

Customers Churn Prediction using Artificial Neural Networks (ANN) in Telecom Industry

Yasser Khan¹, Shahryar Shafiq²
Abid Naeem³, Sabir Hussain⁶
Department of Electrical
Engineering, Iqra National
University, Peshawar Pakistan

Sheeraz Ahmed⁴
Department of Computer Science
Iqra National University
Peshawar Pakistan

Nadeem Safwan⁵
Department of Management Science
Iqra National University
Peshawar, Pakistan

Abstracts—To survive in the fierce competition of telecommunication industry and to retain the existing loyal customers, prediction of potential churn customer has become a crucial task for practitioners and academicians through predictive modeling techniques. The identification of loyal customers can be done through efficient predictive models. By allocation of dedicated resources to the retention of these customers would control the flow of dissatisfied consumers thinking to leave the company. This paper proposes artificial neural network approach for prediction of customers intending to switch over to other operators. This model works on multiple attributes like demographic data, billing information and usage patterns from telecom companies data set. In contrast with other prediction techniques, the results from Artificial Neural Networks (ANN) based approach can predict the telecom churn with accuracy of 79% in Pakistan. The results from artificial neural network are clearly indicating the churn factors, hence necessary steps can be taken to eliminate the reasons of churn.

Keywords—Neural Network; ANN; prediction; churn management

I. INTRODUCTION

To survive in dynamic, high service demanding sector of telecommunication and to achieve operational excellence, it is indispensable to maintain up to date customer relationship management system (CRM). This system plays pivotal role in developing customer satisfaction, loyalty and main interface to interact with our clients [1]. The CRM system could revolutionize the entire system by providing real time provisioning of information, improvement in sale process and, enhancement of customer loyalty, advertisements and increase the effectiveness of marketing tools [2]. This term is basically emerge from relationship marketing defined as – the type of marketing where a lot of emphasis is laid to improve the qualitative, strategic and supportive relationship with already working customers and devise strategy to attract new customers.[3]. In order to survive in extreme competitive market of telecom, the firms must have to adapt with external changes, so CRM is proved to be an important monitoring tools to detect environmental changes on business operation and take tough business decision. CRM system is providing both tangible and non-tangible benefits hence former are providing increasing productivity, increase retention rates and profitability, decrease the marketing cost and faster turnaround time while later benefit is responsible for customer loyalty and satisfaction, take the benefits of segmentation, increase

customer experience, positive effect from word of mouth effects and excellence customer experience [4].

The income can be increased by improving service quality and customer satisfaction, redress the customer problems and timely tackle of complaints registered by customer [5]. The process can be more effective by introduction of automation through CRM system. The business capabilities can be improved by better and integrated knowledge in real time from CRM system and help significantly through better decision-making. By extracting detailed tables and report without spending much time which can better improve working productivity [6]. Excellent and comprehensive analysis can be executed through provisioning of customer service so individual planning and decision making can improved through better learning opportunity. The customer needs, expectations and behavior can be predicted through information derive from CRM [7].

The prime factors behind customer retention can be extracted in order to develop the profitable, loyal and long lasting relationship with their clients. Without running retention campaign, the telecom operators are consistantly losing significant numbers of customers that is 20% to 40% each years [8]. By bringing improvement in retention phenomena the aount £128 million rang from 20% to 10 % can anually be saved by one of british well know telecom operator Orange [9]. Many statistical and datamining techniques are introduced to investigate the customer churn prediction. In contrast to market survey, data mining techniques are analyzing the information obtained from both historical & current data in order to predict the patterns on historical data and future customer attitudes [10].

The most common techniques used for prediction are Decision Tree (DT), Logistic Regression (LR), Support Vector Machine (SVM) and Neural Network (NN). Further more, the decision tree is used for resolution of classification problems to divide the instances into two or more than two classes. Similarly, logistics regression gives the probability by providing input/output fields, set of equation and factors causing customer churn [11].

Most of the companies in today's world are suffered badly due to switching over of dissatisfied customers famously know as customer churn and departure is mostly done to new competitor. The acquisition of new customer costs new

company 6 to 7 times more than retaining the existing customer hence cause lot of profit lose [12]. The most probable reason behind the departure of customer is achieving of cheaper offer from another company, expression of dissatisfaction from existing operator or successful marketing strategy of new company [13].

Predictive Analytics can be divided into classifications and prediction which can further be sub divided into decision tree, logistic regression, neural network and support vector machine given in Fig. 1. To protect the profit, brand name and assets of exiting operator, it is deemed necessary to retain the customers. Hence collection of these customer data and necessary prediction would help identification of these dissatisfied customer and would help utilization of resources for these targeted customers.

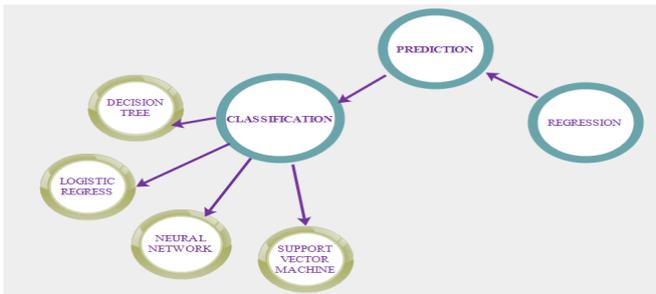


Fig. 1. Churn Prediction Techniques.

In the section-I brief introduction about customer churn is given which is followed by literature review in section-II and related work about churn and predictive analytics techniques is furnished. In the section-III deep learning is explained and illustrated followed by methodology in section-IV. In section-V, all the results of study are given in details.

II. LITERATURE REVIEW

The organization of recent times have significantly moved from product and service centric approach to customer centric behavior. To meet the needs of the customer is merely remain the core function of the organization but telecom operator are fast moving towards service improvement by enhancing customer loyalty and satisfaction. A lot of emphasis is laid to developed strategy to retain and categorize the customers according to the values they are providing to company [14].

