International Journal of Scientific Research in Science, Engineering and Technology (www.ijsrset.com)
© 2020 IJSRSET | Volume 7 | Issue 1 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
DOI : https://doi.org/10.32628/IJSRSET207142

Customer Churn Prediction in Telecommunication Industry Using Data Certainty

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

ABSTRACT

Customer churn is a major problem affecting large companies, especially in telecommunication field. So the telecom industries have to take the necessary steps to retain their customers, to maintain their market value. So companies are seeking to develop methods that predict potential churned customers. We have to find out the factors that increase customer churn for making necessary actions to reduce churn. In the past, different data mining techniques have been used for predicting the churners. Here the most popular machine learning algorithms used for churn predicting are analysed. The conclusions are stated with the help of suitable tables.

Keywords : Regression, Linear Classifier, Random Forest, Decision Tree, CNN, Support Vector Machine, Naive Bayes, Neural networks, Multilayer Perceptron, and Performance metrics.

I. INTRODUCTION

The customers are the most important asset for a business. In companies the customers have numerous choices of service providers, so that they can easily switch the service or even the provider. Such customers are called churned customers. The reasons for customer churn can be due to dissatisfaction, higher cost, low quality, lack of features and privacy concerns. Many organizations e.g. financial service, airline ticketing services, social network analysis, online gaming, banking sector, telecommunication sector were focusing more on maintaining long term relationships with their existing customers. Loyal customers can be considered long term customers that are profitable for the company. CCP in Telecommunication is a well-known domain and popular research problem nowadays.

Companies in TCI have abundant information about their customers including local/international call records, short messages, voice mail, demographics, financial detail, and other usage behavior of the customers. A wide range of methods based on ensemble techniques, probabilistic methods, Support Vector Machine, K-nearest neighbor, Neural, have been developed to identify customers with the highest tendency to churn. These approaches help in deciding, measures from the telecommunication data to prevent their customers from churning by offering them promotions and better deals. However, these ML approaches lack the required effectiveness due to problem complexity. Hence, it is difficult to choose which ML or data mining technique can be chosen for building a CCP model. This is partly because most of the time, customer churn and non-churn have resembling features and behavior which increases the classification error rate. In other words, we can say that the classifier is uncertain about the decision and the level of certainty varies from case to case in the TCI.

Machine learning is a method of data analysis, that iteratively learn from data. Machine learning allows

systems to find hidden patterns without being explicitly programmed. The three types of machine learning approaches are 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 makes use of unlabeled data for training. Semi-supervised learning lies between unsupervised learning and supervised learning.

CUSTOMER CHURN

Churn means the behavior of customers who shift from one service provider to another. 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 and identify the reasons behind their churning. In this study, various machine learning techniques were analyzed for classifying customers data using the labeled datasets, to find which of the algorithm best classifies the customers into the churn and non-churn categories. Finding the churners can help the companies to retain their customers. 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. Because of various factors to be considered, that contribute to customer churn current methods for churn prediction are not effective and need to be improved. By removing some of the irrelevant features through reduction algorithms we can improve the prediction accuracy.

II. RELATED WORKS

Bingquan Huang, et.al [1] proposed a method for the customer churn prediction by introducing a new set of features from the existing. Modeling techniques used are Logistic Regression, Linear Classifications, Naive Bayes, Decision Trees, Multilayer Perceptron Neural Networks, Support Vector Machines and Evolutionary Data Mining Algorithm. Based on the new feature set, the existing feature sets and 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 set. 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. Apart from that, the proposed methodology's performance, measured by F-Score was equivalent to ANN that was the best interior classifier. 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 classification models based on three different classifiers, 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. 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 to identify the churned customers. It also provides the factors behind the churning in the telecom sector. 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 model produced better churn 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. From the results it is shown that the distance factor is strongly correlated with the certainty of the classifier and the classifier obtained high accuracy in the zone with greater distance

factor's value than those with a smaller distance factor's 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. METHODS AND MATERIAL

FEATURE SELECTION ALGORITHMS

Mostly, some of the features are significant and relevant. Hence selecting appropriate and small feature subsets of the original features not only helps to overcome the "curse of dimensionality" but also increases the performance and accuracy of the classification.

