Proceedings of the Third International Conference on Intelligent Sustainable Systems [ICISS 2020]
IEEE Xplore Part Number: CFP20M19-ART; ISBN: 978-1-7281-7089-3

2020 3rd International Conference on Intelligent Sustainable Systems (ICISS) | 978-1-7281-7089-3/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICISS49785.2020.9315951

Machine Learning Based Telecom-Customer Churn Prediction

Pushkar Bhuse
Department of Information Technology
Dwarkadas J. Sanghvi College of Engineering
Mumbai, India
pushkar.s.bhuse@gmail.com

Aayushi Gandhi
Department of Information Technology
Dwarkadas J. Sanghvi College of Engineering
Mumbai, India
aayushigandhi23@gmail.com

Parth Meswani
Department of Information Technology
Dwarkadas J. Sanghvi College of Engineering
Mumbai, India
parthmeswani29@gmail.com

Riya Muni
Department of Information Technology
Dwarkadas J. Sanghvi College of Engineering
Mumbai, India
riya.muni1999@gmail.com

Neha Katre
Department of Information Technology
Dwarkadas J. Sanghvi College of Engineering
Mumbai, India
neha.mendjoge@djsce.ac.in

***Abstract*— Customer churn or attrition refers to the percentage of customers who will discontinue with a company's service during a given timeframe. Churn rate is calculated by dividing the number of customers a company lost over a given period of time by the number of retained customers at the beginning of that time period. Churn prediction is a key predictor of the long term success or failure of a Business. In this research, machine learning and deep learning techniques are explored in order to predict telecom customer churn. Ubiquitous techniques like Random Forest Classifiers and SVMs are compared with relatively newer architectures like XGBoost and Deep Neural networks to classify if a customer will churn or not. The efficiency of these models are further explored by passing them through a grid search. From this experiment, it could be inferred the Random Forest model works best for this particular use case with a prediction accuracy of 90.96% on the testing data before grid searh.**

Keywords— Customer Churn, Machine Learning, Deep Learning

I. INTRODUCTION

Over the years, the telecommunications industry has emerged as one of the world's most rapidly growing industries, having its impact on about 90% of the global population. and has scaled dramatically in recent years. It is one of the sectors where the customer is of utmost importance, and understandably, customer satisfaction played a huge role in the success of organisations in this industry. Due to the widespread nature of this industry, consumers today have a plethora of organisations from which they can choose to receive services. The decision of consumers to choose a particular service provider depends on cost, flexibility and customizability of service. To meet these needs, telecom companies strive to develop policies and services to attract customers and try to achieve market dominance. Hence there is an increasing need to predict the potential churners before they actually leave a provider so that the retention strategies could be targeted upon them and the organization may burgeon by overall revenue maximization. Owing to heavy competition from rival companies due to the competitive rates of various providers, customers often tend to switch between them. Also, it is pretty common to hear people express frustration regarding convoluted billing, unwanted marketing emails, hard-to-navigate customer service, high plan prices, etc. All these factors combined today make the necessity of effective customer churn identification in the telecommunications industry of paramount importance.

Churners are customers who will be switching from one telecom service provider to another. Prediction of telecom churners has been an area of interest for researchers and many researchers have worked on various techniques to predict telecom customer churn. Telecom industry has been battling the threat of losing more than 25% of its customers every year, which is believed to result in huge revenue loss[1]. Another known fact is that adding or acquiring a new customer costs between 5 to 10 times more than retaining an old customer with the company[2]. Hence it is believed that the best marketing strategy is to retain the existing subscribers or to avoid customer churn[3]. Over the years, there has been an increasing need to automate the process of identifying customer churn. This process has become such an expensive affair that generally only 15 percent of revenues earned by mobile companies are spent on network infrastructure and IT while 15 to 20 percent of revenues on the acquisition and retention of customers [4]. AI has been really successful in being able to counter this problem by considerably comprehending large amounts of customer data to draw valuable inferences while reducing the workforce this job demands.

