International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET207427

Customer Churn Prediction in Telecommunication Industry Having Data Certainty

V R Reji Raj, Rasheed Ahammed Azad .V
Department of Computer Science, Govt. Engineering College, Idukki, India

Article Info
Volume 7 Issue 4
Page Number: 113-122
Publication Issue :
July-August-2020

Article History
Accepted : 01 July 2020
Published : 07 July 2020

ABSTRACT

Customer Churn Prediction is a challenging activity for decision makers because most of the time, churn and non-churn customers have similar features. It is one of the major concerns for large companies, especially in the field of telecommunication field. Churn can be considered as a binary classification. The classifiers shows different accuracy levels at different zones of data. In such cases, a correlation can easily be observed in the level of classifier's accuracy and certainty of its prediction. So a mechanism to estimate the classifier's certainty for different zones within the data is needed so that the expected classifier's accuracy can be estimated. Here the classifier's certainty estimation is done using six sigma rule of normal distribution applied on the correlation values of all features in the dataset. Based on this the dataset is grouped into two categories such as (i) data having high certainty, and (ii) data having low certainty. Based on these criteria, classifier accuracy is estimated in the high distance zone. From the different evaluation measures like accuracy, f-measure, precision, recall and Receiving Operating Characteristics (ROC) area, the performance of classifier is evaluated. Then by applying a k fold approach the certainty of the classifier decision is estimated.

Keywords : - Churn prediction, Certainty, Telecommunication, Customer churn

I. INTRODUCTION

Customers are the most important asset in business. The customers have numerous choices of service providers that they can easily switch a service or even the provider. Such customers are called churned customer. The behavior of customers who shift from one service provider to another is known as churn. There are two types of customers – Non churning and Churning customers. Non churn customers remain loyal to the company and are rarely affected by the competitor companies. The second type is churn customers. The proposed model targets churn customers. The reasons for customer churn can be due to dissatisfaction, higher cost, low quality etc. Many organizations focus on establishing and maintaining long term relationships with their existing customers in order to retain the customers by providing them attractive offers. Finding the churners can help the companies to retain their customers.

Different machine learning techniques have been developed for churn prediction. A range of machine learning techniques are applied for classifying customer's data using the labeled datasets to find which of the algorithm best classifies the customers into the churn and non-churn categories. The proposed model targets churn customers and identify the reasons behind their churning. Many researchers have tried to establish various models to predict customer churn, for example the decision tree, support vector machine, neural network, genetic algorithm, logistic regression, linear regression analysis etc. However, current methods for churn prediction are not effective and need to be improved because there are many factors that contribute to customer churn.

In Telecommunication industry customer churn problem exists. Customer Churn Prediction in Telecommunication industry is due to fierce competition, launching new attractive offers by rival companies etc. Acquiring new customer can be more expensive than retention of existing customers. The machine learning communities have abundant information about Telecommunication sector to develop predictive modeling techniques to handle the Customer churn prediction. Information available are local/international call records, short messages, voice mail, demographics, financial detail and other usages behavior of the customers. Churners are predicted by using different data mining techniques. This includes ensemble techniques, probabilistic methods, Support Vector Machine, K-nearest neighbor, Neural Networks etc. These approaches help the companies to prevent their customers from churning by offering them promotions and better deals. But these approaches lack the required effectiveness due to problem complexity. Current methods need to be improved because there are many factors that contribute to customer churn. The prediction accuracy could be improved by removing some of the irrelevant features through reduction algorithms. Mostly the customer churn and non-churn have similar features and behavior, which increases the classification error rate. The classifier is uncertain about the decision and the level of certainty varies.

Machine learning is a method of data analysis using algorithms that iteratively learns from data. Machine learning allows systems to find hidden patterns without being explicitly programmed. There are three types of machine learning approaches unsupervised, semi-supervised and supervised. Supervised learning is the machine learning task of finding the hidden patterns from labeled datasets. Unsupervised learning is a machine learning task for finding the hidden patterns from unlabeled data. Semi-supervised learning is a type of supervised learning task that make use of unlabeled data for training. In this paper customer churn is predicted as well as calculate the level of certainty of the classifier. Classifier's decision falls into customer churn and non-churn with high certainty. The low certainty can be considered as uncertain classification for predicting the customer churns.

