Journal of Positive School Psychology http://journalppw.com
2022, Vol.6, No.4, 4672-4680

Churn Prediction in Telecommunication Business using CNN and ANN

Madhuri D. Gabhane[1]
[1]Research Scholar, Gondwana University, Maharashtra
[1]Madhuri05121983@gmail.com

Dr. Aslam Suriya[2]
[2]Assistant Professor, Janata College, Chandrapur
[2]aslamsuriya@gmail.com

Dr. S.B. Kishor[3]
[3]HoD, Department of Computer Science, S.P. College, Chandrapur
[3]sbkishor@rediffmail.com

Abstract-
Customers play a critical part in the operation of the industries. The consumer's tendency to switch brands can have a variety of consequences. Customer churn forecast must be a critical component of every organization's strategy. As a result, clients who are likely to abandon their membership to a service can be identified more easily. Recently, the telecommunications industry has transitioned from a rapidly expanding business to one that has reached saturation. The goal of telecommunications companies is to transition away from the acquisition of new major customers and toward the retention of existing customers. Therefore, it is beneficial to identify which consumers are likely to switch to a competitor's product or service in the foreseeable future. Deep learning technologies such as Artificial Neural Network (ANN) and Convolutional Neural Network (CNN) are used in the development of the model for churn prediction in telecommunication companies. In this research, the dataset is collected from the Kaggle website. The dataset is then pre-processed using several techniques and the necessary features are extracted. Once the relevant features are extracted, they are given to Deep learning algorithms such as ANN and CNN to develop a model. To evaluate feature extraction and classification, several performance metrics are used, such as True Positive Rate (TPR), True Negative Rate (TNR), False Negative Rate (FNR), False Positive Rate (FPR), and precision.

Keywords— Churn prediction, Telecommunication, Deep learning techniques, ANN, CNN.

1. INTRODUCTION

In a firm, churn is regarded as "whenever a customer suspends a membership to a subscription they have been using." Individuals suspending Amazon/Netflix memberships are a prominent example. So, relying on one‘s utilization of the business, Churn Prediction predicts which customers are the most likely to cancel a membership. It is essential to acquire the above relevant data since attracting new customers is frequently more complicated and expensive than retaining existing customers. As a result, the knowledge acquired from Churn Prediction allows them to focus more on the customers who are most likely to leave. In the particular instance of Churn analysis, the outcome is a definite yes or no. As a result, it is a classification task in which it is necessary to predict 1 if the client is likely to come back and 0 alternatively. The client leaves for several reasons, including new players in the industry, offering lower prices, and so on. As a result, there is never an adequate reason why a client wishes to churn. As a result, one sensible solution is to find patterns in the given data and predict the reasons why clients want to churn. While there are multiple explanations of customer churn, attempting to prevent it is probably easy. It is dependent on how well the company is making the customers feel exceptional and provides personalized experiences to persuade people to stay.It has been assessed that the likelihood of trying to sell to an existing customer is approximately 60-70%, whereas the likelihood of selling to a new customer is as low as 5-20%. Furthermore, the cost of attracting new customers customer is 5 times higher than retaining an existing customer. The ability to determine churn is essential to avoid it. And this is where deep learning comes into play. Companies that solely rely on feedback from customers to predict churn frequently ignore other factors that influence churn. The quantity of information that modern businesses make developing deep learning (DL) models for customer churn much easier. Artificial intelligence (AI) or deep learning (DL)-driven customer churn is much more precise than every classification algorithm presently offered.

Churn prediction is a prevalent way to track how many clients the company has lost. Telecommunications firms frequently lose significant clients and, as a result, profits to competitors. The telecommunications industry has undergone transitions in the latest generations, including the growth of different services, technical advances, and greater competition as a result of activities [1]. One of the most challenging problems in the telecommunication sector presently is trying to retain customers with the highest churn risk [2]. Modern customers have a wide range of churn possibilities due to the growing number of telecom operators and increased competition. As a result, telecommunications industry participants are realizing the value of keeping existing ones rather than attracting new ones [3]. Sustaining existing customers is far more financially viable than attracting new ones [4]. Understandably, attracting new consumers is costly, which is why telecommunication companies have been focusing on retaining existing customers. Client churn is influenced by a variety of considerations. Prepaid clients, with exception of post-paid clients, really aren't obligated by service contracts, and they frequently churn for the most insignificant reasons. As a result, predicting one‘s churn rate is challenging. Some other aspect is consumer loyalty, which would be influenced by the providers' service and product quality. Clients may indeed be influenced to switch brands with a broader reach and good receiver reliability due to problems such as cell service and reception reliability. Gradual or insufficient response to criticism, as well as billing issues, are other considerations that increase the likelihood of clients switching sides to the competition. Clients may suffer damage to the competition due to factors such as packaging costs, insufficient features, and outdated technology. Consumers normally try comparing suppliers and switch to who they believe offers significantly better [5].