Correlation Based Feature Selection

It takes the individual predictive ability of each feature along with the degree of redundancy between them. CFS aims to discover a subset of features that are highly connected with the class having low inter-correlation. The main assumption of CFS is that the features are conditionally independent.

Information Gain

To select the test attribute of each node of the decision tree classification it uses the information gain measure. It determines the importance of an attribute by measuring the information gain relative to the class. The information gain ratio is a variation of the information gain, when choosing the significant attributes it reduces its bias by taking the number and size of branches into account.

CLASSIFICATION MODELS

Regression Analysis

Regression analysis is a popular statistical tool for the prediction of customers. The analysis will provide the relationship between the independent and dependent variables. Multiple regressions which consist of various input features or independent variables were used to build the relationship with respective dependent variables or output.

Logistic Regressions

Logistic regression is a widely used statistical modeling technique for discriminative probabilistic classification. Logistic regression estimates the probability of a certain event taking place. We determine the output for the input feature vector.

Naïve Bayes

Naive Bayes classifier calculates the probability that a given input sample belongs to a certain class. This method calculates a posteriori probabilities for each class. Naive Bayes assumes that the conditional probabilities of the independent variables are statistically independent. The Bayes decision rule is to assign a new record to that class which has a maximum posterior probability. An important assumption in the Naïve Bayes classifier is that the attributes used for describing customers are conditionally independent.

Decision Tree

In Decision Trees, ''divide and conquer'' method is applied to construct a binary tree. The method starts by searching an attribute having the best information gain at the root node and divide the tree into sub-trees. The sub-tree is then further separated recursively following the same rule. When the leaf node is reached or there is no information gain the partitioning stops. This learning algorithm constructs a tree with a training data set in which each node is

an attribute and branches of the nodes are corresponding attribute values.

Random Forest

The Random forest consists of a large number of individual decision trees that operate as an ensemble. Each tree in the random forest results in a class prediction and the class having the majority votes become our model's prediction. In Random Forest, the behavior of each tree is not too correlated with the behavior of any of the other trees in the model.

Support Vector Machines

A support vector machine is a supervised machine learning model that uses classification defined by a separating hyperplane. When a trained labeled data is given, the algorithm outputs an optimal hyperplane which categorizes new examples. A line dividing a plane into two parts where each class lay on either side is called a hyperplane.

K Nearest Neighbor

The k-nearest neighbor's algorithm that can be used to solve both classification and regression problems. KNN uses all the data for training while classification. It uses feature similarity to predict the values of new datapoints ie the new data point will be assigned a value based on how closely it matches the points in the training set.

Deep Learning

Deep learning is a type of artificial intelligence that has networks capable of learning unsupervised data that is unstructured or unlabeled. Deep learning models achieve very high accuracy. The term "deep" usually refers to the magnitude of hidden layers in the neural network. Deep learning can overcome the challenges of feature extraction because it is capable of learning to focus on the right features by themselves thus requiring little guidance from the programmer.

Artificial Neural Networks

Artificial Neural Network is an information processing mechanism composed of a large collection of simple neural units. The way of training or adjusting the strengths of these connections to achieve a desired overall behavior is known as learning process. Their most outstanding feature is their ability to learn automatically from available data to provide a means for predictions.

Multilayer Perceptron Neural Networks

A multilayer perceptron is a type of feed-forward artificial neural network. An MLP consists of three layers of nodes - an input layer, a hidden layer, an output layer. Each node is a neuron that uses a nonlinear activation function except input nodes. During training, it uses a supervised learning technique called backpropagation. A review of existing work on churn prediction models, features, methods, metrics are listed in Table 1.