II. RELATED WORKS

Churn prediction has been performed using various techniques including machine learning, data mining, and hybrid techniques. These techniques support companies to identify, predict and retain churning customers and helps in decision making.

Bingquan Huang, et.al [1] proposed a method for the customer churn prediction by introducing a new set of features from the existing. Logistic Regression, Linear Classifications, Naive Bayes, Decision Trees, Multilayer Perceptron Neural Networks, Support Vector Machines and Evolutionary Data Mining Algorithm techniques are used. Based on the new feature set, the existing feature sets, the seven modeling techniques, comparative experiments were carried out. The experimental results showed that the new feature set is more effective for prediction than the existing feature sets. The modeling technique to choose depends on the objectives of the decision.

Ionut Brandsoiu, et.al [2] proposed four predictive models using the Support Vector Machine algorithm with different kernel functions. The four kernel functions used are Radial Basis Function kernel (RBF), Linear kernel (LIN), Polynomial kernel (POL), Sigmoid kernel (SIG. The Gain measure is used to evaluate and compare the performance. By evaluating the results, the model that uses the polynomial kernel function performs best.

A. Keramatia, et.al [3] proposed an improved churn prediction model using four classification techniques. Artificial Neural Network outperformed the other three, namely K-Nearest Neighbors, Decision Tree and Support Vector Machine. It also proposed a hybrid methodology in which all of the four above mentioned techniques are employed. It was shown that using the proposed technology a telecommunication company can gain a considerably higher accuracy for both Precision and Recall measures. Additionally, a new dimensionality reduction methodology is used to extract the most influential set of features. The high value of Recall and Precision obtained by the hybrid methodology is due to the application of four different classifiers.

Mohammad Ridwan Ismail, et.al [4] proposed a Multilayer Perceptron neural network approach to predict customer churn. From the nine training algorithms from the MLP neural network, the highest percentage of accuracy provided algorithm is used. The results are compared against the Multiple Regression Analysis and Logistic Regression Analysis. The result has proven the supremacy of trained model neural networks over the statistical models in prediction. Results show that a neural network learning algorithm is an alternative to statistical predictive methods in customer churn prediction.

Mohd Khalid Awang, et.al [5] proposed two different feature reduction algorithms - Correlation based Feature Selection and Information Gain and built, namely Bayesian Network, Simple Logistics, Decision Table. Experimental results showed that the feature reduction algorithm can improve prediction accuracy and at the same time producing lower error rates. The best prediction model with the highest prediction accuracy is obtained by the Correlation-based Feature Selection method with the Decision Table classifiers.

Sanket Agrawal, et.al [6] proposed a method for prediction of churn using Deep Learning. A multilayered Neural Network was designed to build a non-linear classification model. Once the model gets trained on the training data, it is verified by using validation data and then tested on the testing set. To determine the major factors that were responsible for the churn, a correlation graph will be generated. Microanalysis will also be done on numeric parameters to check for their effectiveness in the overall churn trends and try to predict patterns. The possibilities of Churn as well as the determining factors are predicted. The trained model then applies the final weights on these features and predicts the possibility of churn for that customer which results in high accuracy. The model also provides the churn factors, which are used by companies to analyze the reasons for these factors and take steps to eliminate them.

Shrisha Bharadwaj, et.al [7] proposed two models that predict customer churn with a high degree of accuracy. the first model is a logistic regression model which is a non-linear classifier, with sigmoid as its activation function. The second model is a Multilayer Perceptron Neural Network with a normalized input feature vector that is stacked with three hidden layers. These models predict customer churn as indicated by the client's behavior and are independent of other clients' data. The Multilayer Perceptron model performs well.

Irfan Ullah, et.al [8] proposed a Churn Prediction Model that uses classification and clustering techniques. Also the factors behind the churning in the telecom sector are also identified. Feature selection is done using information gain and correlation attribute ranking filter. The proposed model first classifies churn customer's data using classification algorithms, in which the Random Forest algorithm performed well. The model categorizes the churn customers in groups using cosine similarity to provide group-based retention offers after classification. Then identified churn factors that are essential in determining the root causes of churn. The results showed that the proposed model produced better classification using the RF algorithm and customer profiling using k-means clustering.