Numerous ML algorithms were used in current churn forecasting methods to improve predictive performance [6]. All of these models solely rely on structured information. With tree-based machine learning algorithms and the last several neural networks method, a few of them seem to have produced good results.Liu S et al.[7] in their paper stated that Unstructured data, such as client phone logs, could be used to start generating useful insight, according to their experiments, as well as open to interpretation machine learning must be used in all kinds of customer information management. By recognizing high churn risk customers slightly earlier and personalizing advertising strategy to attain customer loyalty and retention, the Superannuation company could conceivably end up saving huge amounts of money in profit. Pustokhina et al. [8] in their research presented a new ISMOTE-OWELM to define mobile communications churners which have shown the highest prediction results with the precision of 0.94, 0.92, and 0.909 on datasets I, II, and III, respectively. The ISMOTE-OWELM prototype outperformed the especially in comparison methods, according to the results of the experiment. Irina V., et al. in their research introduced a new model by using CCPBI-TABO method, has unveiled a novel churn predictive aimed at fostering data analytics in telecommunication companies. The CCPBI-TAMO prototype pre-processes company data in phases to improve its performance[9]. Alboukaey et al. in their research postulated two different approaches with using multiple regression time-series data to estimate churn daily. The Feature-Based approach implies retrieving significant features from time-series data using our delineated RFM functions (RFM-based model) or performing mathematical (statistics-based model), after which trying to feed those to a standard machine learning model, including such as Random Forest, to make accurate predictions. The Deep Learning-Based strategy, on the other hand, proposes creating deep neural networks models (CNN-based model and LSTM-based model) to gain knowledge relevant attributes from multidimensional data and estimate churn at the same moment[10].

A dataset of 3333 call details with 21 features, as well as a dependent churn parameter with two values: Yes/No, were utilised by Brandusoiu et al. [11] to demonstrate an advanced data mining methodology to estimate churn for pre-paid clients. Customers may view how many messages they have received and sent, as well as how many voicemails they have left, by logging into their account. The PCA algorithm (principal component analysis) was employed by the author to reduce the number of dimensions in the data set to a minimum. To anticipate churn, models such as neural networks, support vector machines, and Bayes networks were all employed. According to Lee et al. [12], customer happiness and switching prices have an impact on French mobile churn; nevertheless, while customer satisfaction remains constant, the higher the switching cost, the less likely it is that a customer will churn from a mobile service provider. Huang and colleagues [13] investigated the issue of client attrition in the context of a big data platform. It was the researchers' goal to demonstrate that the volume, diversity, and velocity of data can significantly boost the process of anticipating turnover in order to improve customer satisfaction. Data from the Operation Support and Business Support divisions of China's largest telecommunications company needed the use of big data platforms, which were developed in-house. The AUC statistic was used to assess the Random Forest approach.

The following is how the article is built: Section 2 discusses the materials involved in churn prediction. Section 3 discusses the data acquisition and the pre-processing techniques employed in this work such as Scaling, Data balancing, etc are briefed in section 4. ANN and CNN are two of the classification algorithms discussed in section 5. In Section 6, the performance of a variety of feature extraction and classification approaches is compared. Finally, in section 7 the conclusion is given

II. MATERIALS AND METHODS

In this project, initially, the data is collected which consists of several features, and then pre-processed using several methods and analysed using different Deep learning techniques, which ultimately leads to a sensible churn prediction with at most accuracy.Figure 1 given below is the block diagram of the stages involved in churn prediction.