In this paper, a comparative analysis is performed on techniques such as Random Forest, SVM, Extreme Gradient Boosting (XGBoost), Ridge classifier and Neural Networks to predict various customer churn patterns. Through Random Forest, we explore the older but trusted and lightweight divide and conquer methodology making decision trees of numeral type and and using them through a random selection of attributes. Finally, a decision tree is created for classification on test data. Random forest, not only performs well on large datasets but also handles missing variables efficiently[5]. A final decision tree is constructed for prediction of the test dataset. Random forest performs well on a large dataset and handles missing variables without deletion of variables[5]. Support Vector Machine (SVM) models are parametric. Its accuracy and performance is greatly influenced by the initial values of its parameters. Therefore, while tuning the parameters of SVM, a new combined evaluation metric is applied to maximize its effectiveness for churn management[6]. Another recent prediction algorithm as a part of the ensemble method in many machine learning challenges is Extreme Gradient Boosting (XGBoost), which generally performs well with imbalanced-classes data[7]. The XGBoost algorithm uses the exact greedy algorithm to find the best split. Through these algorithms, an attempt is made to direct efforts for the customer churn prediction process onto lightweight models as well. By doing so, larger amounts of data can be processed without the extensive load of computation posed by newer Deep Learning Architectures without harming performance.

The paper has been divided into the following sections. Section II discusses the current work and existing research, where data has been surveyed from different sources. Section III explains about the dataset, and how it is pre-processed. Section IV discusses the different algorithms and methodologies that are being used in this project. Section V and VI displays our findings and conclusions, and Section VII mentions the future scope of our work.

II. Existing work

Ensemble based classifiers[8] like bagging, boosting and random forest were compared with Naive Bayes Classifier[8], Decision Tree[8] and Support Vector Machine(SVM)[8] , the result of which yielded that the ensemble based classifiers had less error rate, low specificity, high sensitivity and greater accuracy as compared to the others. SVM models are parametric and the parameters' initial values have great influence on their accuracy and performance.

Deep neural networks [9] are also used to build a model that fits the data using various hierarchies of concepts to increase the performance of models built. The model has been trained using k-fold cross validation technique [9].

Another idea mainly focuses on customer churn prediction models for identifying the key factors which are crucial and which cause the churn. The set of techniques that are used to do the same, include Logistic regression(LGR), decision tree analysis and artificial neural network. There has also been an amalgamation of models using artificial neural networks that has produced results with a probability of 94% [10][11]. Feature selection is performed by using information gain and correlation attribute ranking filter. [11]

In [12], they investigated and proposed two different feature reduction algorithms which are Correlation based Feature Selection (CFS) and Information Gain (IG) and built classification models based on three different classifiers, namely Bayes Net, Simple Logistic and Decision Table. Experimental results demonstrate that the performance of classifiers improves with the application of features reduction of the customer churn data set. [12]

A churn score is assigned to each customer to detect any signs of churn. A suite of supervised learning algorithms are fed the final feature set which is determined by applying a brute force approach to feature engineering, using feature selection to identify the set of relevant attributes[13]. It was found that a combination of clustering and classification techniques perform better when compared to single clustering and classification data mining techniques.[10]

There has also been significant use of churn prediction to find the retention of social media[14] users by combining the social network and call log information by generating a set of influential users who might influence each other. A thorough study has been conducted on the existing systems, and flaws of each model were identified and after analysis, the best or most suitable fit for the dataset has been chosen.

III. Dataset and preprocessing

The dataset this paper works on pertains to the telecom customer churn, focused on a customer retention program. Each row represents a customer and each column contains the customer's attributes. The 'churn' column depicts the data of customers who relinquished the company's services one month back.

It also includes services a given customer opted for like number of cellular devices, number of lines, internet connectivity, security, backup plans, insurance from damages, customer care services, and streaming TV and movie services. In addition, it also includes information about a customer's account - the time for which they have accessed the company's services, their usag plan, chosen method of payment, billing service (papered or paperless), monthly charges and total charges; and finally the dataset has columns pertaining to data regarding the demographic of the customers vis-à-vis gender, age, their marital status and number of children.

Feature selection is performed on the dataset and a few columns are dropped. The attribute Customer_id is dropped because it has no effect on the indication of churn, similarly the 'gender' column is dropped because the churn value for male and female is approximately the same. Phone Service and Multiple Line do not give any Information regarding Churn hence they too are dropped. There are two types of customers - non churn, the ones who remain loyal to the telecom company and the second type is churn. The proposed model aims to predict if a customer is likely to retain services of a company or not. The telecom dataset used has no noise i.e. no null values. The categorical data is converted into numeric data and MinMax scaler is used to normalise the data.