Adnan Amin, et.al [9] proposed a model for Customer churn prediction in the telecommunication industry using data certainty. In this paper, the classifier's certainty estimation is done using a distance factor. Based on the distance factor the dataset is divided into two categories - data with high certainty and data with low certainty, for predicting customers showing Churn and Non-churn behavior. Different evaluation measures are used. The results showed that the distance factor is highly dependent on the certainty of the classifier. The classifier obtained high accuracy in the zone with greater distance value.

Adnan Amin, et.al [10] proposed a model for Customer Churn Prediction in the Telecom Sector by efficient feature weighting technique using GA. Features reduction methods are used to efficiently handle and address the impact of the challenges faced by high dimensional dataset. Information loss resulting from feature reduction is reduced by automatically assigning more appropriate weights without involving domain experts. For that genetic algorithm is used to automatically assign weights to the attributes based on Naive Bayes classification. The experimental results demonstrated that the proposed technique outperformed as compared to without feature weighting.

III. METHODOLOGY

This section provides detailed descriptions of the proposed model. Fig. 1. visualize the overall process of the framework.

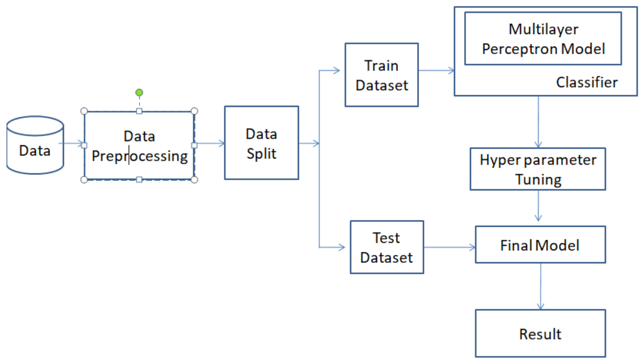


Fig. 1. Visualization of overall framework.

A. PROBLEM STATEMENT

The CCP is a binary classification problem where all the customers are divided into two possible behaviors: (i) Churn, and (ii) Non-Churn. From the existing studies there is no agreement on choosing the best approach to handle CCP problem. This might be because of the indifferent results shown by different classification techniques in different datasets. Therefore, the validity of the classifier, in terms of its certainty or uncertainty, in predicting the customer churn to be found out. There is no state-of-the-art study for checking the validity of the classifier. We propose to do that by considering six sigma rule of normal distribution applied on the correlation values of all features in the dataset for developing CCP model.

B. DATASET DESCRIPTION

We used a publicly available dataset containing churn data in Telecom sector. The dataset consists of 100000 samples and each sample represent individual customer, whereas, the ratio of churn and non-churn customers is 49.6% and 50.43%. This dataset contains 100 different attributes of telecom industry. The target variable is churn which indicates whether the customer will churn or not.

C. DATA PREPROCESSING

1) Noise Removal

It is very important for making the data useful because noisy data can lead to poor results. In telecom dataset, there are a lot of missing values, incorrect values like "Null" and imbalance attributes in the dataset. After analyzing the data, we have to exclude the data that are irrelevant from the dataset. These data has no influence in prediction. A method called imputing can be used to deal with missing values. Imputing refers to using a model to replace missing values with mean, median, or mode.

2) Feature Selection

Feature selection is a crucial step for selecting the relevant features from a dataset based on domain knowledge. A number of techniques exist in the literature for feature selection in the context of churn prediction. In this study, we use Correlation Attributes Ranking Filter techniques for feature selection. The dataset contains 100 attributes and in such a high dimensional dataset, some attributes improve performance measure while others contribute less. The performance of classification increases if the dataset contains highly predictive and valuable variables. Therefore, focusing on selecting significant features and decreasing the number of irrelevant attributes, classification performance can be improved. From this dataset, only the top 20 attributes are selected, having high ranking values.

Correlation Attributes Ranking Filter technique is used for selecting a subset of relevant features. Using this technique, the ranking of the most significant subsets of attributes are selected having low computational cost and reduces the dimensionality problems. After finding the correlation, a cutoff is set using six sigma methodology. In a dataset that is normally distributed will contain approximately 99.7% of the data within 3 standard deviations of the mean, which is a fundamental concept used in statistical process control. It is used as a normality test if the population is potentially not normal.