Table-1 : Literature survey of models, features, methods, metrics used by various churn prediction systems

Year	Title	Features	Methods	Advantages	Disadvantages	Metrics
2011	Customer Churn Prediction in Telecommunication	Demography, Call details, Complaint information, user billing profile, customer account information	Linear Regression, Linear Classifier, Naïve Bayes, Multilayer Perceptron, SVM	The new proposed feature set is more effective in prediction.	More information is needed, imbalanced classification & dimension of input feature to be reduced.	Confusion Matrix, AUC
2013	Churn Prediction in Telecommunication Sector using Support Vector Machine	Call details, Usage data, Customer service details, plan details	Support Vector Machine using 4 different kernel functions.	The model using the polynomial kernel function performs well with 88.56 % accuracy.	SVM is used, more classifiers can be tried.	Confusion matrix, Gain measure

2014	Improved churn prediction in telecommunication industry using data mining techniques	Demography, usage data, user billing data, call details	Decision Tree, Artificial Neural Networks, KNN, SVM. A hybrid method, where all the above classifiers are used.	ANN outperforms the other three. The hybrid method shows higher accuracy in terms of precision & recall measure.	By f1-score it is equivalent to ANN.	Accuracy, Precision, recall, F-score
2015	A multilayer Perceptron Approach for Customer Churn Prediction	Demography, usage data, user billing data, customer relationship data	Neural Network, Logistic Regression, Multiple Regression	Neural Network outperforms with an accuracy of 91.28%.	Various factors of predictors need to be analyzed.	Accuracy, sensitivity, specificity
2017	Improving Accuracy & Performance of Customer Churn Prediction Using Feature Reduction Algorithm	Demography, usage data, user billing data, call details	Correlation-based feature selection, Information Gain Bayesian Network, Logistic Regression, Decision Table	Performance improves with feature reduction. A CFS feature reduction Decision Table gives the highest accuracy of 92.08 %.	Only 2 methods of feature reduction are used, more to be tried.	Accuracy, Root mean square error, time taken
2018	Customer churn prediction modeling based on behavioral patterns analysis using deep learning	Demography, usage data, call details	Multilayered Neural Network	The final accuracy is 80.03%. It also predicts churn factors.	Need more weight adjustment to improve accuracy.	Confusion Matrix, Accuracy
2018	Customer churn prediction in mobile networks using logistic regression & MLP	Demography, call duration, call details	Logistic Regression, Artificial Neural Network	Artificial Neural Network obtained an accuracy of 94.19%.	-	accuracy
2018	A churn prediction model using Random Forest: Analysis of ML techniques for churn prediction & factor identification in the telecom sector	Demography, user billing data, subscription data, duration	RF, J48, NB, ANN, MLP, k-means clustering, information gain, correlation attribute ranking filter	RF performs well	Focus more on feature selection & ensemble methods.	Accuracy, data correlation
2019	Customer churn prediction in the telecom industry using data certainty.	Demography, plan details, call details, user billing information	Naive Bayes	Also finding the certainty of the classifier.	Don't consider the imbalanced data.	Accuracy, Precision, recall, F-score
2019	Features Weight Estimation Using a Genetic Algorithm for Customer Churn Prediction in the Telecom Sector	Demography, plan details, call details	Feature weighting technique using Genetic algorithm used with Naive Bayes.	The proposed technique achieved 89.10% overall accuracy.	Never guarantee an optimal solution.	Accuracy, Precision, recall, F-score

IV.CONCLUSION

In this paper, we introduced the churn prediction problem and the significance of using predictive modeling methods to overcome the customer churn problem in the telecom industry. We examined the existing churn prediction methods in detail and summarized them. Finally, we examined the list of the commonly used metrics proposed in the literature for evaluating the performance of various churn prediction methods.

V. 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 Using Data Certainty", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 7 Issue 1, pp. 252-258, January-February 2020. Available at doi : https://doi.org/10.32628/IJSRSET207142
Journal URL : http://ijsrset.com/IJSRSET207142