Predictive Framework for Advanced Customer Churn Prediction using Machine Learning

Dinesh Kumar Jena

Abhyarthana Bisoyi
Electronics & Instrumentation Engg.
Odisha University of Technology and Research
Bhubaneswar, India

Aruna Tripathy
Electronics & Instrumentation Engg.
Odisha University of Technology and Research
Bhubaneswar, India

ABSTRACT
In recent years, the telecom sector has been burgeoning to satisfy the demand of mobile subscribers and telecom service providers. The increase in number of mobile subscribers and competition among providers, results in the creation of "churners". These are the subscribers who tend to switch from the current telecom service to another. The detection of these churners is called "churn prediction". This prediction has become a major challenge for telecom companies. The main purpose of customer churn prediction is to estimate the number of subscribers those who want to quit the current service provider by providing specific solutions to retain them. This paper proposes methods for the estimation of churners by applying different classification techniques and estimates the differences between them. The performance is measured by taking different parameters like accuracy, precision, recall, etc. In this paper, the various performance measurement and comparison are done by using the dataset collected from American Telecommunication Company. All the proposed work is based on Machine Learning, inculcating the supervised learning. In addition to all, a single test-bed is designed as a user interface to predict the individual customer according to different attributes.

Keywords
BigML, Churning, Customer churn prediction, Decision tree, Machine Learning, Random Forest, Supervised learning, Telecom, User Interface

1. INTRODUCTION
Now-a-days the telecom industry is facing a competition among each other due to the entry of Reliance Jio into the market. In the middle of 2017-19, the loss incurred by service providers other than Jio, touched a high value which never occurred before. The onset of Jio created a competition among the service providers to retain their existing customers for a long-term profit [1-3]. Customer churn prediction plays a vital role to tackle the above said situation. Basically, customer churn refers to the inclination of existing customers of the current service providers towards the any other telecom service providers. The number of customers who leave the current provider in a particular interval of time are referred as churners. In the above scenario—to retain the customers and to keep a track of them—each telecom company should have a good predictive customer churn framework.

Customer churn prediction is a prominent method to check the indirect feedback of customers as well as the profitability. Some statistics show the importance of churn prediction to the customer retain capacity of the company. One of the studies shows that, 1% increase in the customer retains campaign may result in 5% increase in the overall values of the company [5]. In the telecommunication industry, the monthly rate of customer churn is 2.2% and the annual rate of customer churn is 27% [4]. If we consider Europe and America, the yearly cost of customer churn is 4 billion dollars approximately. It is seen that if a company retain 1.5 million customers by increasing the rate at 1%, then this may add a benefit of 54 million dollars for the same year.

Prodigious researches have been conducted for the creation of different models, development of new algorithms, development of existing models and classifiers, comparison of algorithms, efficiency, etc in order to retain the churners.

In 2017, model was proposed by using decision tree classifier to solve unbalanced problem, scatter problem, high dimensional data in telecom data set [1]. In 2015, the rotation forest approach was followed to evaluate the prediction model, which is compared by C4.5 decision tree and ant-miner [2]. In addition, k-means clustering and down sampling to preprocess the data followed by the decision tree C5.0 and random forest for training and evaluation of data set has been reported in [3]. Further an attribute derivation process was implemented to increase the correct prediction rate [4]. Then, a Bayesian Belief network method was attempted in a study which was mentioned in [5]. Further the paper proposed support vector machine to increase the performance and an increased accuracy was obtained by using two different rules extraction method. Two methods like AntMiner+ and ALBA were implemented in [6-8]. A data mining model was developed for for the telecom churners about some churn prediction in the telecom sector in [9-10]. In addition, a hybrid model was developed as a combination of two artificial neural networks and a second hybrid model was a combination of self-organizing maps and artificial neural networks. [11-12]. In 2011, parameters such as F-measure and accuracy, which could be helpful in the development of Machine Learning algorithms were discussed in [13]. Various novel methods using Convolutional Neural Network (CNN) and other supervised learning methods for churn prediction have been reported in the recent time [14-20]. The recent work in churn prediction includes methods based on unstructured data [21-22]. Also, the decision tree has been implemented often in image processing-based application, as seen in [23-24]. These days random forest is being used in myriads of applications, as it provides better accuracy [25-28]. Supervised learning algorithms have been modified to self-supervised algorithms providing better results in image-based applications [29-33]. The major issues arising due to customer churning and its solution lies in determining the factors that leads to churning [34-35].

A predictive framework by decision tree and random forest approach is proposed in this paper after processing the data by random sampling method. And then, by taking the evaluation and result into account, a user interface is prepared to predict the individual churners according to different attributes.

The rest of the paper is organized as follows. Section 2 explains the decision tree and random forest classifiers respectively. Section 3 provides the churn prediction model that includes data set description, data pre-processing, training the model and evaluation. Section 4 discusses with the result obtained from the evaluation. Then an individual user interface is designed in section 5. At last, the conclusion and future work is presented in section 6.

2. DECISION TREE & RANDOM DECISION FOREST

3.1 Decision Tree

We have used the C5.0 decision tree algorithm as one of the classifiers in the presented work as this classifier provides better results than C4.5 and CART algorithm [3]. We have taken the dataset that comprises of nominal as well as ordinal features. Usually, the C5.0 tree algorithm is applicable for nominal data, but here, ordinal variables are treated as nominal ones.

In the proposed model, the dataset is split into small subsets. The leaf node represents the decision node and according to the feature values of instances, the decision tree classification is done. The root node or parent node of decision tree is selected among different attributes based on the information gain. From all the attributes, the attribute having highest information gain is selected as the parent node. Then, in order to calculate the information gain, we have to calculate the total entropy of the dataset.

