An Enhanced Methodology to predict Customer Churn and Retention model using Machine Learning

Mr. V Naveen Kumar,
Assistant Professor,
Department of CS&E,
Bapatla Engineering College

Mr. G Ramesh Babu ,
Assistant Professor,
Department of CS&E,
Bapatla Engineering College

Mr. A Ravi Kishore,
Assistant Professor,
Department of CS&E,
Bapatla Engineering College

Abstract:

Telecom industry has seen a phenomenal growth throughout the globe in recent times. Today companies in this sector are putting their best efforts to retain their churning customers by satisfying them with offers and discounts. It's due to the fact that, acquiring a new customer is way expensive than retaining an existing one. Deep neural network learns on its own in a supervised manner and thus are often utilized in this regard efficiently. Number of hidden layers, epoch, number of neurons, hidden dropout ratio, input dropout ratio and activation function are varied to realize high sensitivity value. Sensitivity is that the percentage of churners who are correctly predicted as churning customers. Our model has achieved sensitivity of 85% and thus the results are satisfactory. Keeping customers satisfied is actually essential for saying that business is successful especially within the telecom. Many companies experience different techniques which will predict churn rates and help in designing effective plans for customer retention since the value of acquiring a replacement customer is far above the value of retaining the existing one. In this paper, 3 machine learning algorithms are accustomed predict churn namely, Naive Bayes, SVM and decision trees using two benchmark datasets IBM Watson dataset, which contains 7033 observations, 21 attributes and cell2cell dataset that contains 71,047 observations and 57 attributes. The models' performance has been measured by the world under the curve (AUC) and that they scored 0.82, 0.87, 0.77 respectively for IBM dataset and 0.98, 0.99, 0.98 respectively for cell2cell dataset. The proposed models also obtained better accuracy than the previous studies using an equivalent datasets.

Keywords:

Churn Prediction, Telecommunication, Modeling, Analysis, SVM, Naïve Bayes, Decision Trees, Telecom Churn Prediction, Deep Learning

Introduction:

Today numerous telecom companies are prompt everywhere the globe. The telecommunications sector has become one among the main industries in developed countries. Telecommunication market is facing a severe loss of revenue because of increasing competition among them and loss of potential customers [1]. Companies are working hard to survive during this competitive market counting on multiple strategies. in this telecommunication, churn may be a major issue where Churn is that

the activity of the telecommunication industry is that the customers leaving the present company and moving to a different telecom company. The technical progress and therefore the increasing number of operators raised the extent of competition. Many companies are finding the explanations of losing customers by measuring customer loyalty to regain the lost customers. to stay up with the competition and to accumulate as many purchasers , most operators invest an enormous amount of revenue to expand their business within the beginning [2]. within the telecmmunication industry each company provides the purchasers with huge incentives to draw in them to modify to their services, it's one among the explanations that customer churn may be a big problem within the industry nowadays. Churn management is extremely important for reducing churns as acquiring a replacement customer is costlier than retaining the prevailing ones [3]. to stop this, the corporate should know the explanations that the customer decides to maneuver on to a different telecom company. Prediction is that the only process to research the matter of churn activity. Customers churn prediction in telecom is usually a challenging task for the massive feature space and imbalanced nature of the dataset. Churn prediction is employed to acknowledge customers who are most probable to churn. Churn prediction and analysis can help a corporation to develop a sustainable strategy for customer retention programs [4]. Churn prediction and management became of great concern to the mobile operators. Mobile operators wish to retain their subscribers and satisfy their needs. Hence, they have to predict the possible churners then utilize the limited resources to retain those customers. the essential layer for predicting future customer churn is data from the past. we glance at data from customers that have already got churned (response) and their characteristics / behavior (predictors) before the churn happened. By fitting a statistical model that relates the predictors to the response, we'll attempt to predict the response for existing customers. This method belongs to the supervised learning category, just just in case you needed another buzzing expression [5]. The minority class comprises of fewer instances within the whole dataset that results in biased training of classifiers thanks to dominating presence of majority class instances. The telecom companies also acquire many information about customers including billing, payments, call records, demographics, etc. which turns sizable amount of features to account for within the dataset. Consequently, the predictors suffer from the curse of dimensionality and imbalanced distribution of the telecom dataset for predicting churners. Thus, a churn prediction method is very desirable which will affectively mitigate the imbalanced nature and high dimensionality issues present within the telecom dataset [6]. Churn occurs mainly due to customer dissatisfaction. Identifying customer dissatisfaction requires several parameters. A customer usually does not churn due to a single dissatisfaction scenario [7]. There usually exist several dissatisfaction cases before a customer completely ceases to do transactions with an organization. Several properties associated with the customer and