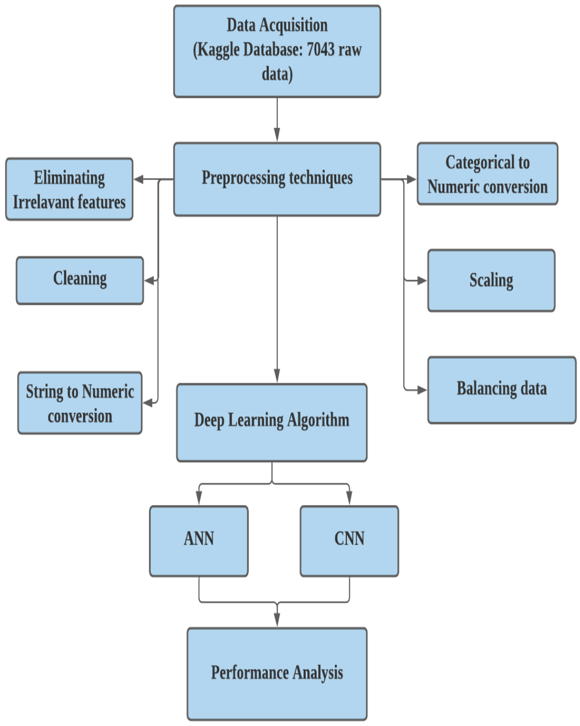

Fig.1 Sample image of the obtained data

From figure 1, it is shown that the dataset is initially collected from the Kaggle website. Then the obtained raw data is pre-processed using several pre-processing techniques such as Scaling, Data cleaning, converting string and categorical data to numeric data, removing irrelevant features, and balancing the data. Later after processing the raw data is analysed using deep learning algorithms such as ANN and CNN. Then the classification values are evaluated using several performance metrics.

III. DATA ACQUISITION

The requisite set of data for this study was collected from the Kaggle website a raw data count of about 7043, that are typically recorded customer information throughout their lifecycle through the use of source code such as Customer relationship management, web services, sentiment analytical techniques, social media listening tools, customer support applications, and others. A sample of the obtained data with various key facts is given below in figure 2.

customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSupport
6366-ZGQGL	Male	0	No	No	1	No	No phone service	DSL	No	...	No	No
1791-PQHBB	Female	0	No	Yes	2	Yes	No	DSL	Yes	...	No	No
8644-XYTSV	Male	0	Yes	No	42	No	No phone service	DSL	Yes	...	No	No
4115-UMJFQ	Male	0	No	No	69	No	No phone service	DSL	No	...	Yes	No
4367-NHWMM	Female	0	No	No	1	No	No phone service	DSL	No	...	No	No
9508-ILZDG	Female	1	No	No	34	Yes	Yes	Fiber optic	No	...	No	Yes
9824-BEMCV	Male	0	Yes	Yes	17	Yes	No	Fiber optic	No	...	No	No

Fig.2 Sample image of the obtained data

Figure 3 gives an idea about the features that are available in the raw data that is collected from Kaggle. The collected data has about 19 various features in it. Then the output will be churn or not. The churn is indicated by 1 and the not churn is indicated by 0.

#	Column	Non-Null Count	Dtype
0	customerID	7043 non-null	object
1	gender	7043 non-null	object
2	SeniorCitizen	7043 non-null	int64
3	Partner	7043 non-null	object
4	Dependents	7043 non-null	object
5	tenure	7043 non-null	int64
6	PhoneService	7043 non-null	object
7	MultipleLines	7043 non-null	object
8	InternetService	7043 non-null	object
9	OnlineSecurity	7043 non-null	object
10	OnlineBackup	7043 non-null	object
11	DeviceProtection	7043 non-null	object
12	TechSupport	7043 non-null	object
13	StreamingTV	7043 non-null	object
14	StreamingMovies	7043 non-null	object
15	Contract	7043 non-null	object
16	PaperlessBilling	7043 non-null	object
17	PaymentMethod	7043 non-null	object
18	MonthlyCharges	7043 non-null	float64
19	TotalCharges	7032 non-null	float64

Fig.3 Features present in the data

IV. PRE-PROCESSING METHODS

During the data preparation process, the collected data is consolidated and cleaned. Converting all of this primary data into a relational database so that the DL algorithm can sequence it is a significant step in data preparation. In this project, the pre-processing methods such as Eliminating irrelevant features, Cleaning, Scaling, String to numeric conversion, etc, are briefed below.

A. Eliminating Irrelevant Features

This approach minimizes the dimension of the data and improves result precision by removing irrelevant data from the raw dataset. Irrelevant features are those that are either completely or partially unrelated to the process. In this project, the dataset contains 19 different features in them. Out of them, the Caller ID feature is a randomly generated feature for every client which is considered to be an irrelevant feature. Hence the Caller ID feature is eliminated from the raw dataset as it is irrelevant for this project.

