	415	1	0	129.1	137	21.95	228.5	83	19.42	208.8	111
1	415	1	0	187.7	127	31.91	163.4	148	13.89	196	94
1	408	1	0	128.8	96	21.9	104.9	71	8.92	141.1	128
1	510	1	0	156.6	88	26.62	247.6	75	21.05	192.3	115
1	415	1	0	120.7	70	20.52	307.2	76	26.11	203	99
0	415	1	0	332.9	67	56.59	317.8	97	27.01	160.6	128
1	408	0	27	196.4	139	33.39	280.9	90	23.88	89.3	75
1	510	1	0	190.7	114	32.42	218.2	111	18.55	129.6	121
1	510	0	33	189.7	66	32.25	212.8	65	18.09	165.7	108
1	415	1	0	224.4	90	38.15	159.5	88	13.56	192.8	74
1	415	1	0	155.1	117	26.37	239.7	93	20.37	208.8	133
0	408	1	0	62.4	89	10.61	169.9	121	14.44	209.6	64

Figure 2: Snapshot of Dataset Used

B. Proposed Model

Model has been built for the churn prediction using the R studio tool in AMD A4-3305M APU with Radeon(tm), RAM 4GB processor and the same has to be tested to determine the performance of the model. The entire process has been schematically presented in Fig 2. The completely connected CNN based churn prediction model formed by joining all the input features with all the nodes of hidden layer. Each node associated with the hidden layer node with a certain weights and weights of each edge represents information with the input feature to the class label.

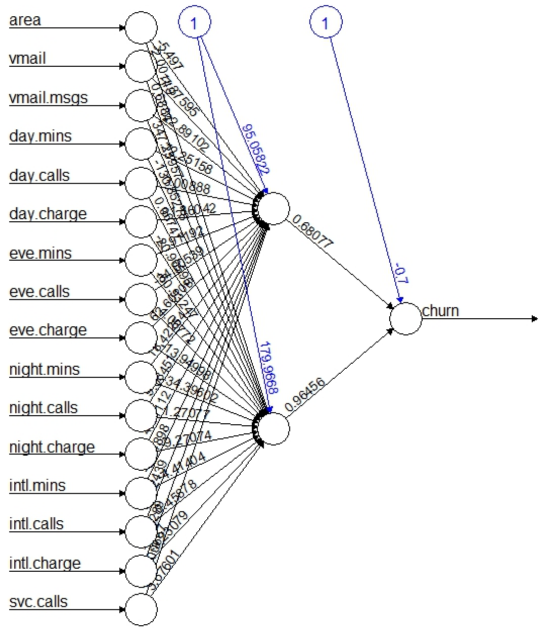

Figure 3: CNN Model for Churn Prediction

C. Performance Metrics

The performance of the CNN based Churn Prediction model can be evaluated by using the equations

$$Accuracy = \frac{TP + TN}{Total} \dots\dots\dots (8)$$

$$Error\ rate = \frac{FP + FN}{Total} \dots\dots\dots (9)$$

$$Precision = \frac{TP}{predicted\ yes} \ldots (10)$$

$$Recall = \frac{TP}{Actual\ yes} \ldots \ldots \ldots (11)$$

$$F - score = \frac{2(Precision * Recall)}{Precision + Recall} \ldots (12)$$

V. RESULT AND ANALYSIS

The result and analysis part describe customer churn prediction model using H2O package which is available in R studio tool. Designing and testing the model is as follows: first, the dataset is divided into training set and test in the ratio (60:40), that is 60% is used for training the model, once the model is ready for prediction, then remaining 40% of data is used to test the model.

Performance metric can be used to evaluate the model efficiency. The confusion matrix for the CNN based churn prediction model is given in the Table I.

TABLE I. CONFUSION MATRIX FOR CNN BASED CHURN PREDICTION MODEL

ACTUAL \ PREDICTED	0	1
0	**70**	**50**
1	**38**	**511**

The class label for the test instances represent the actual values and the model is used for prediction of the class label for test dataset representing the predicted values. These values are represented in a tabular form known as confusion matrix. 00^{th} position represents the True Positive value, 01^{th} position represents the False Positive, 10^{th} position represents False Negative and 11^{th} position represents the True negative value. Now all the performance metric parameters can be evaluated using the equation (8) to equation (12) which is shown in the Table II.