their mode of operations with the organization are recorded by the organizations. This represents the customer‟s behavior data. Analyzing this data would present a clear view of the customer‟s current status [8]. Hence this can be used as the base data for churn prediction. The major difficulty arising from this mode of operation is that the data under discussion tends to be very huge. The hugeness can be attributed to the behavioral nature of the data, depicting all the product lines dealt with by the organization. Further, due to the requirement of structural representation of the data, all the instances are bound to contain all the properties corresponding to a generic customer in the organization [9]. This leads to data sparseness, since customers will be associated with only a few properties and not all the properties pertaining to the organization. The hugeness of data and sparsity acts as the major difficulties in the process of churn prediction. Many research confirmed that machine learning technology is highly efficient to predict this situation. This technique is applied through learning from previous data. Machine learning is subset of artificial intelligence which enables computers to learn (i.e. improvise) from data without any intervention. Machine learning is often coined with the terms pattern recognition and computational learning theory. Machine learning involves constructing algorithms that can learn from data available and can be used to make predictions on data. Here a model is built from the given input data which is then used to make predictions on new data. As the data being collected is drastically increasing each day, this calls for the need of machine learning. In machine learning, the learning process is extracting knowledge from the given data thus there is no need for a human to specify any kind of knowledge [10]. Because of the hierarchy of concepts, complex concepts can be learned by building them from simpler ones thus reducing the complexity. Prediction of churn customer is analyzed and implemented using machine learning approach. Large companies interact with their customers to provide a variety of services to them [11]. Customer service is one of the key differentiators for companies. The ability to predict if a customer will leave in order to intervene at the right time can be essential for pre-empting problems and providing high level of customer service. The problem becomes more complex as customer behavior data is sequential and can be very diverse. Churn is an unavoidable process in any industry. However, though difficult, it is possible to identify the causes of churn using Machine learning approaches. Having a good churn prediction model becomes extremely useful in order to minimize the churn rate because tailored promotions can be offered to specific customers that are not satisfied. Thus, there is no formal notification from the customer of ending a contract term. Our goal is to predict the customer churn. The sooner these changing patterns are detected the more opportunities and time the company will have to retain the customer. For this accuracy Area Under Curve (AUC) is used in classification analysis in order to determine which of the used models predicts the classes best. Although machine learning algorithms are

usually designed to improve accuracy by reducing error, not all of them take into account the class balance, and that may give bad results. In general, classes are considered to be balanced in order to be given the same importance in training.

Literature Survey

Prediction of telecom churners has always been an area of researcher's interest and hence many researchers have worked in this direction to predict telecom customer churn.

Hung, Yen and Wang [11] applied data mining techniques like neural network anddecision tree to predict telco churn. They found out that both decision tree and neural network data mining techniques can predict churning customers accurately using various customer details like demographics,billing information, call details and many more. This paper also showed the effect of inadequate data on model building. Due to unavailability of customer demographics Wei and Chiu [12] proposed a technique to identify churners form customers call pattern and information about contracts obtained from customers call details. In this paper multi classifier class combiner approach to solve the problem of skewed distribution of customers as churners and non-churners. Results suggest that the technique provided satisfactory results, as compared to other demographic based churn prediction systems.

T. Vafeiadis, K.I. Diamantaras, G. Sarigiannidis and K.Ch. Chatzisavvas[13] compared the performance of widely used machine learning algorithms including Artificial Neural Networks, Decision Trees, Support Vector Machines, Naïve Bayes classifiers, and Logistic Regression classifiers, then they applied boosting techniques and compared the performance of the boosted versions. The classifier which performed best among the models was the SVM-POLY usingAdaBoost with accuracy of about 97%.

Chuanqi Wang, Ruiqi Li, Peng Wang andZonghai Chen [14] discussed that customer churn prediction is cost sensitive problem. In contrast to most of papers which considers each misclassification same, this paper presents that misclassification cost of each sample is different. A partition cost-sensitive CART model is proposed in this paper in order to minimize the cost of churn prediction. ArnoDeCaigny, KristofCoussement and Koen W.De Bock [15] proposed a new composite algorithm built on top oflogistic regressionanddecision treesfor predicting customer churn. They have discussed that decision trees and logistic regression have issues likein handlingvariables that are interacting and many more. The logit leaf model (LLM) is a new algorithm to classify data in an effective manner. The LLM model combines the power of both the base algorithms as first decision tree is used to create segments and then for each leaf model is built. The results that were obtained were better than many advanced methods.