B. Cleaning

One of the most crucial processes of data cleaning is eliminating null values from the dataset. Any deep learning algorithm's precision suffers as a result of these missing values. The acquired raw dataset misses certain values and to deal with that problem Data cleaning process is done. The most typical way to analyze datais to remove the rows that contain missing values. The other way is to predict the missing value using an algorithm. In overview, based upon the nature of data and the problem to be solved, we can use various approaches to manage missing input data while data cleaning. In this project, 11 of the null values are completely removed by deleting the row.

C. String to Numeric Conversion

It is necessary to make sure that the values in the dataset are in the correct format to be used for prediction. In figure 3, the total charges are in object type. But the total charge should be numeric. Since the total charges in the acquired data are in the string format and it is necessary to convert them to float values to make them compatible.

D. Categorical to Numeric Conversion

All of the input and output parameters in deep learning models have to be numeric. Numerous learning algorithms are unable to function effectively on label data. Converting categorical data to numeric values can be done in three ways. They are as follows: Dummy Variable Encoding, Ordinal Encoding, and One-Hot Encoding. In this project, the categorical values are converted into Numeric values using the Method for label encoding. Label encoding is the process of converting labels into numerical form so that computers can interpret them. Deep learning methods can then assist in deciding how to use all of those labels. It is a critical pre-processing technique for the structured dataset in supervised learning. Figure 4 gives the available unique categorical values in each feature.

```
gender: ['Female' 'Male']
Partner: ['Yes' 'No']
Dependents: ['No' 'Yes']
PhoneService: ['No' 'Yes']
MultipleLines: ['No phone service' 'No' 'Yes']
InternetService: ['DSL' 'Fiber optic' 'No']
OnlineSecurity: ['No' 'Yes' 'No internet service']
OnlineBackup: ['Yes' 'No' 'No internet service']
DeviceProtection: ['No' 'Yes' 'No internet service']
TechSupport: ['No' 'Yes' 'No internet service']
StreamingTV: ['No' 'Yes' 'No internet service']
StreamingMovies: ['No' 'Yes' 'No internet service']
Contract: ['Month-to-month' 'One year' 'Two year']
PaperlessBilling: ['Yes' 'No']
PaymentMethod: ['Electronic check' 'Mailed check' 'Bank transfer (automatic)'
 'Credit card (automatic)']
Churn: ['No' 'Yes']
```

Fig.4 Sample image of Categorical feature

E. Scaling

A methodology for normalizing high-dimensional data or data characteristics is called scaling. In processing data, it's also referred to as data normalization, and it's typically done,mostly during the pre-processing step. A few deep learning models do not function reliably without scaling due to the obvious wide range of values in raw data. The original data in this project has a high degree of variability, which is a disadvantage because it takes the deep learning algorithm a long time to classify the data and the precision has also been tampered with. As a result, the values are converted to a 0-1 value range to make the data analysis easier for the deep learning algorithm. Figure 5 given below is a sample image of how the data values are converted into a 0-1 value range.

gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	OnlineSecurity	OnlineBackup	DeviceProtection	...	InternetService_DSL
1	0	1	0	0.000000	0	0	0	1	0	...	1
0	0	0	0	0.464789	1	0	1	0	1	...	1
0	0	0	0	0.014085	1	0	1	1	0	...	1
0	0	0	0	0.619718	0	0	1	0	1	...	1
1	0	0	0	0.014085	1	0	0	0	0	...	0
1	0	0	0	0.098592	1	1	0	0	1	...	0
0	0	0	1	0.295775	1	1	0	1	0	...	0
1	0	0	0	0.126761	0	0	1	0	0	...	1

Fig.5 Scaling the data values

F. Balancing Data

Having a well-balanced set of data for a classification model will indeed result in more accurate results, as well as a higher balanced precision and identification rate. As a result, a balanced data set is essential for a classification algorithm. The raw data obtained for this project is approximately 7043, and the data after cleaning is approximately 7032.And when the data after cleaning is processed with the ensembling method the data count is about 10092. Figure 6 given below gives the graphical representation of the data count in each process.