Acquiring new customer is many times costlier than to retain the existing customer so strategic goal is set to work on dissatisfied customers who are intended to stop the service and leave the company and phenomena is termed as customer churn management. This involves prediction of those profitable customers in advance, are unable to continue with existing firm due to marketing strategy failure, company tariffs, poor customer care or frequent technical issues [15].

In this scenario data mining techniques are used to process and analyze the demographic details, customer care service data, billing and account receivables, customer credit scoring, customer usage behavior patterns, value added services data, customer relationship data. Hence model is developed and evaluated for potential churners and arrange to develop effective business solution to beat the competitor in the face of

extreme competition. This activate the CRM department to convince churning customer by redressing their issues, offering extra service and discount in monthly bills so loss could be prevented. The six steps initiating from data collection to devising of churn policy [16] is depicted in Fig. 2.



Fig. 2. The Six Steps for Customer Churn Prediction.

Churn prediction and analysis are performed through different techniques and covered mostly by data mining tools. Many different studies are conducted by researchers and telecom professional to construct churn prediction models with varying accuracy and precision on different data sets. Support Vector Machine (SVM), Logistic Regression (LR) and Multilayer Perceptron are being used for model creation from data set of Taiwan Logistic Company [17]. Similarly, decision tree algorithm (DT-ID3) is used for construction of prediction model for Malaysian Telecommunication Company [18]. Another study conducted on the basis of Decision Tree (DT) and logistic regression (LR) by taking the data from renowned Telecommunication Company by using survival analysis [19]. Research is conducted by using the data from Oracle Company's data set and techniques are Naïve Bayesian Theorem for building of churn prediction model [20]. In China Telecommunication Company, data was analyzed for churn prediction by using Decision Tree (DT), C5.0 algorithm BPN and Neural network [21].

III. DEEP LEARNING MODEL

This is the phenomena involve the processing from features selection to ultimate process of classification or prediction. There are three main requirement for effective deep learning mode. These are availability of abundant data to train the model, adequate system computational power lead to introduction of GPU's and innovative algorithm to make the process fast [22].

A. Artificial Neural Network

In human brain, all decision are taken through neural networks provided naturally in our body which are composed of basic building block 'neuron'.

The biological neuron as given in Fig. 3 is composed of dendrites responsible for receiving of data from other neuron, cell body sums all the inputs received inside and output the data through axon outside from neuron [23]. All the communications and processing is performed in electrical signals through synapses—a connection point between dendrites and axon from preceding neuron.

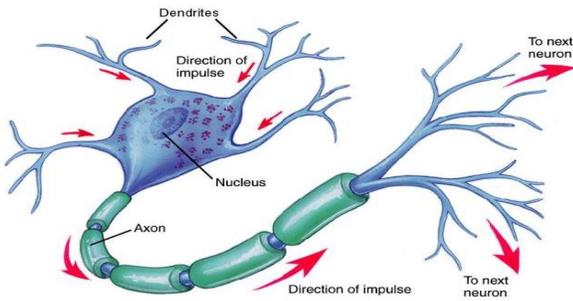


Fig. 3. The Biological Neuron.

Similarly, in artificial neuron inputs $X_1, X_2 \dots X_n$ are taken by each neuron and inserted to for summation and activation/ transfer function for decision making. The output is taken outside on basis of joint decision taken by entire neural network given in Fig. 4.

The neuron is composed of three main layers (input, hidden & output layers) provided in Fig. 5. When input X_i is applied to neuron, the weight is added to and results:

$$O_k = F(\sum W_i X_i + b_j) \tag{1}$$

W_i is weight for each input data X_i with b_j is the bias for each perceptron and O_k is the output.

The model is comprised of networks of neuron interconnecting with each other through weights and output from a particular perceptron is obtained from experience after completion of training process of a model [24].

Multilayer perceptron are perceptron consists of a scores of layer including a number of hidden layer in between input and output layer. In the training process, the error is consistently reduced calculated during comparison of actual and desired output and back feed to input side and weight for each input value is regularly adjusted.

$$E_i = A - \hat{A} \tag{2}$$

where E_i is the Error between actual and predicted outputs

A = Actual Output

\hat{A} = Predicted/ Modeled Output

B. Activation Function

In artificial neural network all the inputs are multiplied against their weights, sum and activation function is applied. There are lots of expectation from ANN to perform non-linear, complicated, high dimensional mapping between inputs and response output being considered as universal function approximates.

Sigmoid activation function or S-shaped curve range from 0 to 1. The function has some shortcomings of gradient problem vanishing and output is not zero centered hence push the gradient in far too different directions that's why having slow convergence.

On another hand, Hyperbolic Tangent function (tanh) is zero centered and best for optimization which make it superior over sigmoid function. The third and most popular function is Rectified Linear Units (ReLU) which has improved the

performance six times Tanh and completely remove the problem of vanishing gradient. The function is frequently used and limited to only hidden layer of neural network.

$$R(x) = \max(0, x) \tag{3}$$

If $x < 0, R(x) = 0$ And $x > 0, R(x) = x$

In addition, the modified version of ReLU is Leaky ReLU which has covered the shortcoming of ReLU and gradient stopped working during training phase and can die.

1) Sigmoid function :

$$f(z) = \begin{cases} 0 & \text{for } z < 0 \\ z & \text{for } z \geq 0 \end{cases}$$

2) Hyperbolic tangent function:

$$\tanh(z) = \frac{2}{1 + e^{-2x}} - 1$$

3) ReLU function:

$$f(z) = \frac{1}{1 + e^{-z}}$$

C. Convolution Neural Network

The use of convolution neural network has largely being extended for understanding of image contents, image recognition, detection and segmentation.

The main reason behind this fact is capability of CoVNet depicted in Fig. 6, for scaling the network to millions of attributes and labeled data set help greatly in learning process. The technique has also caught the attention of classification of large scale videos classification where the network has excellent performance not only in static images but also complicated temporal evolution.