Entropy of the total samples is given by [3],

$$Entropy\ (Ent) = -\sum_{i=1}^{n} [p_i^+ \log_2(p_i^+) + p_i^- \log_2(p_i^-)] \quad (1)$$

where, p_i^+ and p_i^- represent the positive as well as negative target variable in the dataset.

The information gain of each attribute is given by;

$$Gain = Ent - [weighted\ avg * Ent\ of\ each\ feature] \quad (2)$$

Here, the info gain calculation and entropy measurement are the basic theory of decision tree C5.0. [5]

3.2 Random Decision Forest

Random forest is an ensemble of decision tree which is widely used for classification and regression problems [25-28]. In this paper, the random forest is used as the second classifier to evaluate the model. The random forest method is widely accepted due to its flexibility and accuracy [3].

Random forest initiates the tree building procedure to form several decision trees by taking different attributes. In case of random forest, the split node is chosen randomly during the construction of decision tree. The method comprises of several decision tree having arbitrarily selected attributes. The error in this method is reduced by the enhancement of discrepancy of trees. After the forest is formed, each tree will decide its own decision for each sample. Finally, the outcome of random forest is obtained after considering the average vote of all trees.

3. PROPOSED MODEL FOR CHURN PREDICTION

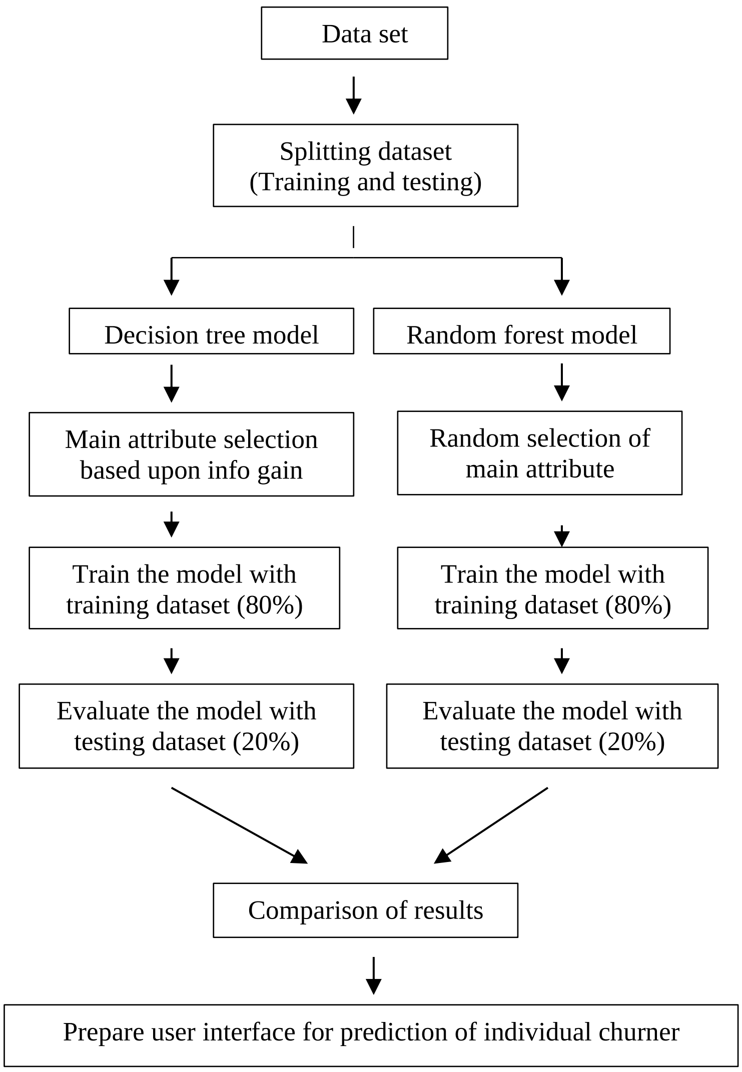

Figure 1: Model flow of the customer churn prediction scheme

3.1 Dataset

The dataset [7] used in this paper has been obtained from the UCI repository of machine learning databases at the university of California, Irvine. The dataset is comprised of the information about 3333 customers. It contains 20 variables of the customer along with one target variable ‘churn’ an indication of whether the customer left the company or not.

3.2 Data Preprocessing

The model flow of the customer churn prediction scheme is shown in figure 1. Here, the main dataset is split into training data (80%) as well as testing data (20%). From the training dataset 2286 instances (85.75%) are not churners i.e., their churn status is false and 380 instances (14.25%) are churners i.e., their churn status is true.

3.3 Training the model

In this work, the important attributes for classification of decision tree are selected according to the information gain principle as outlined in Eq (1) and Eq (2). The attribute having high information gain is considered as parent node and the condition of splitting starts from that node. The bar graph in figure 2 represents the training field importance in case of decision tree.