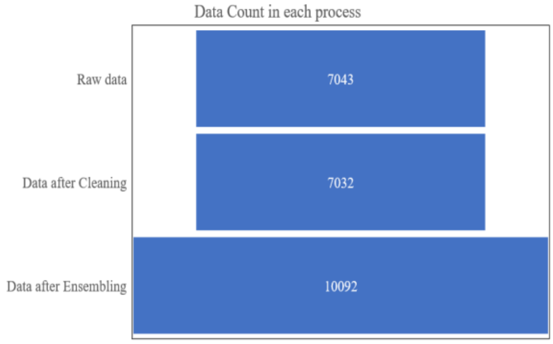

Fig.6 Data count in each process

V. CLASSIFICATION METHODS

Once the data is pre-processed using several methods the next step is to classify the processed data using Deep learning methods. In this project, the Deep Learning methods used are Artificial Neural networks (ANN) and Convolutional Neural networks (CNN). These categorization methods utilized in this project are detailed in the section below.

A. ANN

Regression analysis can be thought of as a single perceptron (or neuron). At every layer of an Artificial Neural Network, there are numerous perceptron's/neurons. The stimulation of the neuron is determined by the feature of the neuron. Stimulation is accomplished using a multitude of roles. The sigmoid [14] is used as activation function for binary classification in output layer. It is referred to as a Feed-Forward Neural Network because the input data is only analysed in the forward direction when it is fed into the network. An ANN is made up of three major layers: the input layer, the hidden input layer, and the output layer. Hidden layers are in charge of analysing and synthesising the information that is provided to them through various channels.In essence, each layer is attempting to learn specific weights. Any non-linear feature can be learned by an Artificial Neural Network. As a result, these networks are typically referred to as Universal Function Probabilistic models. ANNs can learn weights that correspond to any input to the desired output. To start with this method initially, the weight values are picked at random. The training, or learning, the process then commences. There are two types of training techniques: unsupervised and supervised training. Supervised training entails either manually "evaluating" the network's effectiveness or offering the desired outcome with the input data to provide the network with the target values. Unsupervised training is when the network is left to its own devices to make some sense of the input data. Supervised training is used by the vast majority of networks [15]. Artificial neural networks offer performance that is tough to compare with some other technology solutions in so many complicated issues, and ANN-based alternatives are exceedingly effective in advancement resources and time.

B. CNN

In a variety of fields, including pattern classification, the urge for deeply hidden layer upon layer has officially started to overperform old techniques. The Convolutional Neural Network is a type of deep learning model that is extensively used. Convolution is the name given to an algorithmic sequential procedure among matrices. CNN has several layers, including a convolutional layer, a non-linearity layer, a pooling layer, and a fully-connected layer. Variables aren't present in the pooling and non-linearity layers, but they appear in the convolutional and fully-connected layers. The CNN accomplishes admirably in deep learning problems. The convolutional layer and pooling layer combinations are stacked in the extraction of features neural networks. The convolution layer, as its names imply, uses the convolution to process the images. It's a combination of filter coefficients. The pooling layer joins adjacent pixels to create a single pixel. As a result, the pooling layer reduces the picture's depth. The activities of the convolution and pooling layers are abstractly in a two-dimensional plane, as the image is CNN's top consideration. One of the distinctions between CNN and other neural networks is this.

VI. RESULT AND DISCUSSION

For this project, the Kaggle database was used to acquire the raw data. The raw data count was about 7043 which is later pre-processed using several processing techniques such as removing irrelevant data, converting string and categorical values into numerical values, scaling, data balancing, etc. The data count is reduced to 7032 after all this process and the data is balanced using the ensembling methods resulting in 10092 as the final data value. Later after these steps, the data is subjected to classification using Deep learning methods such as CNN and ANN. The prediction output metrics such as True positive, True negative, False positive, and False-negative are then calculated. Figure 7 and 8 given below depicts the result of ANN and CNNconfusion matrix values.