For selecting features, a cutoff point is made in the normal distribution curve. With one standard deviation from the spearman correlation coefficients of all features plotted against target variable 'churn' both in two directions. The spearman coefficient values were spread over the curve with mean zero and end points +1 and -1. The features lying above and below this cutoff have coefficients very close to +1 and -1. Those features lying in this range are having very high correlation with target variable churn. Those features are highly dependent either positively or negatively. These features are extracted from this range. Now our dataset contains only these highly dependent attributes.

D. CUSTOMER CLASSIFICATION AND PREDICTION

Classification is the process of predicting to which class an item belongs. The proposed model targets only churn customers. In the proposed method deep learning is used. Deep learning is a type of machine learning, which utilizes a hierarchical level of artificial neural networks, to carry out the process of machine learning. The artificial neural networks works like the human brain, with neuron are connected together like a web. The hierarchical function of deep learning system enables machines to process data with a nonlinear approach. This hierarchical or multi-layered structure forms the predictive capability of neural networks.

Multilayer Perceptron model is used for Churn Prediction. Multilayer Perceptron consists of multiple layers of computational units, interconnected in a feed-forward way. In a feed forward network, each neuron in one layer has directed connections to the neurons of the subsequent layer. The input neurons are connected to hidden neurons. At the end of the graph, are the output neurons. The output of these neurons is defined to be the output of the entire network. Simple learning algorithms can be used to train Perceptrons.

The architecture of the Multilayer Perceptron model is shown in Fig. 2. The model is build using Keras. Its sequential method allows building the model layer by layer. A dense input layer is created with 512 neurons. Five dense hidden layers were constructed with 256, 128, 64, 32 and 16 neurons in each layer. Output layer contains only one neuron. The activation function and dropout are specified in each layer. ReLU function is used. The main advantage of using the ReLU function is that it does not activate all the neurons at the same time. If the output of the linear transformation is less than 0, then only the neurons will be deactivated Since only a certain number of neurons are activated, the ReLU function is more computationally efficient when compared to other functions.

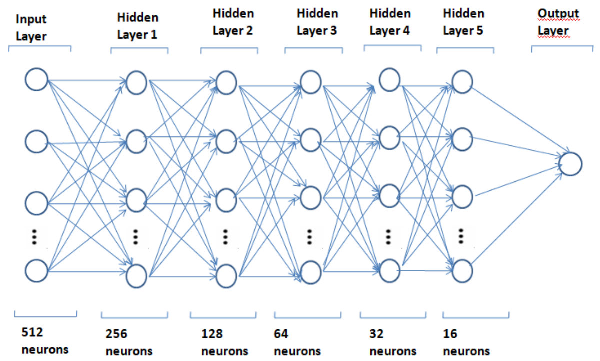


Fig. 2. Architecture of Multilayer Perceptron model.

For training the model, in forward pass we just pass the input to the model. The network processes the input forward, activating neurons as it goes to finally produce an output value. The output that we got from the model is called the Predicted output. Already we have expected output. This predicted output is compared to the expected output and an error or loss is calculated. This error is then send back through the network. The weights are updated according to the error calculated. The weights can be updated from the errors calculated for each training example. This is called the back propagation algorithm. The process is repeated for all of the examples in the training data. The process of updating the network for the entire training dataset is called an epoch. A network may be trained depending on the number of epochs set. The amount that weights are updated is controlled by configuration parameters called the learning rate. It is also called the step size and controls the step or change made to network weight for a given error.

After calculating the loss, it is back propagated and weights are updated using gradient. This is the main step in the training of the model. In this step, weights will adjust according to the gradient flow in that direction. After trained the neural network it can be used to make predictions. The predictions can be made on test or new data. Predictions can be made by providing the input to the network and performing a forward-pass, producing an output that can be used as a prediction.

When we train the model by updating each of its weights, it become too dependent on the dataset we are using. Therefore, when this model has to make a prediction, it will not give satisfactory results. This is known as over-fitting. To overcome this problem, a technique called dropout is used. Dropout is a technique where randomly selected neurons which are ignored during training. Applying dropout increases the training time.