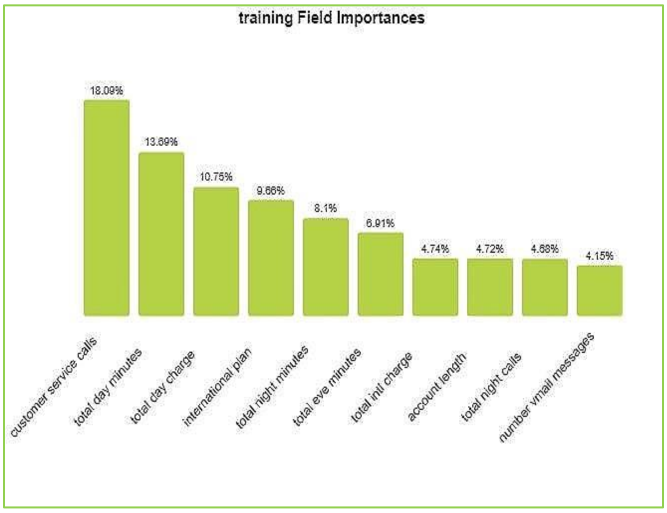

Figure 2: Training field importance in case of decision tree

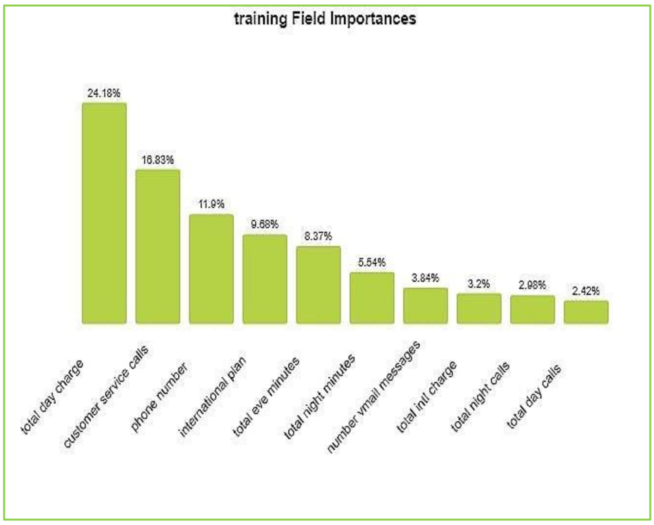

Figure 3: Training field importance in case of random forest

In case of random decision forest, the several subsets of data are extracted from the original training dataset arbitrarily. Along with the random dataset, it takes the random selection of features rather than all features. The bar graph in figure 3 represents the training field importance in case of random forest.

3.4 Model evaluation

Evaluation or testing is a method to check the efficiency of the model with the help of confusion matrix as mentioned in table 1 and table 2. A common method to analyse the predictive performance of the model is the confusion matrix. Here, the model is compared with the testing dataset (20%) which is split in the data pre-processing stage. The confusion matrix consists of four variables: true positive (TP), false positive (FP), true negative (TN), false negative (FN) which are described in the tables below. The above four variables are observed properly to retain the customers and to avoid the unnecessary promotions for the false customers.

Table-1 represents the testing dataset comprised of 667 subscribers. Among them, 103 subscribers are churners (already switched from one service provider to other) and 564 subscribers are non-churners (continuing with the present service provider). After the evaluation, it is predicted that among 667 subscribers, 126 subscribers are going to switch to another service provider and 541 will retain with the current service provider. The confusion matrix result elucidates that there is a difference of 23 subscribers in actual and predicted values.

Table 1. Class confusion matrix for decision tree

		PREDICTED		**TOTAL**
		Churn customer	***Non-Churn Customer***	
ACTUAL	Churn Customer	TP (81) True Positive	FN (22) False Negative	Actual Positive (103)
	Non-Churn Customer	FP (45) False Positive	TN (519) True Negative	Actual Negative (564)
TOTAL		Predicted (126) Positive	Predicted (541) Negative	Total Customer (667)

Table 2. Class confusion matrix for random forest

		PREDICTED		**TOTAL**
		Churn customer	***Non-Churn Customer***	
ACTUAL	Churn Customer	TP (86) True Positive	FN (17) False Negative	Actual Positive (103)
	Non-Churn Customer	FP (8) False Positive	TN (556) True Negative	Actual Negative (564)
TOTAL		Predicted Positive (94)	Predicted Negative (573)	Total Customer (667)

Table-2 represents the testing dataset comprised of 667 subscribers using the random forest classifier. Among them, 103 subscribers are churners (already switched from one service provider to another) and 564 subscribers are non-churners (continuing with the present service provider). After the evaluation, it is predicted that among 667 subscribers, 94 subscribers are going to switch to another service provider and 573 will retain in the current service provider. From the confusion matrix result, it is clear that there is a difference of 9 subscribers in actual and predicted values.

There are several classification measures available to evaluate a classification model. In this paper: accuracy, precision, recall, F-measure and phi-coefficient are considered for evaluation purpose. All the aforementioned classification measures are described below by using the variable used in confusion matrix.

Equations (3), (4), (5), (6) and (7) represent the above said measures; [4]

$$Accuracy = \frac{TP + TN}{Total\ instances} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Precision = \frac{TP}{TP + FN} \quad (5)$$

$$\text{F} - mesaure = \frac{2 * precision * recall}{precision + recall} \quad (6)$$

$$phi - coefficient = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TN + FN)(TN + FP)(TN + FN)}} \quad (7)$$

In equation (3), the accuracy indicates the number of correctly classified instances to the total instances evaluated. In equation (4) and (5), the precision and recall indicate the percentage of correctly predicted instances over the total instances predicted for the positive class and the total actual instances for the positive class. High value of all the three measures indicates better performance of the model. In equation 6, the F-measure indicates the balance harmonic mean between precision and recall. It ranges from 0 to 1. High value indicates better performance. In equation 7, the phi-coefficient depicts the correlation coefficient between predicted and actual values. It ranges between -1 to 1 and the coefficient of 1 indicates a perfect prediction.

4. RESULT DISCUSSION

All the results are observed by taking positive class true into account—the customers who churned from one provider to other. The random forest method is observed to perform better than the decision tree method in terms of accuracy, precision, recall, F-measure, phi-coefficient as it produces higher values.

Table 3 shows the comparison of the two models in terms of the five above mentioned classification measures. Furthermore, it has been observed that random forest approach is a better approach among the two classifying methods. As accuracy does not reflect the truth in all the cases, the other four variables can be taken into consideration. As far as classification model is concerned, all the five parameters play a significant role for the evaluation of the model.