The data is then upsampled and is split into dependent and independent variables; ready to be trained and tested.

The dataset is divided into a train set and a test set. These sets are in a 80:20 ratio with a total of 206950 rows in the train set and a total of 51750 in the test set.

IV. METHODOLOGY

A. Algorithms

1) Ridge classifier: The Ridge Classifier, based on Ridge regression method, converts the label data into [-1, 1] and solves the problem with regression method. The highest value in prediction is accepted as a target class and for multiclass data multi-output regression is applied. L2 regularization solves the overfitting issue.

$$min(||Y - X(\theta)||_2^{\,2} + \gamma||\theta||_2^{\,2}) \qquad (1)$$

Given above is the cost function for ridge regression, which can be controlled by changing the alpha parameter (λ). It is a convex objective function with a global minima. Note that λ=0 gives the ordinary least squares solution (no regularization).

2) Random Forest: Training is performed on 80% of the dataset with 100 estimators, i.e. it builds 100 trees and then takes the average of the predictions. The maximum depth of each decision tree is taken as 10. In this method, important features are ranked and predicted based on the voting done on the decision trees. Thus we get high accuracy with Random Forest.

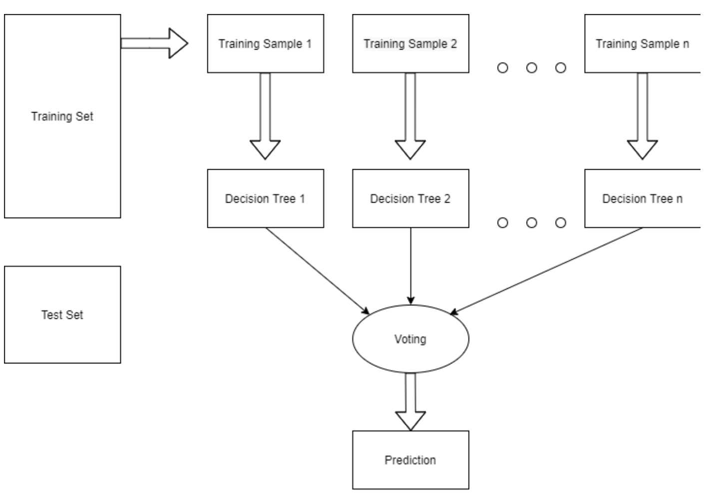

Fig. 1. The decision tree algorithm for classification

3) XG Boost: The XGBoost library implemented in Python is selected as one of the classifiers. XGBoost stands for Extreme Gradient Boosting, which is a combination between gradient descent and boosting. Boosting is a supervised and ensemble-learning algorithm that works by assigning different weights for training data distribution for each iteration. The XGBoost algorithm uses the exact greedy algorithm to find the best split. Due to its good cache optimization, the XG boost classifier gives good results in the prediction model but it takes more training time for the iteration process. It has good execution speed and model performance. GBM uses second order gradient statistics to minimize following regularized objectives that shown in equation (2).

$$L(\Phi)=\sum_i l\,(\widehat{yi}, yi)+\sum_k \Omega(fk), \qquad (2)$$
$$where\ \Omega(f) = \gamma T + \tfrac{1}{2}\lambda||w||^2$$

with l is a differentiable convex loss function that measures the difference between the prediction $\widehat{yi}$ and the target y_i , and Ω penalizes the complexity of the model.[7]

4) K-nearest neighbors (KNN): KNN is one of the most famous classification algorithms. As the KNN algorithm does not assume anything about the underlying data, it is considered to be a non-parametric learning algorithm. It is a supervised learning algorithm without a specialised phase for training. To predict the values of new data points, it uses 'feature similarity' which further means that the position of the new data point will change depending on its proximity to the other points in the training set. We need to preprocess the features so as to prevent any major differences in measurement scale. Adjusted normalization is used to handle features that have a continuous distribution:

$$x_{pn} = (2(x_{pn}^{\,0} - min(x_p^{\,0})) / max(x_p^{\,0}) - min(x_p^{\,0})) -1 \qquad (3)$$

where x_{pn} is the normalized value, in case n for feature p, $x_{pn}^{\,0}$ the original value assigned for case n, min(x^0_p) and max((x^0_p) in all training cases depicts the minimum and maximum value respectively. [15][16]