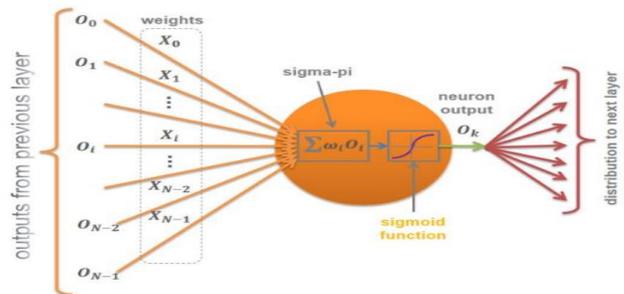


Fig. 4. The Artificial Neuron.

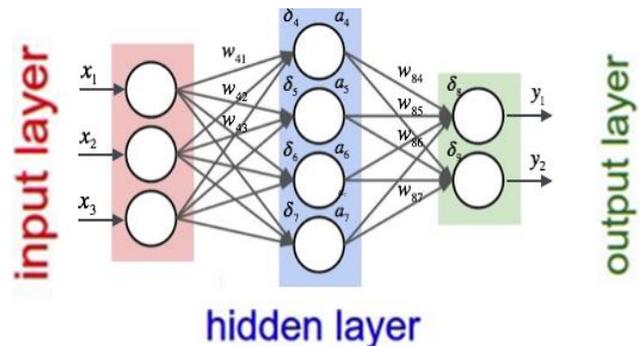


Fig. 5. The Layers of ANN.

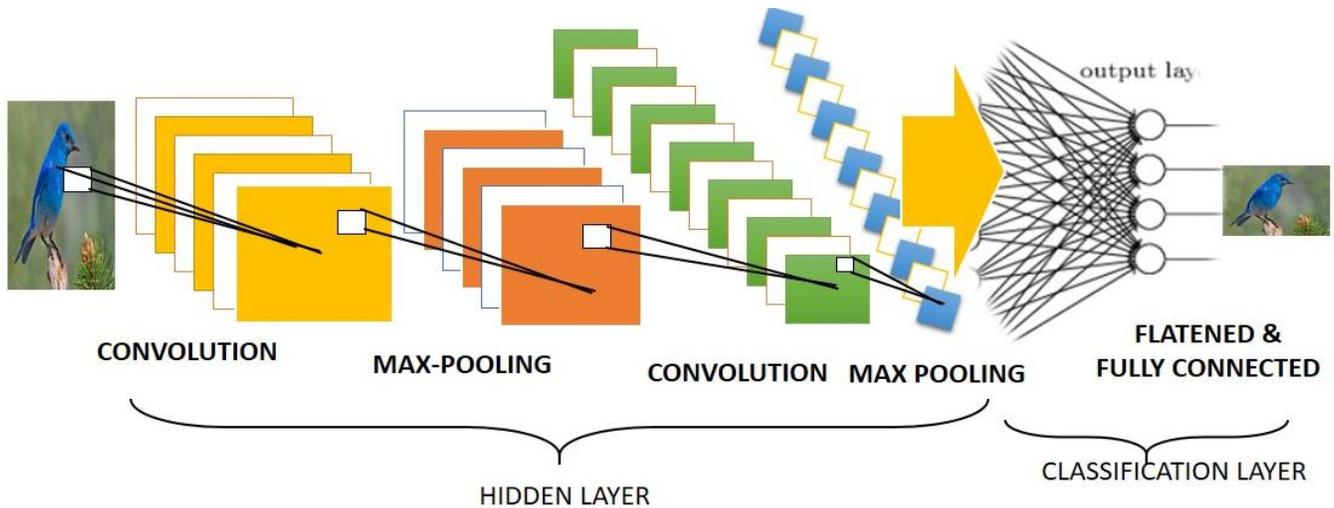


Fig. 6. Convolution Neural Network Layers Diagram.

IV. METHODOLOGY

A. Collection of Data

The collection of data were carried out from all cellular companies of Pakistan telecom market in Year 2018-19. There are 20468 instances comprise of 26 attributes selected during period of October to December 2018. The demographics information of the survey is illustrated in Fig. 7 indicating gender, education, marital and age information. However, Table I is detailed with descriptive statistics while the correlation of the data is furnished with another illustration, Table II.

B. Variable used for Design of Artificial Neural Network

1) *Predictor (Independent) variables:* The independent variable considered for this research studies is as below:

- (i) Age (ii) Gender (iii) Monthly income (iv) Call Drop (v) call failure Rate (vi) Calling number (vii) customer ID (viii) customer suspension status (ix) Education (x) Home Owner (xi) Marital Status (xii) Monthly Billed Amount (xiii) Number of Complaints (xiv) Number of month Unpaid (xv) Number of days contract equipment's (xvi) Occupation (xvii) Area (xviii) Total minutes used in last minutes (xix) Unpaid Balance (xx) Either use internet service (xxi) Use voice service (xxii) Percentage of call outside network (xxiii) Total Call duration (xxiv) Avg. call duration (xxv) Churn (xxvi) Years& Months

2) *Target variable:* The dependent variable in this research is churn of customer and will exhibit dichotomous outcome with two variables i.e. 0 or 1. The value 1 indicate that customer has switched from the company and moved to another operator while value 0 shows customer is still stick to existing firm.