Table 3. Classification Results

CLASSIFICATION MEASURES	**DIFFERENT ALGORITHMS**	
	Decision Tree	***Random Forest***
Accuracy (%)	89.96	96.25
Precision (%)	64.29	91.49
Recall (%)	78.64	83.49
F-measure	0.71	0.87
Phi coefficient	0.65	0.85

The increased value of all the parameter in case of random forest compared to the decision tree signifies a better classifier performance. According to different thresholds, different confusion matrices along with different plots are obtained through Python simulations. A better method to evaluate the model is to plot the different measures in different charts. In this paper, four curves have been have considered i.e., region of convergence (ROC), precision-recall, gain and k-s statistic, and lift for the model evaluation.

The ROC curve is the plot between false positive rate (FPR) to true positive rate (TPR). The TPR is obtained by calculating the recall and the FPR is equivalent to (1-precision). [4]

$$FPR = \frac{FP}{TN+FP} \quad (8)$$

A higher Area Under the Curve (AUC) indicates a better classifier performance. Here, area under the convex hull (AUCH) is calculated taking into account the convex shape of the curve where no other points lay above the curve.

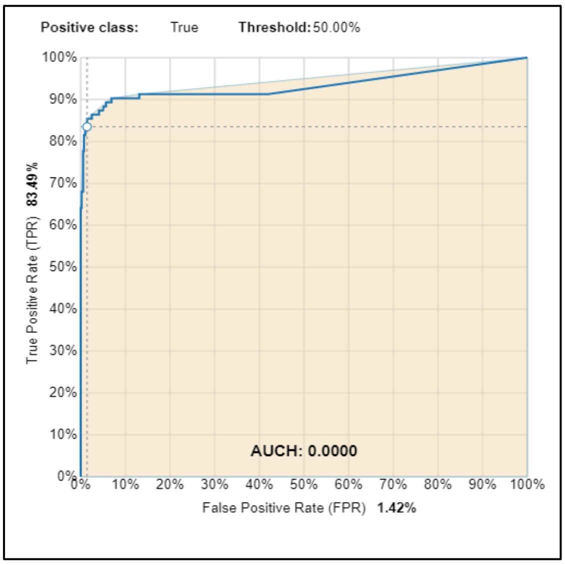

Figure 4: AUCH for TPR&FPR in Random Forest

Figure 4 shows the AUC between TPR and FPR in case of random forest. The x-axis represents an FPR with 1.42% and the y-axis represents TPR with 83.49% which are taken from the evaluation of classification measures. From the figure it can be observed that the intersection point lies in the upper left part of the graph and after 83% of the TPR, FPR starts increasing. Calculation of the AUCH by the help of BigML yielded a value of 0.9458 by taking the convex hull into account.

Figure 5 shows the area under the curve between TPR and FPR in case of decision tree. The x-axis represents an FPR with 7.98% and the y-axis represents TPR with 78.64% which are taken from the evaluation of classification measures.

It is observed that the FPR in case of decision tree is more than the random forest method and the intersection point is not as closer as random forest, which decreases the performance of the decision tree model. Calculation of the AUCH by the help of BigML yields a value of 0.8748 by taking the convex hull into account. As the AUCH in case of random forest is greater than that of the decision tree, the performance of the model is better in case of random forest.

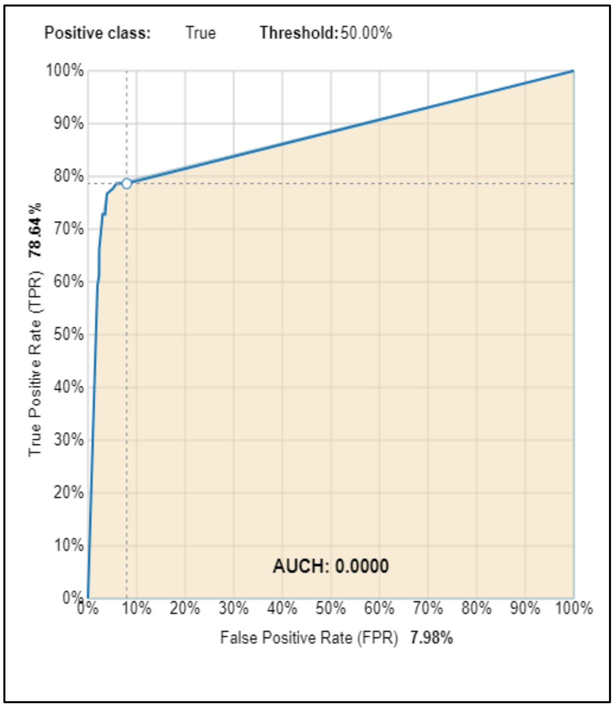

Figure 5: AUCH for TPR&FPR in decision tree

Similarly, precision-recall curve represents the trade- off between both measures in terms of positive class. High precision-recall represents that the AUC is bigger and hence represents the better classifier.