Kiran Dahiya and Surbhi Bhatia [16] proposed a new framework for the customer churn prediction model and which is implemented using WEKA Data Mining software. This paper also compares the performance of decision tree and logistic

regression using framework. The results obtained showed that decision tree had much higher accuracy than the logistic regression technique.

Y Huang, F Zhu, M Yuan, K Deng, Y Li, B Ni, W Dai, Q Yang and J Zeng[17] showed how telco big data make prediction churners easier. They have showed that large volume of data, large variety of features and large incoming data improves performance of churn prediction. The proposed system was deployed in one of the biggest mobile operators in china. The system provided prepaid customers who are likely to churn in the next month, having 0.96 precision for the top 50000 predicted churners.

K Kim, C Jun and J Lee [18] proposed a procedure that predicts customer churn by examining communication patterns among customers. The subscribers of telecom companies are connected with other customers, and network properties may affect the churning customers. Hence it makes the use of social network in predicting the churning customers. This paper clearly showed that there was improvement in predicting churners by introducing network analysis as compared to techniques taking personal customer information into consideration's.

Verbeke, D Martens, C Mues and B Baesens [19] discussed about two advanced data mining techniques which are AntMiner+ and ALBA. They had used these two because: AntMiner+ is high performing technique which allows to incorporate domain knowledge. ALBA brings about high accuracy prediction of a non-linear SVM. The results showed that use of ALBA results in increased performance by improving learning of classification techniques and AntMiner+ resulted in accurate and comprehensible model.

Y Zhang, R Liang, Y Li, Y Zheng and M Berry[20] presented prediction system based on behaviour to predict telecom churn. This system derives attributes from services that the customer avails. Hence only usage pattern is taken into consideration to make prediction, eliminating the problems like feature selection, missing values and many more. Also this technique was free from the problems that traditional systems have to face like correlations among inputs.

Related Work

The method used in this paper has been summarized in Figure 1 and it has been explained in detail in the next paragraphs.

Datasets Visualization

There are two datasets used in this study. The first dataset consists of 7034 samples and 20 attributes while the second dataset contains 71,047 samples and 57 attributes. Datasets details are as shown in Table 1. Both datasets have been visualized using Orange.

The samples from IBM dataset shown in Table 2 are the features which have been used in prediction models.

Figure 1. The research strategy.

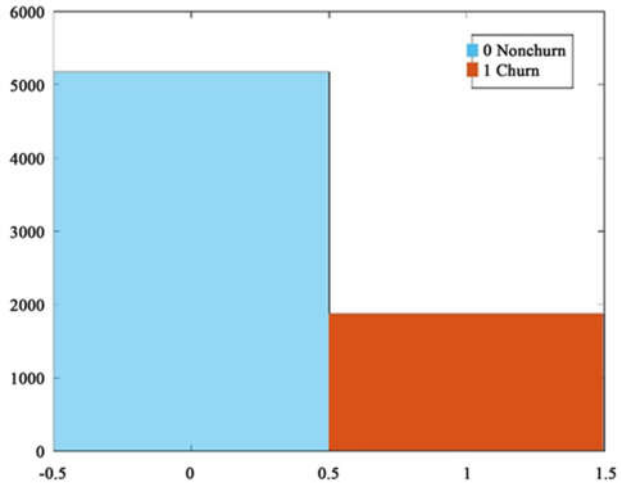
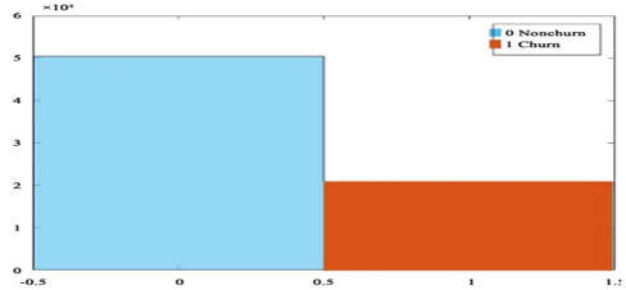


Table 1. Datasets used.



IBM Dataset Visualization and Preprocessing

The dataset is for customers who left within the last month. The column is called Churn where it contains the below attributes [14] :

• Services that each customer has signed up, internet, online security, online backup, device protection, tech support, and streaming TV and movies;

• Customer account information how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges;

• Demographic info about customers—gender, age range, and if they have partners and dependents.