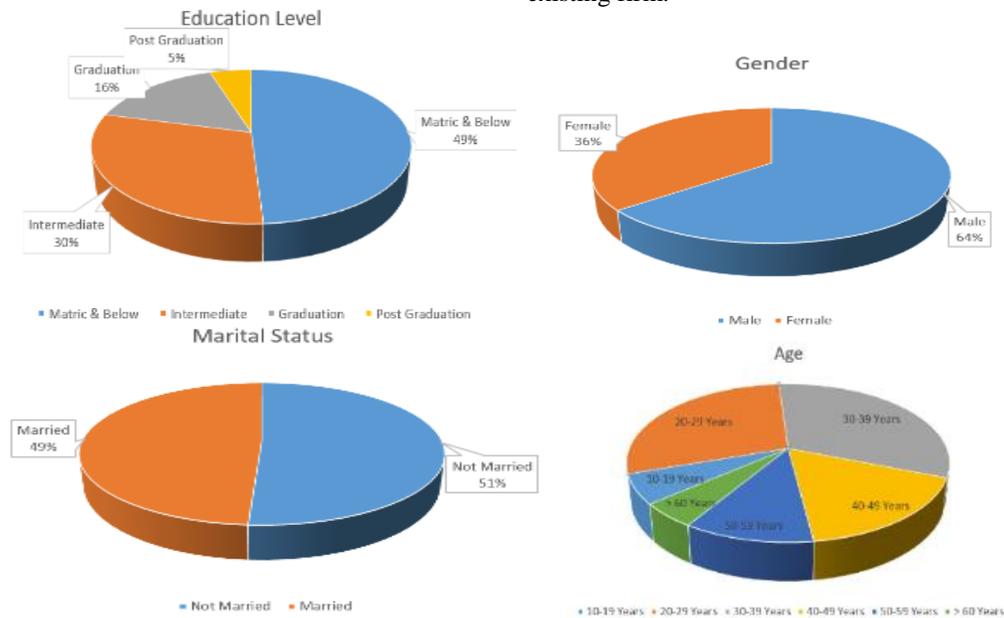


Fig. 7. The Pie Chart (Gender, Education Level, Marital Status and Age).

TABLE. I. DESCRIPTIVE STATISTICS

Descriptive Statistics									
	N	Mean		Std. Deviation		Skewness		Kurtosis	
	Statistics	Statistics	Std. Error	Statistics	Std. Error	Statistics	Std. Error	Statistics	Std. Error
Age	20468	45.33	.137	19.625	0.008	.008	.017	-1.199	.034
Education	20468	1.75	.006	0.887	.933	.933	.017	-1.101	.034
Gender	20468	0.49	.003	0.500	.47	.047	.017	-1.998	.034
Marital Status	20468	0.49		0.500	.041	.041	.017	-1.998	.034
Valid N(listwise)	20468								

TABLE. II. THE CORRELATION AMONG AGE, GENDER, TENURE AND MONTHLY CHARGES

Correlations					
		age	gender	tenure	MonthlyCharges
age	Pearson Correlation	1	.661**	.623**	.791**
	Sig.(2-tailed)		.000	.000	.000
	N	20468	20468	20468	20468
gender	Pearson Correlation	.714**	1	.685**	.822**
	Sig.(2-tailed)	.000		.000	.000
	N	20468	20468	20468	20468
tenure	Pearson Correlation	.565**	.598**	1	.656**
	Sig.(2-tailed)	.000	.000		.000
	N	20468	20468	20468	20468
MonthlyCharges	Pearson Correlation	.786**	.822**	.756**	1
	Sig.(2-tailed)	.000	.000	.000	
	N	20468	20468	20468	20468
** Correlation is significant at the 0.01 level (2-tailed).					

C. Design and Setup of Artificial Neural Network

The artificial neural network model was built in multilayer perceptron (MLP) in IBM SPSS statistics 21. The ANN is trained through back-propagation learning algorithm and synaptic weights were adjusted through gradient decent for lowering the error through transformation function. The required data were assigned to training and testing in the percentage of 80% and 20% respectively. Special care were made not to over-train the model in the training mode and reduce the error to minimum possible level. Before initialization of training, all the covariates were normalized with values of 0 and 1.

The transformation function used for this model is hyperbolic tangent function returned the values all the ways between -1 and +1 efficiently.

Hyperbolic Tangent Function Output:

$$O_j = \tanh(H_j) = \frac{e^{H_j} - e^{-H_j}}{e^{H_j} + e^{-H_j}} \quad (4)$$

ranges from [-1 , +1]

Soft max Function Output:

$$O_j = \sigma(H_j) = \frac{e^{H_j}}{\sum_{k=1}^m e^{H_k}} \quad (5)$$

ranges between [0 , 1]

The function can be termed as probability distribution and the output O_j is called as network's pseudo-probability or estimated probability of input classification function. There are two mode for optimization through gradient descent algorithm and the most commonly used method is batch mode which updates all synaptic weights of all records in training data set. Another method is online or incremental mode which do the same work however with lower efficiency than preceding method. For solution of linear equations through iteration, the technique used for in the category of batch mode is scaled conjugate gradient has become more optimized. Hence by each repetitive iteration, the synaptic weights are continuously updated, the algorithm bring the error surface to more minimum values compare to previous iteration.

The scaled conjugate gradient algorithm has four parameters for development of model that is initial lambda, sigma, interval center and offset. The Hessian Matrix when acting as negative then controlling parameter is lambda. When size of weight change by sigma parameter, Hessian estimate change the derivative of first order of error function.

Similarly, the annealing algorithm boosts the interval center a_0 and $a_0 - a$ and $a_0 + a$. So the initial setting of all parameters are:

Initial lambda = 0.0000005, Initial sigma = 0.00005 & Interval center = 0 and Interval offset = 0.5

The SPSS prefer to use cross-entropy error function rather than squared error function whenever softmax activation function is selected and applied on output layer. The mathematical form of cross entropy error function is as below:

$$\text{Error} = -\sum_{i=1}^n T_j \ln O_j \quad (6)$$

where n= total number of output nodes

T_j =target value (output node j)

O_j = Actual output value (output node j)

V. RESULTS

The main purpose of this research the generation of multilayer perceptron neural network (MLP-NN) model that can actually predict the telecommunication churn by processing of data obtained from telecom industry according to factors. Below mentioned Table III details the information of dataset applied for building ANN model.

TABLE. III. THE CASE PROCESSING SUMMARY

Case Processing Summary			
		N	Percent
Sample	Training	16421	80.20%
	Testing	4047	19.80%
Valid		20468	100.00%
Excluded		0	
Total		20468	

In the Table IV provide the total number of neurons used in input, hidden & output layer and all predictor variables (income per month, drop call rate, failure call rate, education, monthly bill amounts, complaint lodges for resolution, unpaid numbers, call duration- total, call duration- average, internet subscribers, Voice service used, total calling, occupation and age).

Moreover, the number of nodes calculated in the hidden layer is 3 while in the output layer is 2, the activation function taken for hidden layer is selected as hyperbolic tangent (tanH) and softmax function in the output layer and error function used as cross-entropy of network information table.