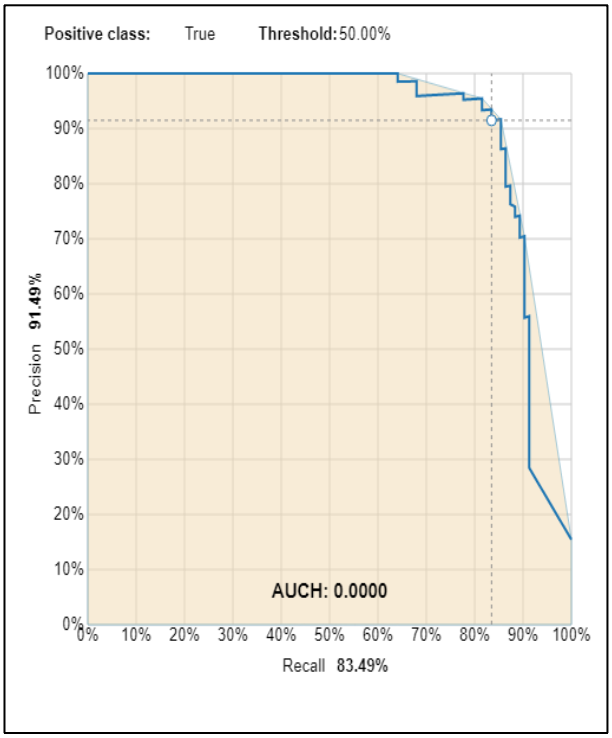

Figure 6: Precision-recall AUCH in Random Forest

Figure 6 shows the AUC between precision and recall in case of random decision forest. The x-axis represents the recall with 83.49% and the y-axis represents precision with 91.49%. All the values are taken from the evaluation of classification measures. From this figure it is also depicted that the convex shape of the curve begins from the intersection point of precision and recall. i.e., in terms of convex shape there is a sudden fall of precision after 83.49% of recall. On calculating the AUCH by the help of BigML a value of 0.9280 has been observed taking the convex hull into account.

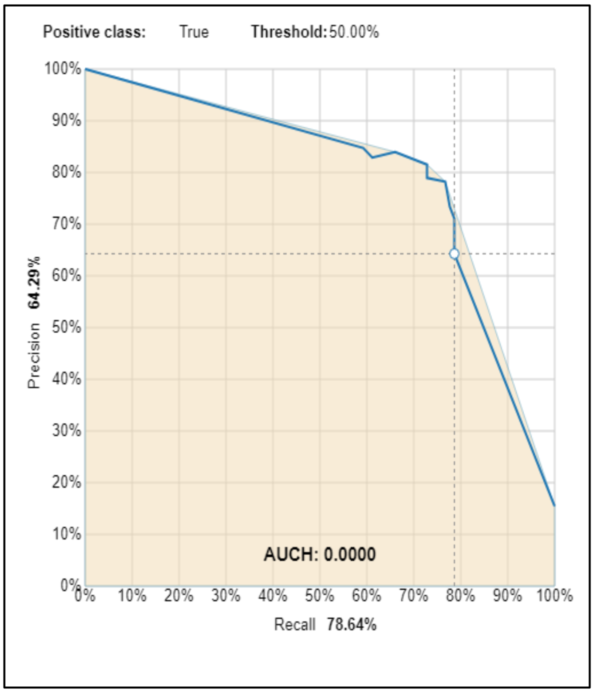

Figure 7: Precision-recall AUCH in Decision Tree

Figure 7 shows the area under the curve between precision and recall in case of decision tree. The x-axis represents the recall with 78.64% and the y-axis represents precision with 64.29%. All the values are taken from the evaluation of classification measures. From this figure it is also depicted that the intersection point of precision and recall is not much closer to the point (1,1) as compared to the random forest classifier. On calculating the AUCH (with the help of BigML), it shows 0.8025 taking the convex hull into account. So, it is concluded that as the AUCH in case of random forest is greater than the decision tree, the performance of the model is better in case of random forest.

Similarly, the K-S statistic is an indicator of how well the model separates the positive from the negative classes. A K-S statistic of 100% indicates a perfect separation and a model that classifies everything correctly. Higher values for the K-S statistic indicate a higher quality model.

Figure 8 represents the gain curve between true positive rate (Gain) and percentage of positive instances in case of random decision forest. On calculating the (K-S) value with the help of BigML, it shows 84.02% for random forest. The above value indicates a good separation between the positive classes from the negative classes. It cannot be neglected that the prediction of the model depends on the classification of classes, as in how much the classes are separated from each other.

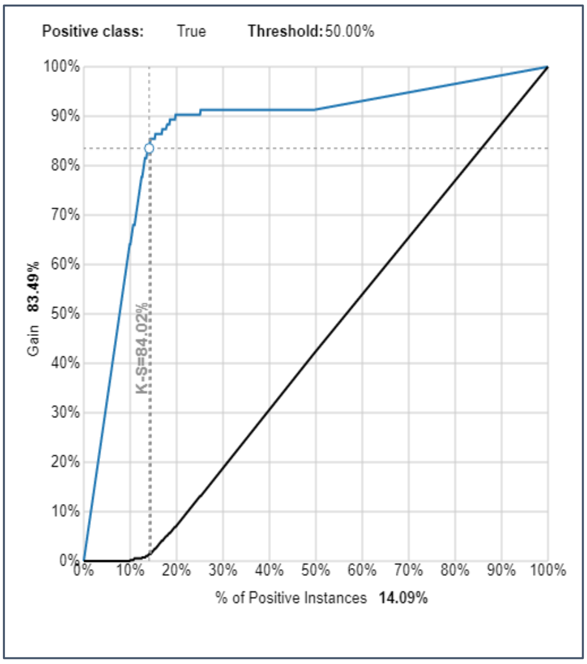

Figure 8: Gain curve in Random Forest

Figure 9 represents the gain curve between true positive rate (Gain) and percentage of positive instances in case of decision tree classifier. On calculating the (K-S) value with the help of BigML it shows 72.80% for decision tree.

So, it can be observed that on comparison to random decision forest, decision tree is not able to separate the positive classes from the negative classes. In case of random decision forest, the K-S value is greater than the decision tree, indicates the model well separates the positive and negative classes.