Figures 2-11 show the attributes and their distributions according to the churn class where orange color indicates the churn customers and the blue for non-churn. As noticed that:

• Most churned customers have internet service type Fiber optic;

• They use paperless billing;

• Most of the customers were dependents;

• Their payment method was electronic check;

• They don't use "device protection" or "online backup" services, rather they use phone service;

• Their tenure was less than 14 months.

Cell2cell Dataset Visualization and Preprocessing

Cell2cell is the 6th largest wireless company in the US, Cell2cell dataset consists of 71,047 signifying whether the customer had left the company two months after observation and 57 attributes [17]. The histograms Figures 13-17 show the attributes and their distributions according to the churn in similar way as done with IBM dataset visualization. What has been noticed on cell2cell dataset that:

• Churned customers have average (mean) monthly minutes of use which is less than 530' minute;

• They have service for only 11 - 15 months;

• The numbers of days for their equipment were between 300 & 361 day;

• The numbers of models issues are less than 2;

• Their prizm code refer to town;

• Their handsets have web capability.

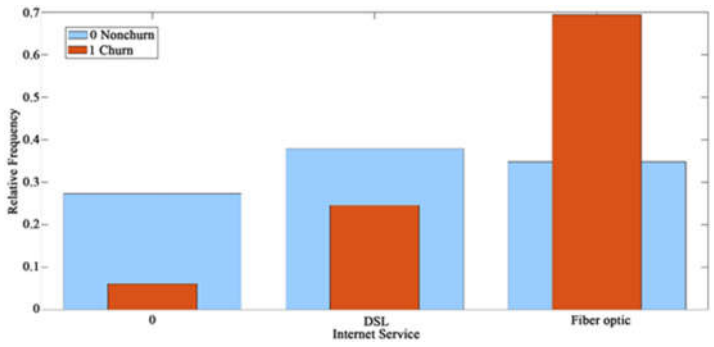


Figure 2. Internet service.

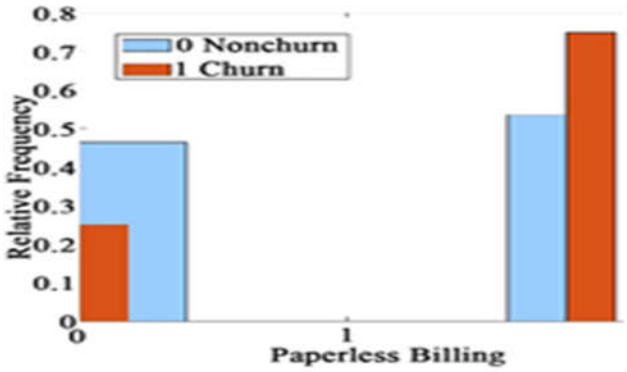


Figure 4. Paperless billing.

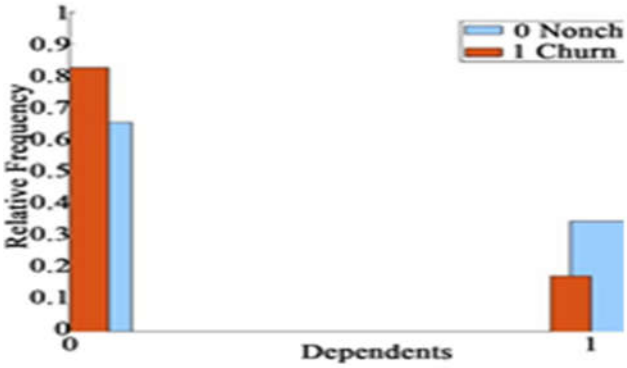


Figure 5. Dependents.

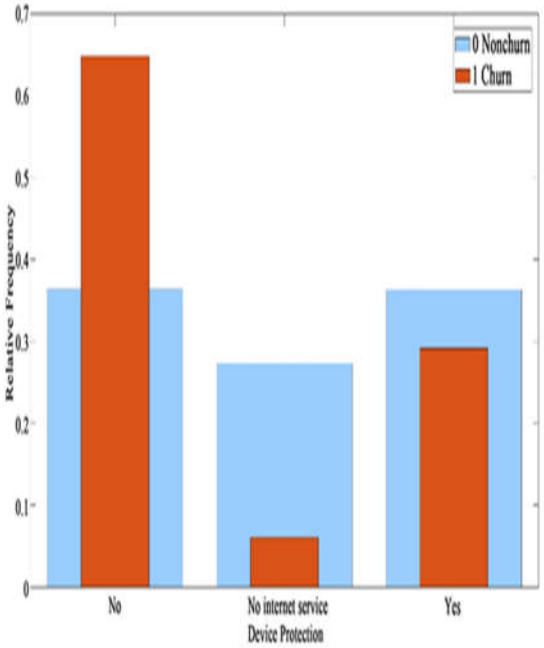


Figure 6. Payment method.