5) Support Vector Classifier(SVC): The dataset is divided into 2 groups - churn and non churn- to identify customers from each group. The data points are made linearly separable by mapping them to a higher dimensional plane. Hyper plane is the plane which divides data points. It can be used for small dataset to present an optimal solution. For noisy data, SVC cannot be more effective.The hyperplane equation is given as follows:

$$Y = mx+c \qquad (4)$$

Then the equations of decision boundary become:

$$mx+c= +p \qquad (5)$$

$$mx+c= -p \qquad (6)$$

Thus, any hyperplane that satisfies our SVR should satisfy:

$$-p < Y - mx + c < +pc \tag{7}$$

6) Deep Neural Network : To explore the impact of deep learning algorithms for this use case, we use an artificial neural network with multiple layers, also known as Deep Neural Network. Deep Neural Networks are computing systems based on structure and functioning of biological neural networks[14]. For obtaining a Deep Neural Network, we take a Neural Network with only one hidden layer, and then keep on adding more layers. In Deep Neural Networks, a feature hierarchy of increasing abstraction and learning complex concepts is created by training each layer of neurons on the features/outputs of the previous layer.

V. RESULTS

All the above mentioned techniques were performed on the given churn dataset. Grid search algorithm was implemented on the algorithms to increase its performance. For example, initially the accuracy of random forest was 90.96% but after using grid search it was observed that the accuracy increased to 91.26%. The graph displayed in Figure 2 compares the accuracy of all the algorithms on the training set and testing set.

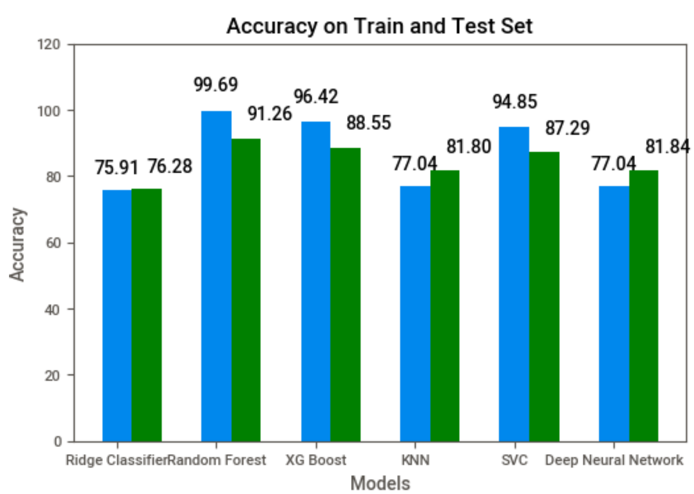

Fig. 2. Accuracy on train and test set

In the above graph shown in Figure 2, a comparative analysis between train and test set is made to measure the accuracy of the models. Initially, it was observed that the Accuracy on Training Set was slightly more and thus it represented Overfitting. It was possible to decrease the depth of the tree to prevent the model from Overfitting, however it would also decrease the accuracy. Therefore, it is necessary to carefully optimize the parameters, since parameter tuning is a very critical part. It is possible to increase the number of trees which would help the Model to be more generalized and reduce Overfitting. Another way is to do Cross Validation which allows us to use every sample in Train Set and Test Set. Hence Grid search Cross validation is implemented which increases the accuracy of the model. An overall accuracy of 90% was achieved after implementing it. The final result showed that the accuracy of Random Forest was the highest on both train set as well as test set.

VI. CONCLUSION

The telecommunication industry has been subjected to major changes in recent years. As a result of being a growing industry, it has now become a competitive market. Due to having increased options of services in this industry, customers tend to switch between these services. Hence, in order to retain their customers, the industry requires a way to understand and predict the customer churn pattern which can be done using churn modelling. Our research was focused on implementing some of these techniques: Random Forest, SVM, Extreme Gradient Boosting (XGBoost), Ridge classifier, K-nearest neighbors (KNN) and Deep Neural Networks. The dataset used was focused on a customer retention program which included various customer attribute fields and also a column of customer churn. We carried out pre-processing steps on this dataset and the same was used as input data for implementing all the techniques. A comparison between all the mentioned techniques was made to classify if a customer will churn or not. The efficiency of these models was further explored by passing them through a grid search. Hence, it was concluded that the Random Forest model works best for this particular use case with a prediction accuracy of 90.96% on the testing data before grid search. Through this research, a relevant inference can be drawn regarding the effectiveness of older but more lightweight models in the prediction of customer churn.