E. HYPERPARAMETER TUNING

Hyper parameter tuning is the method for choosing a set of optimal hyper parameters for a learning algorithm. Hyper parameter optimization is a big part of deep learning. One method for optimizing hyper parameters is Grid Search. In GridSearchCV approach, machine learning model is evaluated for a range of hyper parameter values. It searches for the best set of hyper parameters from a grid of hyper parameters values. It works by searching exhaustively through a specified subset of hyper parameters. The benefit of grid search is that it is guaranteed to find the optimal combination of parameters supplied. The drawback is that it can be very time consuming and computationally expensive. Hyper parameter tuning improved the models performance.

The parameters to be tuned are the following. Batch size, which is the number of patterns shown to the network before the weights are updated. Epoch, which is the number of times that the entire training dataset is shown to the network, during training. Learning rate, which controls how much to update the weight at the end of each batch weight update. By using hyper parameter tuning the accuracy can be improved.

IV. RESULTS AND DISCUSSION

In this section, the results of the proposed study were explored and evaluated through the evaluations measures such as precision, recall, f-measure and accuracy. The precision, measures to what proportion the predicted positives are truly positive. Recall on the other hand represents what proportion of actual positives is correctly classified. The harmonic mean between precision and recall is F-measure. F-measure is the mean of an individual's performance, based on two factors i.e. precision and recall. The mathematical equations of these evaluation measures are shown below.

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{F-measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Further, the stability of the model is to be validated. Some kind of surety is needed to show that it's low on bias and variance. The simplest method used to evaluate the performance of a machine learning algorithm is to use different training and testing datasets. Here we use a k-fold cross validation type approach for analysis. The dataset is divided into three different training and testing partitions. The percentage of selecting the training and testing data are different. These types of three evaluations are done separately. In addition to specifying the size of the split, random seed is also specified. By specifying the random seed we ensure that we get the same random numbers each time, we run the code.

In order to evaluate the performance of the classification problem, AUC (Area under the Curve) ROC (Receiver Operating Characteristics) curve is used. This is the most important evaluation metrics for checking classification model's performance. ROC represents the probability curve and AUC represents degree or measure of seperability. It tells how much the model is capable of distinguishing between classes. The ROC curve is plotted with TPR (True Positive Rate) on y-axis and FPR (False Positive Rate) on the x-axis. An excellent model has AUC near to the 1 which means it has good measure of seperability. A poor model has AUC of 0 which represents the worst measure of seperability.

Accuracy is another evaluation criteria used. It is the proportion of true results among the total number of cases examined. It is a valid choice of evaluation for classification problems. The AUC-ROC curve of the Multilayer Perceptron model is shown in Fig. 3.

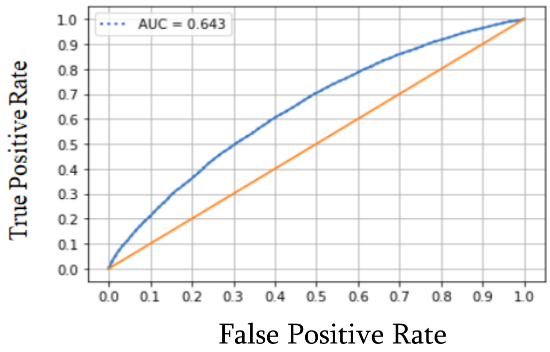


Fig. 3. AUC-ROC curve of Multilayer Perceptron model.

Tuning is the process of maximizing a model's performance without over-fitting or creating too high variance. In machine learning, this is accomplished by selecting appropriate "hyper parameters." After Hyper parameter tuning the model's performance is improved. The AUC - ROC curve of the model after Hyper parameter tuning is shown in Fig. 4.

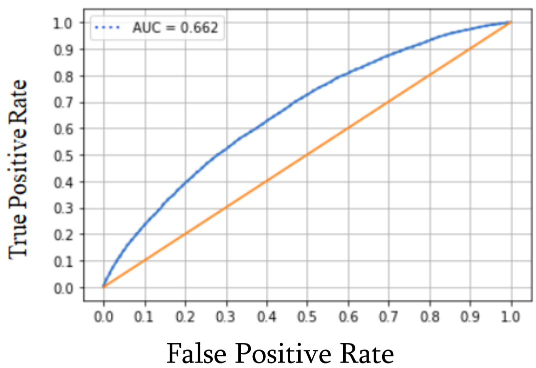


Fig. 4. AUC-ROC curve after Hyper parameter Tuning.