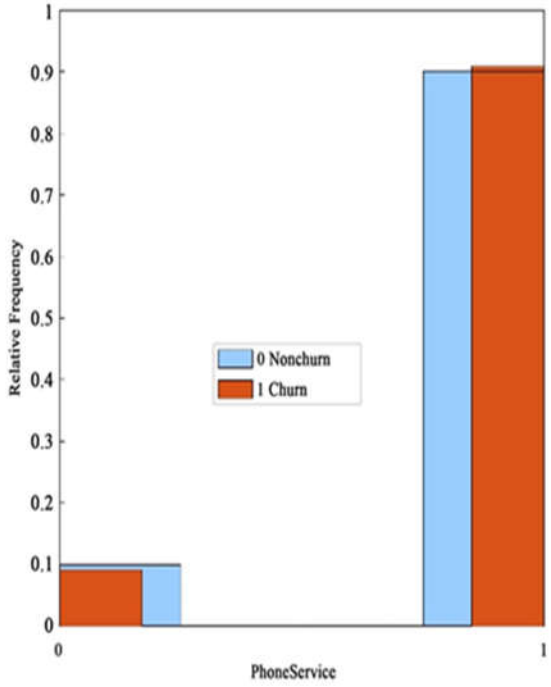


Figure 7. Phone service.

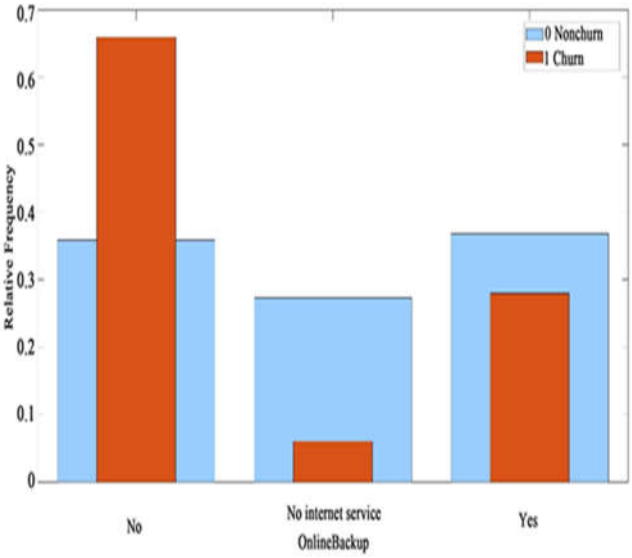


Figure 8. Monthly charges & total charges.

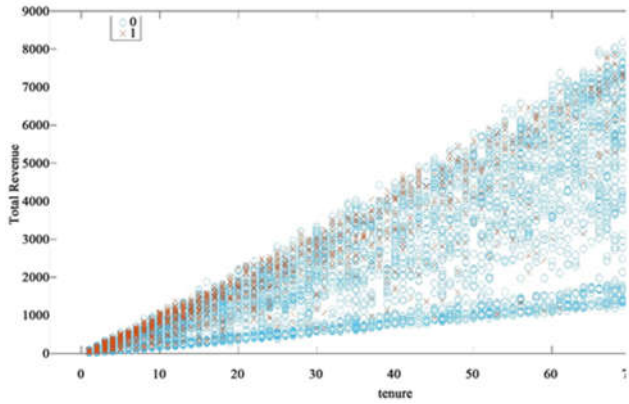


Figure 9. Tenure & total charges.