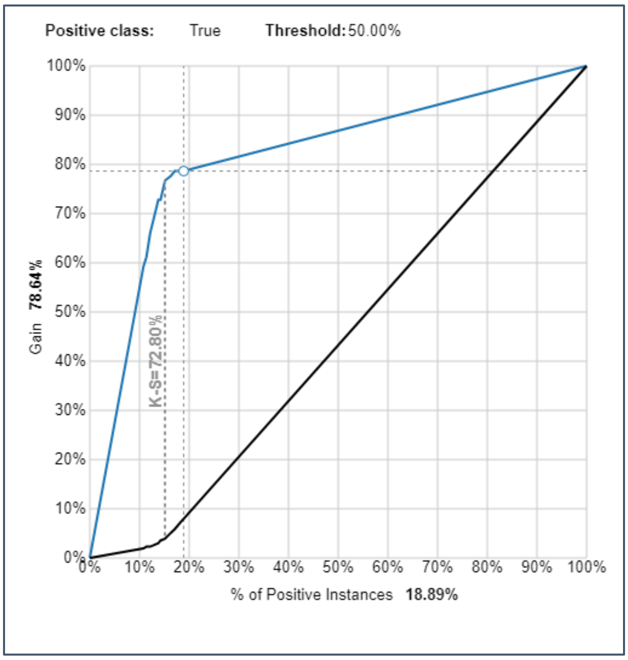

Figure 9: Gain curve in Decision tree

The Lift curve shows the goodness of fit of the model compared to a random class assignment given in a sample of positive instances. The Lift is plotted in the y-axis and it is calculated as the ratio between the result predicted by the model and the result using no model. The x-axis represents again the percentage of correct predictions. The horizontal line in the chart indicating a 100% lift represents a model that makes random predictions.

$$\text{Lift} = \frac{precision}{\frac{positive\ instances}{total\ instances}} = \frac{\frac{TP}{TP+FP}}{\frac{TP+FN}{TP+FP+TN+FN}} \tag{9}$$

$$\%\ \text{of positive instances} = \frac{TP+FP}{TP+FP+TN+FN} \tag{10}$$

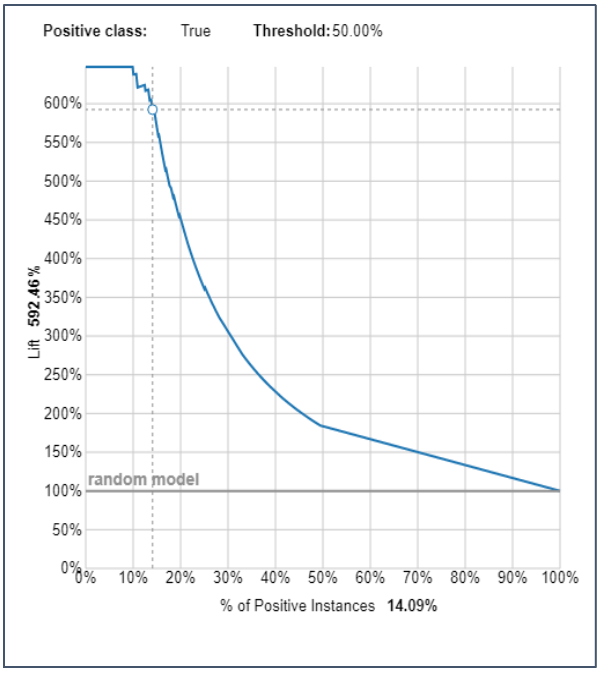

Figure 10: Lift curve in Random Forest

Figure 10 represents the List curve in the y-axis with the percentage of positive instances in x-axis for random forest classifier. From the calculation it is seen that 592.46% the model lift compared to the random model classifier. From the intersection point, it is inferred that when the positive instances less than 14% the lift curve suddenly rises and attains the maximum value. It indicates that a classification model comprised of random forest algorithm is better than the arbitrary model by 492.46%.

Figure 11 represents the Lift curve in the y-axis with the percentage of positive instances in x-axis for decision tree classifier. From the calculation it is seen that 416.30% the model lift compared to the arbitrary model classifier. From the intersection point it infers that when the positive instances less than 18% the lift curve suddenly rises and attains the maximum value. It indicates that a classification model comprised of decision tree algorithm is better than the arbitrary model by 316.30%.

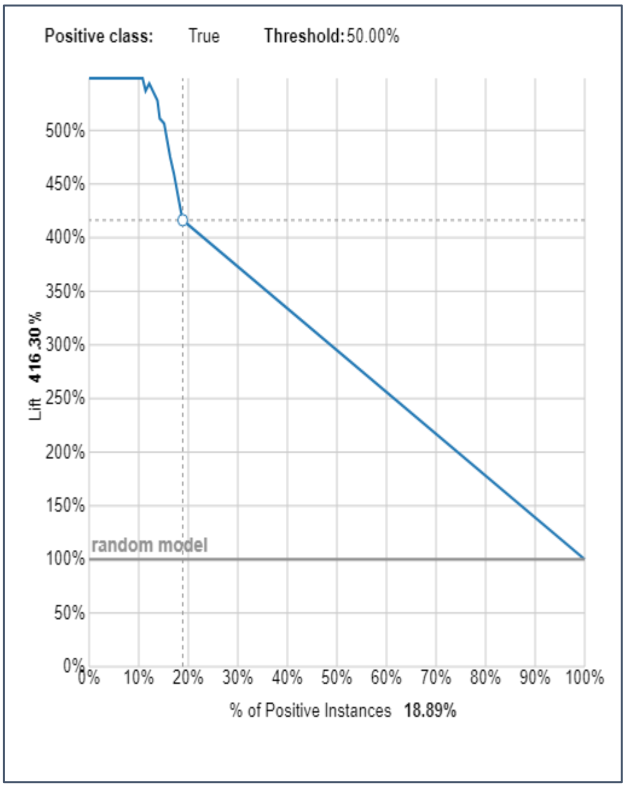

Figure 11: Lift curve in Decision tree

5. USER INTERFACE

Taking the above result into account a user interface shown in figure 12 is prepared to distinguish the individual churners with the help of Python tool. The interface represents different attributes which are used for training as well as testing the dataset. If a user, having the attribute of a single individual and wants to predict whether the individual is going to churn or not, then it can be possible by the help of above interface. The figure below represents the user interface for customer churn prediction.