In network diagram extracted from SPSS results, the telecom customer churn (Churn=No, Churn=Yes) from inputs tenure.

Total charge. Similarly, there are three nodes in the hidden layer and two output nodes determined as churn, No-churn depend on results.

The network diagram of artificial neural network is furnished in Fig. 8 which are comprised of input layer, hidden layer and output layer after putting the data in software. The parameter estimates table is given in Table V while model summary in Table VI provide details information about training and testing data set. For both training and testing while the cross entropy error is brought down to minimum possible level during training processing. The power of the model to forecast the outcome of the model is calculated through small value of cross entropy error is 1623.861. The percentage of incorrect prediction is 22.7%. There are 10 consecutive steps completed was performed where there is no more reduction in error function during testing of sample is performed categorical dependent variable outcome.

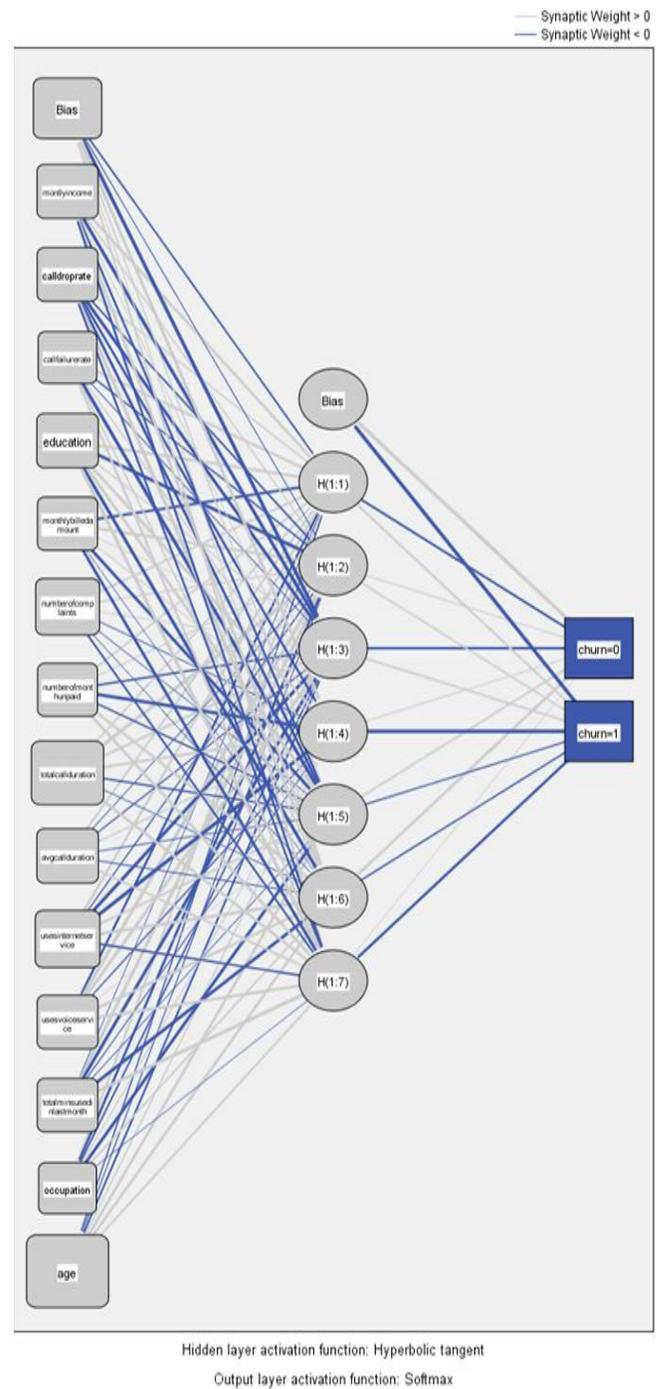


Fig. 8. The Network Diagram of ANN.

TABLE. IV. THE NETWORK INFORMATION

Network Information			
Input layer	Covariates	1	monthly income
		2	Call drop rate
		3	Call failure rate
		4	education
		5	Monthly billed amount
		6	number of complaints
		7	Number of month unpaid
		8	Total call duration
		9	Avg. call duration
		10	Uses internet service
		11	uses voice service
		12	Total mins used in last month
		13	occupation
		14	age
Number of Units			
Rescaling Method of Covariates		Standardized	
Hidden Layers(s)	Number of Hidden Layers		
	Number of Units in Hidden Layer1		
	Activation		Hyperbolic tangent
Output Layer	Dependent Variables	1	Customer churn
	Number of Units		
	Activation Function		Softmax
	Error Function		Cross-Entropy
a. Excluding the bias unit			

TABLE. V. THE PARAMETERS ESTIMATES

Parameter Estimates											
Predictor		Predicted									
		Hidden Layer1							Output Layer		
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	[churn =0]	[churn =1]	
Input Layer	(Bias)	-.202	-.046	-.559	.401	.151	.051	.245			
	monthly income	.362	.130	-.457	.036	-.275	.370	-.278			
	calldroprate	.166	-.201	-.436	-.311	-.361	.459	-.263			
	callfailurerate	.313	-.164	-.038	.125	-.467	.200	.311			
	education	.353	-.426	.106	.384	.376	-.441	.454			
	monthlybilledamount	-.308	.349	.020	.065	-.394	-.002	-.372			
	Number of complaints	.249	.135	.044	-.050	-.017	-.008	-.247			
	numberofmonthunpaid	.205	.110	-.203	-.426	-.180	.345	.368			
	Total call duration	.236	.473	.279	.328	-.195	-.032	.364			
	avgcalleduration	.255	-.037	-.117	.172	-.052	-.058	.262			
	usesinternetservice	.331	-.100	-.439	-.487	.278	.379	-.220			
	uses voice service	-.295	.492	.439	-.119	.002	.337	.363			
	totalminsusedinlastmonth	-.027	.488	-.438	-.056	-.027	-.423	.443			
	occupation	.280	-.336	-.194	.075	-.406	.215	-.005			
age	-.127	-.057	-.243	.164	.366	.355	.249				
Hidden Layer1	(Bias)								.799	-1.132	
	H(1:1)								-.274	.258	
	H(1:2)								.073	.183	
	H(1:3)								-.291	.222	
	H(1:4)								.086	-.490	
	H(1:5)								.238	-.138	
	H(1:6)								.367	-.180	
	H(1:7)								.007	-.354	