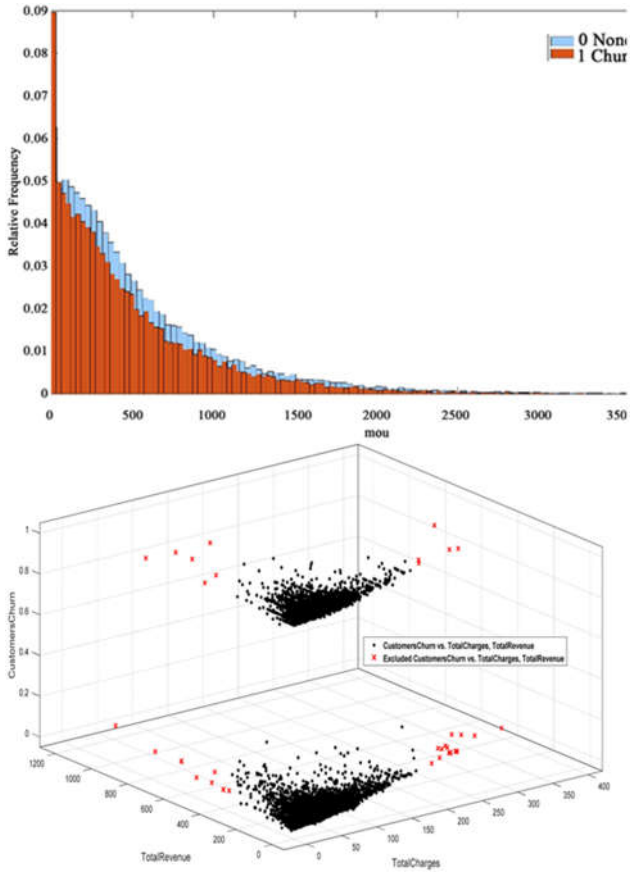


Figure 10. Churn vs. total revenue, total charge.

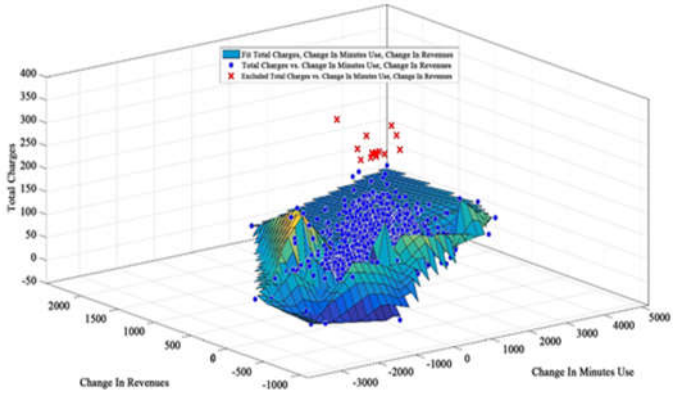


Figure 11. Total charges vs. change in miute use, change in revenues.

Support Vector Machine Algorithm

SVM algorithm for the classification of both linear and nonlinear data. It transforms the original data into a higher dimension, from where it can find a hyperplane for data separation using essential training tuples called support vectors [19]. The SVM binary classification algorithm searches for an optimal hyperplane that separates the data into two classes. For separable classes, the optimal hyperplane maximizes a margin (space that does not contain any observations) surrounding itself, which creates boundaries for the positive and negative classes. The data for training is a set of points (vectors) x_j along with their categories y_j. For some dimension d, the xj∈Rdxj∈Rd, and the $y_j = \pm 1$. The equation of a hyperplane is [20]

f(x)=x'β+b=0f(x)=x′β+b=0(3)

where β∈Rdβ∈Rd and b is a real number.

As the data used is not allow for a separating hyperplane, the SVM used a soft margin, meaning a hyperplane that separates many, but not all data points. There are two standard formulations of soft margins. Both involve adding slack variables ξ=(ξ1,ξ2,⋯,ξN)ξ=(ξ1,ξ2,⋯,ξN) and a penalty parameter C.

• The L^1-norm problem is:

minβ,b,ξ(12β'β+C∑jξj)minβ,b,ξ(12β′β+C∑jξj)(4)

such that
yjf(xj)≥1−ξjξj≥0yjf(xj)≥1−ξjξj≥0(5)
• The L^2-norm problem is:
minβ,b,ξ(12β′β+C∑jξj2)minβ,b,ξ(12β′β+C∑jξj2)(6)
In these formulations, it can be used C places more weight on the slack variables ξj, meaning the optimization attempts to make a stricter separation between classes. Equivalently, reducing C towards 0 makes misclassification less important.
The propsed SVM model standardizes the predictors using their corresponding weighted means and weighted standard deviations. Means it standardizes predictor j (x_j) using
xj∗=xj−μj∗σjxj∗=xj−μj∗σj(7)
μj∗=1∑kwk∗∑kwk∗xjkμj∗=1∑kwk∗∑kwk∗xjk(8)
X_{jk} is observation k (row) of predictor j (column).
(σj∗)2=v1v12−v2∑kwk∗(xjk−μj∗)2(σj∗)2=v1v12−v2∑kwk∗(xjk−μj∗)2(9)
v1=∑jwj∗v1=∑jwj∗(10)
v2=∑j(wj∗)2v2=∑j(wj∗)2(11)