VII. FUTURE SCOPE

The future scope of this paper will focus on ways to achieve better results such as using different hyper-parameter optimization techniques for the same algorithms in a smaller time frame. Different combinations of attributes can be used in the future to determine the customer retention policies. Also the performance factors can be improved using different deep learning approaches.

REFERENCES

[1] Chih Ping Wei, I-Tang Chiu, "Expert Systems with Applications", Volume 23, Issue 2, August 2002, Pages 103–112.

[2] J Lu., "Modeling Customer Lifetime Value Using Survival Analysis: An Application in the Telecommunications Industry", Data Mining Techniques, 2003, SUGI 28.

[3] Golshan Mohammadi, Reza Tavakkoli Moghaddam, and Mehrdad Mohammadi, "Hierarchical Neural Regression Models for Customer Churn Prediction", Journal of Engineering, Volume 2013 , 2013, Article ID 543940, 9 pages.

[4] How Costly Is Customer Churn in the Telecom Industry?, The European Business Review, Aug 2020, Accessed on:Nov1,2020.[Online]. Available: https://www.europeanbusinessreview.com/how-costly-is-customer-churn-in-the-telecom-industry/.

[5] Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. ul, & Kim, S. W. (2019)."A Churn Prediction Model using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector.", IEEE Access, 1–1. doi:10.1109/access.2019.2914999.

[6] Ali Rodan, Hossam Faris, Jamal Alsakran, and Omar Al-Kadi, "A Support Vector Machine approach for Churn Prediction in Telecom Industry", Information -An International Interdisciplinary Journal, Vol. 17(8), 2014.

[7] Iqbal Hanif, "Implementing Extreme Gradient Boosting (XGBoost) Classifier to Improve Customer Churn Prediction", Proceedings of the 1st International Conference on Statistics and Analytics, ICSA 2019, 2-3 August 2019, Bogor, Indonesia.

[8] Abinash Mishra, U. Srinivasulu Reddy, "A Comparative Study of Customer Churn Prediction in Telecom Industry Using Ensemble Based Classifiers " , Proceedings of the International Conference on Inventive Computing and Informatics (ICICI 2017) IEEE Xplore Compliant - Part Number: CFP17L34-ART, ISBN: 978-1-5386-4031-9.

[9] Bhawna Nigam, Himanshu Dugar, Niranjanamurthy M, "Effectual Predicting Telecom Customer Churn using Deep Neural Network", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958, Volume-8 Issue-5, June 2019.

[10] Pradeep B, Sushmitha Vishwanath Rao, Swati M Puranik, Akshay Hegde, "Analysis of Customer Churn prediction in Logistic Industry using Machine Learning", International Journal of Scientific and Research Publications, Volume 7, Issue 11, November 2017 ISSN 2250-3153

[11] Xin Hu, Yanfei Yang, Lanhua Chen, Siru Zhu, "Research on a Customer Churn Combination Prediction Model Based on Decision Tree and Neural Network", 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics.

[12] 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, e-ISSN: 2289-8131 Vol. 9 No. 2-3.

[13] Khan, Muhammad & Manoj, Johua & Singh, Anikate & Blumenstock, Joshua, "Behavioral Modeling for Churn Prediction: Early Indicators and Accurate Predictors of Custom Defection and Loyalty", 10.1109/BigDataCongress.2015.107.

[14] Reena Pagare, Dr. Akhil Khare, "Churn Prediction by Finding Most Influential nodes in Social Network", 2016 International Conference on Computing, Analytics and Security Trends (CAST).

[15] Ionut B. BRANDUSOIU & Gavril TODEREAN "Predicting Churn in mobile telecommunications industry", Volume 54, Number 3, 2013 ACTA TECHNICA NAPOCENSIS Electronics and Telecommunications

[16] Universitatea Tehnica, Accessed on Sept 17, 2020 [Online]. Available: https://users.utcluj.ro

[17] Van Gerven, M., & Bohte, S.(Eds.). (2018), "Artificial neural networks as models of neural information processing", Frontiers Media SA. 2018