TABLE. VI. THE MODEL SUMMARY

Model Summary		
Training	Cross Entropy Error.	4856.634
	Percent Incorrect Predictions.	9.10%
	Stopping Rule Used.	1 consecutive step(s) with no decrease in error a
	Training Time.	00:00.7
Testing	Cross Entropy Error.	1196.824
	Percent Incorrect Predictions.	9.10%
Dependent Variable: Customer Churn		

For, classification (confusion matrix) is drawn in Table VII. When the predicted outcome is greater than 0.5 the outcome is determined as Yes (churn=Yes). The over percentage of training, 77.3% classification of training data was performed.

The pseudo-probability is calculated in box plot are drawn in Fig. 9. The dependent variable (Churn=Yes or No), the pseudo-probabilities obtained from whole dataset is displayed in box plots and each value greater than 0.5 display correct predictions. The first box plot extended from left to right are customer No churn in the category of No churn, the second plot indicates the classification of churn churn=No, although the values related are Yes category.

TABLE. VII. THE CLASSIFICATION

Classification				
Sample	Observed	Predicted		
		No	Yes	Percent Correct
Training	No	14926	0	100.00%
	Yes	1495	0	0.00%
	Overall Percent	100.00%	0.00%	90.90%
Testing	No	3679	0	100.00%
	Yes	368	0	0.00%
	Overall Percent	100.90%	0.00%	90.90%
Dependent Variable: Customer Churn				

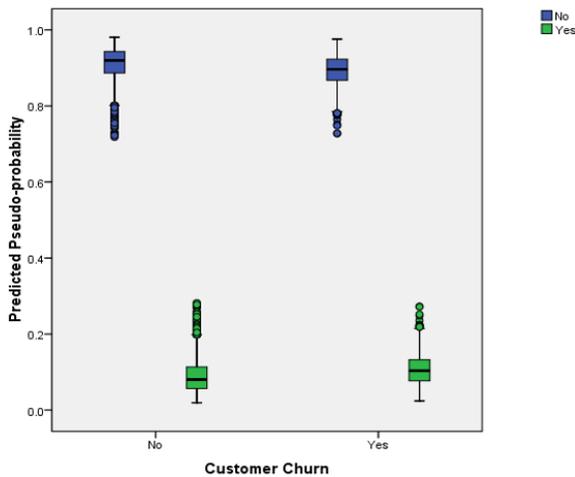


Fig. 9. The Box Plot.

Similarly, the third box plot actually relates to Yes category although predicted in Category churn=No and fourth plot indicated probability and actual are classified as customer churn=Yes.

To determine the possible cutoff point, the sensitivity versus specificity is classified through receiver operating characteristic curve (ROC) is illustrated in Fig. 10.

The combine sensitivity and specificity (1-false positive rate) is showing he random diagonal line is drawn from lower left side to upper right side at 45 degree and greater the accuracy can be achieved depend on the distance from 45 degree base line through classification process.

The mathematical determination of the area under the curve can determine in Table VIII. The probability predicted as 79% displays in the model having the customer churn= Yes and churn=No are randomly selected clearly reveal the pseudo-probability of churn prediction in the category churn=Yes.

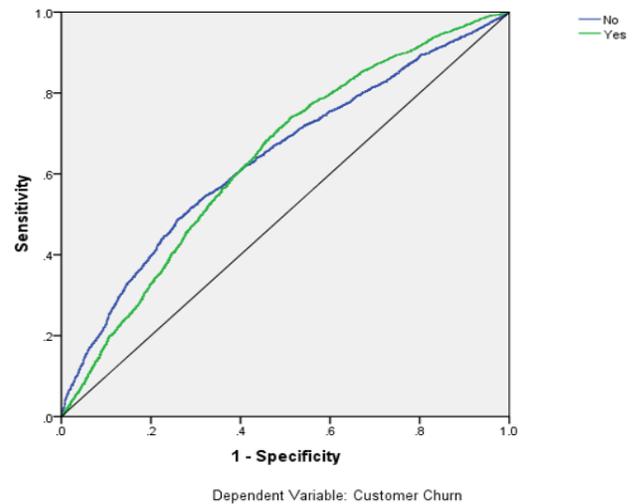


Fig. 10. The Receiver Operating Characteristic Curve.

TABLE. VIII. THE CURVE FOR AREA UNDER (AOC)

Area Under The Curve		
	Area	
Churn	No	0.79
	Yes	0.79

The cumulative gain that displays the classification of telecom customer churn calculated through artificial neural network against classification of prediction through chance. The fifth point in the curve at category churn=No (50% and 40%), meaning if dataset is rated and churn cases are predicted through pseudo-probability of category churn=No, then it is not difficult to determine that top 40% cases contain 50% of total cases actually take churn=No.

Simply, the gain given in Fig. 11 is the determination of effectiveness of classification model that the correct prediction out of total model against the prediction determine without using a model. The greater the distance of baseline curve main incline line the greater the gain a model have, which measure the excellent performance.

On other hand, lift curve drawn in Fig. 12 also evaluate the performance of the model according to the portion of population and give clear view of benefit of using the model to scenario where there is no predictive modeling. By comparing the gain curve with lift curve, it is determine that 93% value of gain curve, the lift factor is determine as 2.5 on lift curve.