Decision Tree Algorithm

Decision tree induction is the learning of decision trees from class-labeled training tuples. A decision tree is a flowchart-like tree structure, where each internal node (nonleaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label. The topmost node in a tree is the root node [19]. The Classification Tree splits nodes based on either impurity or node error. Impurity means one of several things, depending on the Split Criterion name-value pair argument:
• Gini's Diversity Index (gdi)—the Gini index of a node is
Gini(D)=1−∑i=1mP2i,Gini(D)=1−∑i=1mPi2,(12)
where the sum is over the classes i at the node, and p(i) is the observed fraction of classes with class i that reach the node. A node with just one class (a pure node) has Gini index 0; otherwise the Gini index is positive. So the Gini index is a measure of node impurity.
• Deviance ("deviance")—with p(i) defined the same as for the Gini index, the deviance of a node is
∑ip(i)log2p(i)∑ip(i)log2p(i)(13)
A pure node has deviance 0; otherwise, the deviance is positive.

Experiments and Result Analysis

For the experiment we have taken dataset containing 30 predictor variables that are: gender, Senior Citizen, Partner, Dependents, tenure, Phone Service, Paperless Billing, Monthly Charges, Total Charges, Multiple Lines No phone service, Multiple Lines Yes, Internet Service Fiber optic, Internet Service No, Online Security No internet service, Online Security Yes, Online Backup No internet service, Online Backup Yes, Device Protection No internet service, Device Protection Yes, Tech Support No internet service, Tech Support Yes, Streaming TV No internet service, Streaming TV Yes, Streaming Movies No internet service, Streaming Movies Yes, Contract One year, Contract Two year, Payment Method Credit card (automatic), Payment Method Electronic check, Payment Method Mailed check and 1 variable to be predicted i.e. Churn (having 2 possible values: yes or no).In total the dataset has 4914 rows which is divided into training set and test set as 70% and 30%.In the experiment we have chosen a base model. In order to check the models, they have been compared with previous papers which used similar datasets. ApurvaSree, G., et al. [4] and Induja, S. &

D. V. P. Eswaramurthy [21] used IBM Waston dataset with different algorithms including SVM for the first paper & Naïve Bayes for the second mentioned paper, both results are similar to the results obtained from this paper. However, our proposed method obtained higher accuracy by using SVM model on IBM dataset with k-fold partition, k value = 10 to produce an area under curve reached to 0.86548. As for cell2cell dataset, the papers in [21] [22] [23] [24] also used different algorithms including SVM where the best accuracy for previous studies was 94.13 for AUC whereas the AUC in the proposed model using SVM is 0.99 .

Conclusion

Hence we have obtained the optimal model for prediction of telecom churn by varying the parameters. It is taken care that the model does not over fit the training data and also has a good sensitivity value (i.e. model is generalized) on the test dataset. Hence we have obtained satisfactory results as compared to traditional methods for predicting customer churn. Telecom industry has seen a phenomenal growth throughout the world in recent times. Today companies in this sector are putting their best efforts to retain their churning customers by satisfying them with offers and discounts. It is due to the fact that, acquiring a new customer is far more expensive than retaining an existing one. The AUC, which obtained using SVM algorithm, is better compared with the previous papers. As noticed that the churned customers have some similar services, which means that any telecom company can detect the predictors and retain their customers. The paper concluded that telecom operators can get best predictive models if they analyzed their whole records and tracked the customers' behavior so they can build different marketing approaches to retain the churners based on the predictors which can be detected when analyzing the historical customer's records. All churn prediction models in this paper can be used in other customer response models as well, such as cross-selling, up-selling, or customer acquisition.

References

1. Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. IBM Journal of research and development, 3(3), 210-229. 1959
2. KB, S. K., Krishna, G., Bhalaji, N., &Chithra, S. (2018). BCI cinematics–A pre-release analyser for movies using H2O deep learning platform. Computers & Electrical Engineering. 2018
3. Goodfellow, I., Bengio, Y., Courville, A., &Bengio, Y. (2016). Deep learning (Vol. 1). Cambridge: MIT press. 2016
4. Tompson, J. J., Jain, A., LeCun, Y., &Bregler, C. (2014). Joint training of a convolutional network and a graphical model for human pose estimation. In Advances in neural information processing systems (pp. 1799-1807).2014
5. Zhang, Y., Chan, W., &Jaitly, N. (2017, March). Very deep convolutional networks for end-to-

end speech recognition. In Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on (pp. 4845-4849). IEEE. March 2017