The first k fold type model is trained and evaluated using 75% training data and 25% testing data. The AUC - ROC curve of the first model with 0.25% split data is shown in Fig. 5.

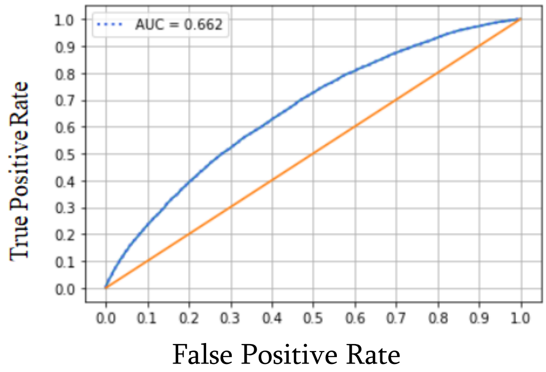


Fig. 5. AUC-ROC curve of model with 0.25% split data.

The second model is trained and evaluated using 80% training data and 20% testing data. The AUC - ROC curve of the second model with 0.20% split data is shown in Fig. 6.

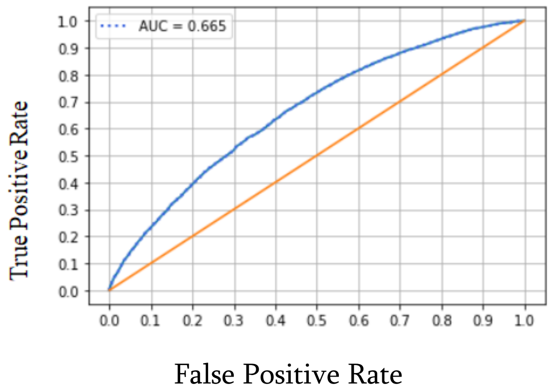


Fig. 6. AUC-ROC curve of model with 0.20% split data.

The third model is trained and evaluated using 90% training data and 10% testing data. The AUC-ROC curve of the third model with 0.10% split data is shown in Fig. 7.

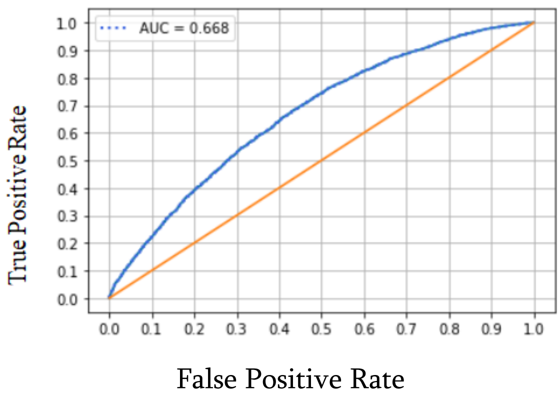
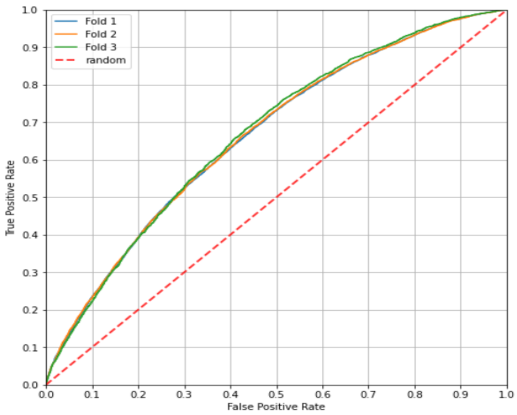


Fig. 7. AUC-ROC curve of model with 0.10% split data.

Comparison of the AUC - ROC curve between the models with different split data is shown in Fig. 8.

True Positive Rate



False Positive Rate

Fig. 8. Comparison of the ROC curve between the models with three different split data.

.

The other performance metrics such as precision, recall and f-measure obtained are shown in Table.1. The three models using the different data split criteria indicates that the models performance is stable even after different training and testing splits. This indicates that the classifier decision is certain. We can see that the estimated accuracy for the model was approximately 67%.