TABLE II. PERFORMANCE MEASURE FOR CNN BASED CHURN PREDICTION MODEL

Performance measure	CNN (in percentage)
Accuracy	86.85
Error rate	13.15
Precision	91.08
Recall	93.07
F-score	92.06

The churn prediction model using the convolutional neural network is efficient in terms of Accuracy, Error rate, Precision, Recall and F-score

VI. CONCLUSION AND FUTURE SCOPE

In this research paper, Customer Churn Prediction in Telecom Industry was proposed using Convolutional Neural Network (CNN). The experimental results showed that the CNN is best Classifier for the Churn Prediction Problem in terms of all the performance measures like accuracy, error rate, precision, recall and F- score.

Every customer is expecting good service or reward points from the service providers. Enabling the prompt services for the valid customers is more tedious task, because it is very difficult to predict the genuine customers for the company. The early churn prediction can prevent the company loss by predicting the customer behavior. In future, the model performance can be improved using tensor flow package with respect to time and accuracy.

VII. ACKNOWLEDGMENT

The authors would like to thank MHRD for the financial support. The authors would like to acknowledge the infrastructure support provided by Machine Learning & Big Data Analytics Laboratory, Department of Computer Applications, National Institute of Technology, Tiruchirappalli.

REFERENCES

[1] Natwar Modani, Kuntal Dey, Ritesh Gupta and Shantanu Godbole, " CDR Analysis Based Telco Churn Prediction and Customer Behavior Insights: A Case Study," Springer-Verlag Berlin Heidelberg, Vol. 11, pp. 256–269, 2013.

[2] Huang, Yiqing, Fangzhou Zhu, Mingxuan Yuan, Ke Deng, Yanhua Li, Bing Ni, Wenyuan Dai, Qiang Yang, and Jia Zeng, "Telco churn prediction with big data, " In Proceedings of the 2015 ACM SIGMOD international conference on management of data, pp. 607-618. ACM, 2015.

[3] V. Umayaparvathi, K. Iyakutti, "Automated Feature Selection and Churn Prediction using Deep Learning Models," International Research Journal of Engineering and Technology (IRJET), Vol. 4 (3), 2017.

[4] Maryam M Najafabadi, Flavio Villanustre, Taghi M Khoshgoftaar, Naeem Seliya Randall Wald and Edin Muharemagicl,"Deep learning applications and challenges in big data analytics," Springer Journal of Big Data, Vol. 2 (1), pp. 1-21, 2015.

[5] Niall McLaughlin, Jesus Martinez del Rincon, BooJoong Kang, Suleiman Yerima,Paul Miller, Sakir Sezer, "Deep Android Malware Detection, Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy, "pp.301-308, 2017

[6] Federico Castanedo, Gabriel Valverde, Jaime Zaratiegui, and Alfonso Vazquez, "Deep Learning to Predict Customer Churn in a Mobile Telecommunication Network," , pp 1-8, 2014.

[7] Rong Zhang, Weiping Li, Tong Mo and Wei Tan, "Deep and Shallow Model for Insurance Churn Prediction Service, " IEEE 14th International Conference on Services Computing, pp. 346-353, 2017.

[8] Joerg Evermann, Jana-Rebecca Resheb, Peter Fettkeb, "Predicting Process Behaviour using Deep learning," Elsevier Decision Support System, pp 129-140, 2017.

[9] Ali, Afan, and Fan Yangyu, "Unsupervised feature learning and automatic modulation classification using deep learning model," Elsevier Physical Communication Vol. 25, pp. 75-84, 2017.

[10] Carlos Affonso, AndréLuis Debiaso Rossi, Fábio Henrique Antunes Vieira and AndréCarlos Ponce de Leon Ferreira de Carvalho, "Deep learning for biological image classification," Elsevier Expert System with Applications, Vol. 85, pp. 114-122, 2017.