6. Prashanth, R., Deepak, K., &Meher, A. K. (2017, July). High accuracy predictive modelling for customer churn prediction in telecom industry. In International Conference on Machine Learning and Data Mining in Pattern Recognition (pp. 391-402). Springer, Cham. July 2017
7. Araque, O., Corcuera-Platas, I., Sanchez-Rada, J. F., & Iglesias, C. A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Systems with Applications, 77, 236-246. 2017
8. Young, T., Hazarika, D., Poria, S., & Cambria, E. (2017). Recent trends in deep learning based natural language processing. arXiv preprint arXiv:1708.02709. 2017
9. Wang, Y., & Xu, W. (2018). Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. Decision Support Systems, 105, 87-95. 2018
10. Ahmad, A.K., Jafar, A. and Aljoumaa, K. (2019) Customer Churn Prediction in Telecom Using Machine Learning in Big Data Platform. Journal of Big Data, 6, 28. Andrews, R., et al. (2019) Churn Prediction in Telecom Sector Using Machine Learning. International Journal of Information Systems and Computer Sciences, 8, 132-134.
11. ApurvaSree, G., et al. (2019) Churn Prediction in Telecom Using Classification Algorithms. International Journal of Scientific Research and Engineering Development, 5, 19-28.
12. Tata Tele Business Services (2018) Big Data and the Telecom Industry
13. Kayaalp, F. (2017) Review of Customer Churn Analysis Studies in Telecommunications Industry. Karaelmas Science Engineering Journal, 7, 696-705.
14. Umayaparvathi, V. and Iyakutti, K. (2016) A Survey on Customer Churn Prediction in Telecom Industry: Datasets, Methods and Metrics. International Research Journal of Engineering and Technology, 3, 1065-1070.
15. Kaur, S. (2017) Literature Review of Data Mining Techniques in Customer Churn Prediction for Telecommunications Industry. Journal of Applied Technology and Innovation, 1, 28-40.
16. Ahmed, A. and Linen, D.M. (2017) A Review and Analysis of Churn Prediction Methods for Customer Retention in Telecom Industries. 4th International Conference on Advanced Computing and Communication Systems, Coimbatore, 6-7 January 2017, 1-7.
17. Amin, A., et al. (2019) Customer Churn Prediction in Telecommunication Industry Using Data Certainty. Journal of Business Research, 94, 290-301.

18. Gerpott TJ, Rams W, Schindler A. Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. Telecommun Policy. 2001;25:249-69.
19. Wei CP, Chiu IT. Turning telecommunications call details to churn prediction: a data mining approach. Expert Syst Appl. 2002;23(2)1 03-12.
20. Qureshii SA, Rehman AS, Qamar AM, Kamal A, Rehman A. Telecommunication subscribers' churn prediction model using machine learning. In: Eighth international conference on digital information management. 2013. p. 131-6.
21. Ascarza E, lyengar R, Schleicher M. The perils of proactive churn prevention using plan recommendations: evidence from a field experiment. J Market Res. 2016;53(11:46-60. S. Bott. Predicting customer churn in telecom industry using multilayer preceptron neural networks: modeling and analysis. Igarss. 2014;1 1(1):1-5.
22. Umayaparvathi V, lyakutti K. A survey on customer churn prediction in telecom industry: datasets, methods and metric. Int Res J Eng Technol. 2016;3(4):1065-70.
23. Elisabetta E, Meyerhenke H, Staudt CL. Approximating betweenness centrality in large evolving networks. CoRR. 2014. arxiv:1409.6241.
24. Brandusoiu I, Toderean G, Ha B. Methods for churn prediction in the prepaid mobile telecommunications indus-try. In: International conference on communications. 2016. p. 97-100.
25. He Y, He Z, Zhang D. A study on prediction of customer churn in fixed communication network based on data mining. In: Sixth international conference on fuzzy systems and knowledge discovery, vol. 1. 2009. p. 92-4.
26. Idris A, Khan A, Lee YS. Genetic programming and adaboosting based churn prediction for telecom. In: IEEE international conference on systems, man, and cybernetics (SMC). 2012. p. 1328-32.
27. Huang F, Zhu M, Yuan K, Deng EO. Telco churn prediction with big data. In: ACM SIGMOD international confer-ence on management of data. 2015. p .607-18.
28. Makhtar M, Nafis 5, Mohamed M, Awang M, Rahman M, Deris M. Churn classification model for local telecom-munication company based on rough set theory.