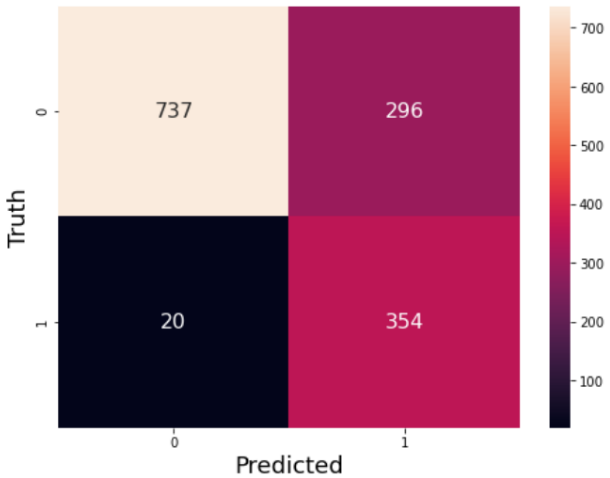

Fig.7 Result of churn predication using ANN

From figure 7 given above it is indicated that the ANN model precisely predicted that about 737 consumers will not churn from the company and about 354 consumers will churn from the company. The model also mistakenly predicated about 316 outcomes, which is relatively less when compared to the correct predictions.

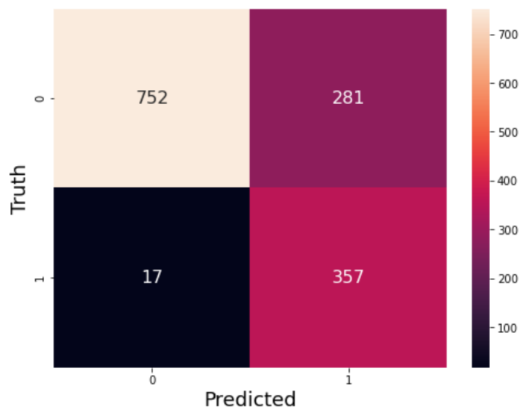

Fig.8 Result of churn predication using CNN

From figure 8 given above it is clear that the CNN model precisely predicted that about 752 consumers will not churn fromthe company, and about 357 consumers will churn from the company. The model also mistakenly predicated about 298 outcomes, which is relatively less when compared to the correct predictions.The final resulting values such as precision, True positive, True Negative rate are given in table 1 below.

Table 1: Analysis of Churn Predication various classifier

Model	TNR	TPR	FNR	FPR	PRECISION
ANN	94.65241	71.3456	28.6544	5.347594	97.35799
CNN	95.45455	72.79768	27.20232	4.545455	97.78934

From table 1 it is evident that the CNN methods yield high precision with about 97.78% whereas ANN yields a precision of about 97.35%. The true positive and negative rate is about 95.45% and 72.79% respectively, yielded by the CNN method whereas the ANN method yields about 94.65% and 71.34%. the False positive and negative rate is comparatively low for CNN with about 4.54% and 27.20% and the ANN with about 5.34% and 28.65% respectively. Figure 9 given below gives the graphical representation of the churn prediction using deep learning.