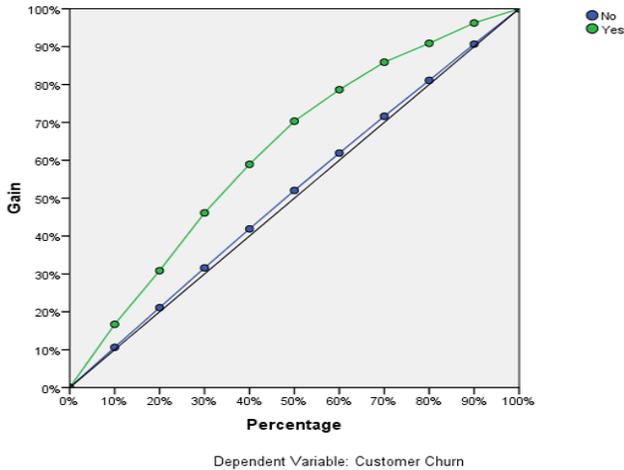


Fig. 11. The Gain Curve.

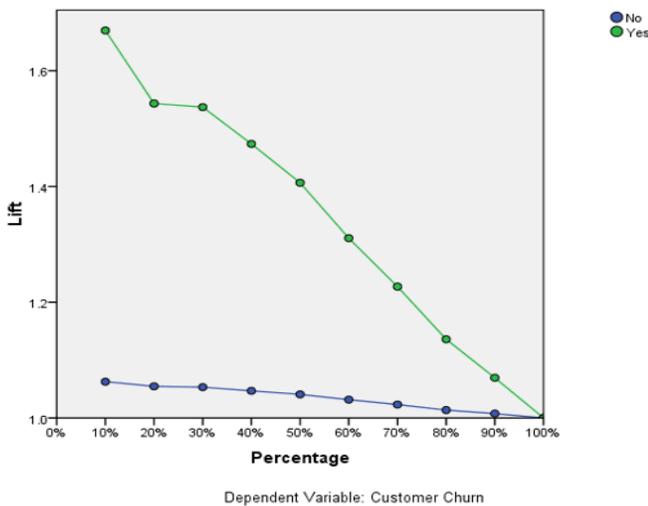


Fig. 12. The Lift Curve.

This figure indicate the sensitivity of model according to change of each input variable. The greatest impact from independent variable that is tenure/subscription of customer with company classify the customers to either churn =Yes or Churn= No. On second position, the churn is mostly affected by total charges.

Table IX is giving the importance of each variable used in main data set. The importance along with percentage is given in more detail for each variable where age has got the highest values equal to 100% followed by call duration (Total) equal to 63%, complaint lodge 33.3%, call duration(Avg) 25%, education 24%, income per month 23.30%, Drop call rate 21.9%, monthly income 20.1%, unpaid numbers 19.1% and occupation 11.4% at lowest level. All these percentage values are illustrated in bar chart provided in Fig. 13 for better understanding and arrange in descending orders for easy understanding.

TABLE. IX. THE INDEPENDENT VARIABLE IMPORTANCE

Independent Variable Importance		
	Importance	Normalized Importance
Income Per Month	.053	23.30%
Drop Rate Call	.050	21.9%
Failure Call Rate	.032	14.0%
Education	.057	24.7%
Monthly Bill Amount(Rs.)	.046	20.1%
Complaints Lodge	.076	33.3%
Unpaid Numbers	.044	19.1%
Call Duration(Total)	.147	63.9%
Call Duration (Avg.)	.057	25.0%
Internet Subscribers	.083	36.2%
Voice Service Used	.051	22.1%
Total Calling Minutes	.049	21.3%
Occupation	.026	11.4%
Age	.229	100.0%

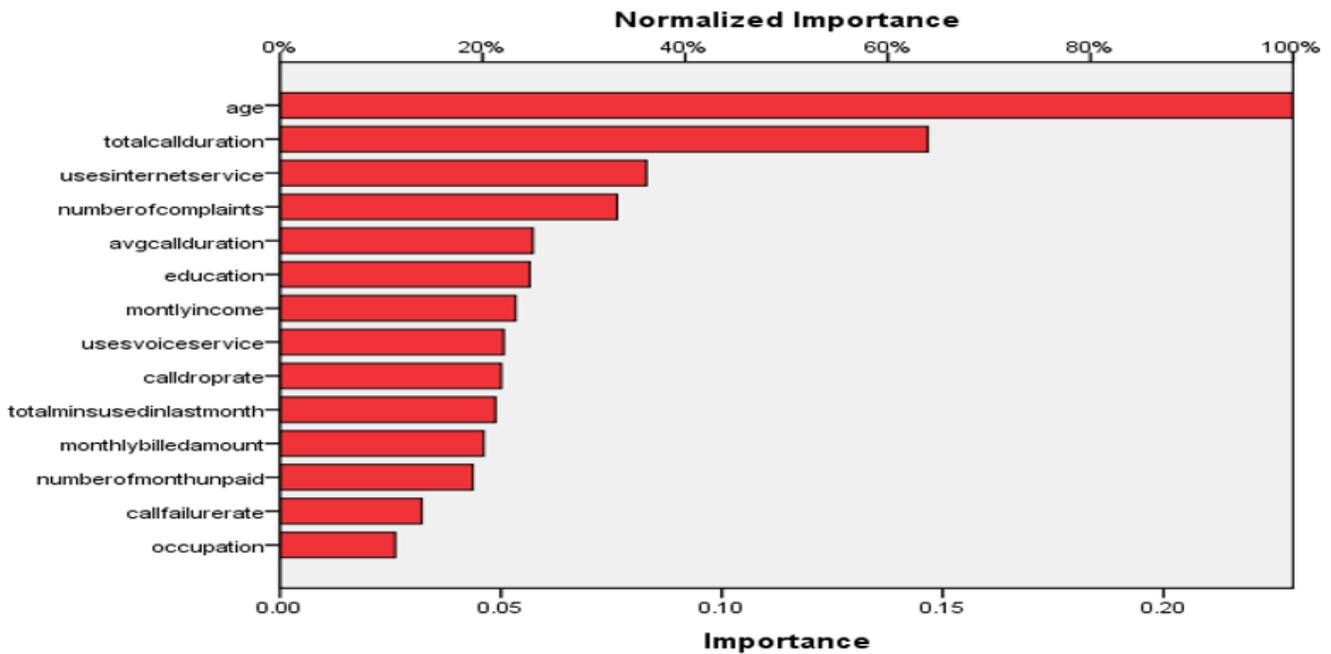


Fig. 13. The Normalized Importance.