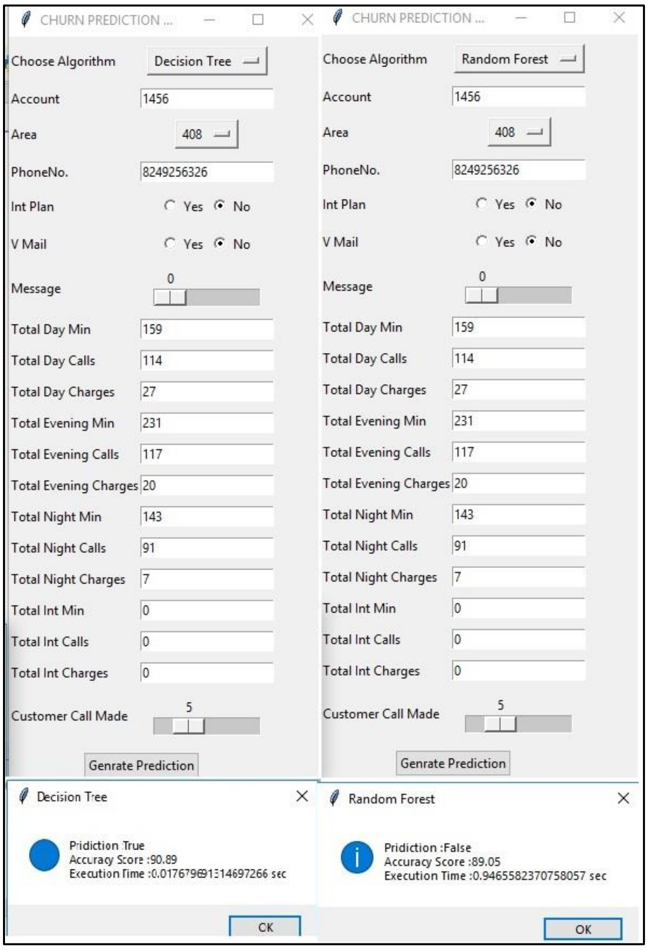

Figure 12: User interface for customer churn prediction

6. CONCLUSION

As per the results obtained—as mentioned in section 4—the random decision forest is better than C5.0 decision tree, because more the number of true predictions of churn customers, less the chances of churning. Also, the random forest method is observed to perform better than the decision tree method in terms of accuracy, precision, recall, F-measure, phi-coefficient as it produces higher values. On comparing the results obtained from random forest and decision tree, it is seen that the random forest accuracy increased by 6.30%, precision increased by 27.20%, recall increased by 4.85%, f-measure increased by 0.17 and phi-coefficient increased by 0.20. Apart from that in case of evaluation curves, K-S statistics in case of random forest increased by 11.22% than the decision tree classifier and the lift curve gain 176.16% over the random classifier. Altogether, the process of churn prediction is applicable to telecom industries, as they have been losing their loyal customers. In fact, not only its importance lies in detection of churners but also an alert for the telecom industries to adapt the competition in the market. The above-mentioned algorithm will act as a foreseer and the telecom services should come up with alternative plans to save the customers from being churners.

Since we had developed a supervised learning model, we are now in the verge of developing an unsupervised learning model. Certainly, the model development will be cumbersome but the efficiency will be higher than the supervised learning. In future, we will consider more machine learning algorithms i.e., logistic regression, neural network et.al. for the investigation of high value of data such as daily, monthly and yearly data. Along with this different clustering algorithms will be tested as well. Also new supervised learning algorithms can be developed for optimizing the global minima (with proper mathematical model).

7. REFERENCES

[1] P. Bhanuprakash and G. S. Nagaraja, "A review on churn prediction modeling in telecom environment", 2nd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions, 2017.

[2] M. Yildiz and S. Albayrak, "Customer churn prediction in telecommunication", 23rd Signal Processing and Communications Applications Conference (SIU), 2015.

[3] H. Li, D. Yang, L. Yang, Y. Lu and X. Lin, "Supervised Massive Data Analysis for Telecommunication Customer Churn Prediction", Guangdong Province Key Laboratory of omputational Science, Guangzhou, China. ISBN 978-1-5090-3936- 4/16.

[4] B. Huang, M. T. Kechadi, and B. Buckley, "Customer churn prediction in telecommunications", Expert Systems with Applications, 39(1), pp. 1414-1425, 2012.

[5] P. Kisioglu and Y. I. Topcu, "Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey", in Expert Systems with Applications, pp. 7151-7157, June 2011.

[6] Y. Zhao, B. Li and X. Li, "Customer churn prediction using improved one-class support vector machine", Lecture Notes in Artificial Intelligence, pp. 300-306, 2005.

[7] C. L. Blake and C. J. Merz, "Churn Data Set, UCI Repository of Machine Learning Databases", University of California, Department of Information and Computer Science, Irvine, CA, 1998, kdd.ics.uci.edu/. Accessed March 17, 2014.

[8] W. Verbeke, K. Dejaeger, D. Martens, J. Hur and B. Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach", European Journal of Operational Research, Vol. 218, No. 1, pp. 211-229, 2012.

[9] S. Y. Hung, D. C. Yen, and H. Y. Wang, "Applying data mining to telecom churn management", Expert Systems with Applications, Vol. 31, pp. 515-524, 2006.

[10] N. Kamalraj and A. Malathi, "A Survey on Churn Prediction Techniques in Communication Sector", International Journal of Computer Applications, Vol. 64, No. 5, pp. 39-42, February 2013.