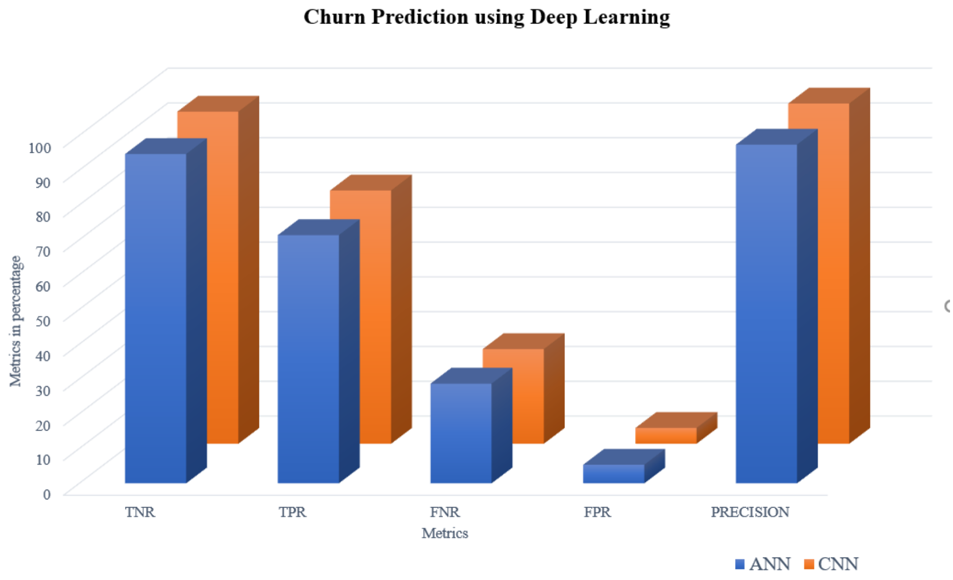

Fig. 9 Graphical representation of churn predication using deep learning

VII. CONCLUSION

The telecom industry is primarily reliant on the feedback provided by its customers. In the modern business world, customer unhappiness is no more something that can be pushed off. Due to the importance of customer retention in the telecommunications industry, organizations must concentrate their efforts on understanding the causes of customer churn so that they may put in place the required adjustments. The use of machine learning technologies in the telecom industry can help the company better forecast client attrition. One of the key objectives of this project is to examine the application of machine learning in a big data platform to estimate customer attrition in the telecommunications industry. Customer attrition in the telecommunications business can be anticipated utilizing artificial neural networks (ANNs), CNNs, and big data. For this project, the collected raw data from the Kaggle website were pre-processed using several techniques and then classified using two Deep learning methods. Out of the two, the CNN algorithm showed higher precision than the ANN algorithm with a value of 97.78%. This precise churn prediction with help the company to retain their existing clients skillfully and will also help them to formulate ideas on how to stop clients from churning.

REFERENCES

[1] Y. Huang, B. Huang, and M. T. Kechadi, "A rule-based method for customer churn prediction in telecommunication services," in *Advances in Knowledge Discovery and Data Mining, Springer*, pp. 411–422, 2011

[2] V. L. Miguéis, D. Van den Poel, a. S. Camanho, and J. Falcão e Cunha, "Modeling partial customer churn: On the value of first product-category purchase sequences," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11250–11256, Sep. 2012.

[3] A. Idris and A. Khan, "Customer churn prediction for telecommunication: Employing various various features selection techniques and tree based ensemble classifiers," in *Multitopic Conference (INMIC)*, 2012, pp. 23–27.

[4]Lu N, Lin H, Lu J, Zhang G (2014) A customer churn prediction model in telecom industry using boosting. *IEEE Trans Industr Inf* 10(2):1659–1665

[5]B. Balle, B. Casas, A. Catarineu, R. Gavaldà, and D. Manzano-Macho, "The Architecture of a Churn Prediction System Based on Stream Mining," in *Artificial Intelligence Research and Development: Proceedings of the 16th International Conference of the Catalan Association for Artificial Intelligence*, 2013, vol. 256, p. 157.

[6]Almana, A.M., Aksoy, M.S. and Alzahrani, R. A survey on data mining techniques in customer churn analysis for telecom industry. *International Journal of Engineering Research and Applications*, *4(5)*, pp.165-171,. 2014

[7]Vo, Nhi NY, Shaowu Liu, Xitong Li, and Guandong Xu. "Leveraging unstructured call log data for customer churn prediction." *Knowledge-Based Systems 212* (2021): 106586.

[8]Pustokhina, Irina V., Denis A. Pustokhin, Phong Thanh Nguyen, Mohamed Elhoseny, and K. Shankar. "Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector." *Complex & Intelligent Systems (2021)*: 1-13.

[9]Pustokhina, I.V., Pustokhin, D.A., Aswathy, R.H., Jayasankar, T., Jeyalakshmi, C., Díaz, V.G. and Shankar, K., 2021. "Dynamic customer churn prediction strategy for business intelligence using text analytics with evolutionary optimization algorithms". Information Processing & Management, 58(6), p.102706.

[10] Alboukaey, N., Joukhadar, A. and Ghneim, N., 2020. Dynamic behavior based churn prediction in mobile telecom. Expert Systems with Applications, 162, p.113779.

[11] Brandusoiu I, Toderean, "Churn prediction in the prepaid mobile telecommunications industry". *In: International conference on communications.* 2016. p. 97–100.

[12] J. Lee, J. *Lee, and L. Feick, ""The impact of switching costs on the customer satisfaction-loyalty link: mobile phone service in France,"* *Journal of Services Marketing*, vol. 15, no. 1, pp. 135–148, 2001.

[13]Huang F, Zhu M, Yuan K, Deng EO. "Telco churn prediction with big data*". In: ACM SIGMOD international conference on management of data.* 2015. p .607–18.

[14] Kukreja, H., Bharath, N., Siddesh, C.S. and Kuldeep, S., 2016. "An introduction to artificial neural network". Int J Adv Res Innov Ideas Educ, 1, pp.27-30.

[15]Maind, S.B. and Wankar, P., 2014. "Research paper on basic of artificial neural network". *International Journal on Recent and Innovation Trends in Computing and Communication, 2(1)*, pp.96-100.