VI. CONCLUSION

The prediction and management of customer churn has become more important task due to liberalization of cellular market. Timely prediction of loyal customers intended to leave the company can help identification and subject to the proactive action in order to retain them. Therefore, building of accurate and precise churn model is necessary not only for telecom companies' owner and practitioners as well. To determine a promising solution for maintaining strong customer baseline, telecom churn prediction has taken a shape of modern day research problem to issue an early warning system for switching over subscribers. The main theme of this paper is application of artificial neural network modeling in customer churn prediction in most volatile market of Pakistan telecommunication industry of the data obtained from all cellular and fixed operators of Pakistan. The research is already in line with literature review and has declared the Artificial Neural Network (ANN) performing excellent than other classification techniques. The already prevailing technique of back propagation algorithm is used to train the model for prediction of telecom customer churn. In the last part of the research the impact of each variable on telecom customer churn in term of normalized importance is provided to calculate the impact of each aspect.

REFERENCES

- [1] Hung, Chihli, and Chih-Fong Tsai. "Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand." *Expert systems with applications* 34, no. 1 (2018): 780-787.
- [2] Berry, Michael JA, and Gordon S. Linoff. *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons, (2013) pp. 12-19.
- [3] Laudon, Kenneth C., and Jane P. Laudon. "Managing the digital firm." *Managing Information Systems* (2014): 197-200.
- [4] Kim, Sangkyun. "Assessment on security risks of customer relationship management systems." *International Journal of Software Engineering and Knowledge Engineering* 20, no. 01 (2010): 103-109.
- [5] Burez, Jonathan, and Dirk Van den Poel. "CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services." *Expert Systems with Applications* 32, no. 2 (2017): 277-288.
- [6] Khodakarami, Farnoosh, and Yolande Chan. "Evaluating the success of customer relationship management (CRM) systems." In *Proceedings of the European Conference on Information Management & Evaluation*, (2015) pp. 253-262.
- [7] Uturytė-Vrubliauskienė, Laura, and Mantas Linkevičius. "Application of customer relationship management systems in Lithuanian mobile telecommunications companies/*Science-Future of Lithuania* 5, no. 1 (2013): 29-37.
- [8] Lee, Jonathan, Janghyuk Lee, and Lawrence Feick. "The impact of switching costs on the customer satisfaction-loyalty link: mobile phone service in France." *Journal of services marketing* 15, no. 1 (2011): 35-48.
- [9] Aydin, Serkan, and Gökhan Özer. "The analysis of antecedents of customer loyalty in the Turkish mobile telecommunication market." *European Journal of marketing* 39, no. 7/8 (2015): 910-925.
- [10] Wei, Chih-Ping, and I-Tang Chiu. "Turning telecommunications call details to churn prediction: a data mining approach." *Expert systems with applications* 23, no. 2 (2012): 103-112.
- [11] Ahn, Jae-Hyeon, Sang-Pil Han, and Yung-Seop Lee. "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry." *Telecommunications policy* 30, no. 10-11 (2016): 552-568.
- [12] Van den Poel, Dirk, and Bart Larivière. "Customer attrition analysis for financial services using proportional hazard models." *European journal of operational research* 157, no. 1 (2014): 196-217.
- [13] Petrova, Ekaterina, Pieter Pauwels, Kjeld Svidt, and Rasmus Lund Jensen. "In search of sustainable design patterns: Combining data mining and semantic data modelling on disparate building data." In *Advances in Informatics and Computing in Civil and Construction Engineering*, Springer, Cham, (2019) pp. 19-26.
- [14] Nabavi, Sadaf, and Shahram Jafari. "Providing a customer churn prediction model using random forest and boosted trees techniques (case study: Solico Food Industries Group)." *vol 3* (2013): 1018-1026.

- [15] Mestre, Maria Rosario, and Pedro Vitoria. "Tracking of consumer behaviour in e-commerce." In Proceedings of the 16th International Conference on Information Fusion, pp. 1214-1221. IEEE, 2013.
- [16] V. Umayaparvathi, K. Iyakutti,, "Attribute Selection and Customer Churn Prediction in Telecom Industry", Proceedings of the IEEE International Conference On Data Mining and Advanced Computing,(2019).
- [17] Chen, Kuanchin, Ya-Han Hu, and Yi-Cheng Hsieh. "Predicting customer churn from valuable B2B customers in the logistics industry a case study." *Information Systems and e-Business Management* 13, no. 3 (2015): 475-494.
- [18] Arsanl, Taner, and Safa Çimenli. "Churn Analysis and Prediction." *International Journal of Computer Science and Information Security* 14, no. 8 (2016): 550.
- [19] Junxiang Lu, "Predicting Customer Churn in the Telecommunications Industry -- An Application of Survival Analysis Modeling Using SAS", In SAS Proceedings, SUGI27, pages 114-127, 2012.
- [20] Nath, Shyam V., and Ravi S. Behara. "Customer churn analysis in the wireless industry: A data mining approach." In Proceedings-annual meeting of the decision sciences institute, vol. 561., (2013) pp. 505-510.
- [21] Liu, Peng, Naijun Wu, Chuanchang Huang, Bingrong Wei, Libo Wang, and Zhen'an He. "Research on Applying Data Mining to Telecom CRM." In International Forum of Information System Frontiers-2016 Xian International Symposium.
- [22] Jayne, Chrisina, Andreas Lanitis, and Chris Christodoulou. "Neural network methods for one-to-many multi-valued mapping problems." *Neural Computing and Applications* 20, no. 6 (2016): 775-785.
- [23] Adebayo, adelaja oluwaseun, and mani shanker chaubey. "data mining classification techniques on the analysis of student's performance." *gsj* 7, no. 4 (2019).
- [24] Andreu, David, and Marc Torrent. "Prediction of bioactive peptides using artificial neural networks." In *Artificial Neural Networks*, Springer, New York, (NY, 2015) pp. 101-118.