[11] C. F. Tsai and Y. H. Lu, "Customer churn prediction by hybrid neural networks", Expert Systems Application, Vol. 36 issue 10, pp.12547-12553, 2009.

[12] K. W. Bock and D. V. Poel, "An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction", Expert Systems with Applications, Vol. 38 issue 10, pp.12293-12301, 2011.

[13] Powers and David. M.W., "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation", Journal of Machine Learning Technologies, Vol. 2 issue 1, pp-37-63, 2011.

[14] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," in IEEE Access, vol. 7, pp. 60134-60149, 2019.

[15] E. Lee, B. Kim, S. Kang, B. Kang, Y. Jang and H. K. Kim, "Profit Optimizing Churn Prediction for Long-Term Loyal Customers in Online Games," in *IEEE Transactions on Games*, vol. 12, no. 1, pp. 41-53, March 2020.

[16] Y. Zhang, S. He, S. Li and J. Chen, "Intra-Operator Customer Churn in Telecommunications: A Systematic Perspective," in *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 948-957, Jan. 2020.

[17] S. Wu, W. -C. Yau, T. -S. Ong and S. -C. Chong, "Integrated Churn Prediction and Customer Segmentation Framework for Telco Business," in *IEEE Access*, vol. 9, pp. 62118-62136, 2021.

[18] Y. Deng, D. Li, L. Yang, J. Tang and J. Zhao, "Analysis and prediction of bank user churn based on ensemble learning algorithm," *2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA)*, pp. 288-291, 2021.

[19] R. K. VLN and P. Deeplakshmi, "Dynamic Churn Prediction using Machine Learning Algorithms - Predict your customer through customer behaviour," *2021 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1-6, 2021.

[20] R. Yahaya, O. A. Abisoye and S. A. Bashir, "An Enhanced Bank Customers Churn Prediction Model Using a Hybrid Genetic Algorithm And K-Means Filter And Artificial Neural Network," *2020 IEEE 2nd International Conference on Cyberspac (CYBER NIGERIA)*, pp. 52-58, 2021.

[21] H. Karamollaoğlu, İ. Yücedağ and İ. A. Doğru, "Customer Churn Prediction Using Machine Learning Methods: A Comparative Analysis," *2021 6th International Conference on Computer Science* and Engineering (UBMK), 2021, pp. 139-144.

[22] Y. E. Özköse, A. Haznedaroğlu and L. M. Arslan, "Customer Churn Analysis with Deep Learning Methods on Unstructured Data," *2021 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 2021, pp. 1-5.

[23] S. Alex, K. J. Dhanaraj and P. P. Deepthi, "Private and Energy-Efficient Decision Tree-Based Disease Detection for Resource-Constrained Medical Users in Mobile Healthcare Network," in *IEEE Access*, vol. 10, pp. 17098-17112, 2022.

[24] S. Sun, T. Liu, Y. Wang, G. Zhang, K. Liu and Y. Wang, "High-Rate Underwater Acoustic Localization Based on the Decision Tree," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-12, 2022.

[25] J. Sun, H. Yu, G. Zhong, J. Dong, S. Zhang and H. Yu, "Random Shapley Forests: Cooperative Game-Based Random Forests with Consistency," in *IEEE Transactions on Cybernetics*, vol. 52, no. 1, pp. 205-214, Jan. 2022.

[26] F. Tong and Y. Zhang, "Exploiting Spectral–Spatial Information Using Deep Random Forest for Hyperspectral Imagery Classification," in *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1-5, 2022.

[27] V. Jain, A. Phophalia and S. K. Mitra, "HML-RF: Hybrid Multi-Label Random Forest," in *IEEE Access*, vol. 10, pp. 22902-22914, 2022.

[28] N. Haque, A. Jamshed, K. Chatterjee and S. Chatterjee, "Accurate Sensing of Power Transformer Faults from Dissolved Gas Data Using Random Forest Classifier Aided by Data Clustering Method," in *IEEE Sensors Journal*, vol. 22, no. 6, pp. 5902-5910, *2022.*

[29] G. Albi, S. Bicego and D. Kalise, *"Gradient-augmented Supervised Learning of Optimal Feedback Laws Using State-Dependent Riccati Equations," in* IEEE Control Systems Letters, vol. 6, pp. 836-841, 2022.

[30] X. He, L. Fang, M. Tan and X. Chen, *"Intra- and Inter-Slice Contrastive Learning for Point Supervised OCT Fluid Segmentation," in* IEEE Transactions on Image Processing, vol. 31, pp. 1870-1881, 2022.

[31] H. Li et al., *"Global and Local Contrastive Self-Supervised Learning for Semantic Segmentation of HR Remote Sensing Images," in* IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-14, 2022.

[32] Z. Song, X. Yang, Z. Xu and I. King, *"Graph-Based Semi-Supervised Learning: A Comprehensive Review," in* IEEE Transactions on Neural Networks and Learning Systems (Early Access).

[33] Y. Liu, J. Wu, L. Qu, T. Gan, J. Yin and L. Nie, *"Self-supervised Correlation Learning for Cross-Modal Retrieval," in* IEEE Transactions on Multimedia (Early Access).

[34] Y. Bharambe, P. Deshmukh, P. Karanjawane, D. Chaudhari and N. M. Ranjan, "Churn Prediction in Telecommunication Industry", *2023 International Conference for Advancement in Technology (ICONAT), Goa, India*, 2023, pp. 1-5.

[35] R. Srinivasan, D. Rajeswari and G. Elangovan, "Customer Churn Prediction Using Machine Learning Approaches," *2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF)*, Chennai, India, 2023, pp. 1-6.