Table 1. The performance of CCP model

Classifier	Accuracy	Precision	Recall	F-measure	ROC -AUC
Initial model	60.14	58.81	66.01	62.20	64.28
Hyper parameter Tuned model	61.28	59.99	66.23	62.96	66.16
Model with 0.25% split data	61.52	60.13	67.65	63.67	66.40
Model with 0.20% split data	61.68	60.21	67.42	63.61	66.48
Model with 0.10% split data	62.27	60.80	68.85	64.58	66.81

V. CONCLUSION

A deep learning Multilayer Perceptron model is created for the Customer Churn Prediction having data certainty. Certainty is estimated using a multilayered Perceptron having five hidden layers in high distance zone. The number of hidden layers is increased so as to increase the feature space. The level of certainty of classifier decision is done by selecting features having the highest tendency to churn or nonchurn. The produced results are evaluated using a k-fold cross validation approach. In high distance zone the certainty of classifier decision is ensured by using this approach. Through this approach, we have extracted the level of certainty of classifier decision and also categories the customers into different customer groups based on lower zone and the upper zone. Also we evaluated the level of certainty of classifier before the classification of customers churn and non-churn. The performances of the resulting models were evaluated using the different evaluation metrics. In future, focus on more feature selection methods. Also by increasing the number of hidden layers, no of neurons in each layers, accuracy of the model can be improved.

VI. REFERENCES

[1]. Bingquan Huang, Mohand Tahar Kechadi, Brian Buckley, “Customer Churn Prediction in Telecommunication”, Expert Systems with Applications 39 (2012) 1414–1425.

[2]. Ionut Brandusoiu, Gavril Toderean, "Churn Prediction in Telecommunication Sector using Support Vector Machine", May 2013.

[3]. A.Keramatia, R.Jafari-Marandia, M.Aliannejadib, I.Ahmadianc, M. Mozaffaria, U. Abbasia Makhtar, “Improved churn prediction in telecommunication industry using data mining techniques”, Applied Soft Computing 24 (2014) 994–1012.

[4]. Mohammad Ridwan Ismail, Mohd Khalid Awang, Mohd Nordin Abdul Rahman, Mokhairi

Makhtar, "A multilayer Perceptron Approach for Customer Churn Prediction", Article in International Journal of Multimedia and Ubiquitous Engineering, July 2015.

[5]. Mohd Khalid Awang, Mokhairi Makhtar, Mohd Nordin Abdul Rahman, "Improving Accuracy and Performance of Customer Churn Prediction Using Feature Reduction Algorithms", Journal of Telecommunication, Electronic and Computer Engineering,Vol.9,2017.

[6]. Sanket Agrawal, Aditya Das, Sudhir Dhage, Amit Gaikwad, "Customer Churn Prediction Modelling Based on Behavioural Patterns Analysis using Deep Learning", International Conference on Smart Computing and Electronic Enterprise, 2018 IEEE.

[7]. Shrisha Bharadwaj, Adhiraj Pahargarh, Anil B.S., P S Gowra, Sharath Kumar, "Customer Churn Prediction in Mobile Networks using Logistic Regression and Multilayer Perceptron", Second International Conference on Green Computing and Internet of Things.

[8]. Irfan Ullah, Basit Raza, Ahmad Kamran Malik, Muhammad Imran, Saif Ul Islam, Sung Wonkim," A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector", 2019 IEEE Translation.

[9]. Adnan Amin, Feras Al-Obeidat, Babar Shah, Awais Adnan, Jonathan Loo, Sajid Anwar, " Customer churn prediction in telecommunication industry using data certainty", Journal of Business Research, 94 (2019) 290-301.

[10]. Adnan Amin, Babar Shah, Ali Abbas, Sajid Anwar, Omar Alfandi, Fernando Moreira, "Features Weight Estimation Using a Genetic Algorithm for Customer Churn Prediction in the Telecom Sector", Springer pp. 483-491, 2019.

Cite this article as :

V R Reji Raj, Rasheed Ahammed Azad .V, "Customer Churn Prediction in Telecommunication Industry Having Data Certainty", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 7 Issue 4, pp. 113-122, July-August 2020. Available at doi : https://doi.org/10.32628/IJSRSET207427
Journal URL : http://ijsrset.com/IJSRSET207427