Formosa Journal of Computer and Information Science (FJCIS)
Vol.2, No.1, 2023: 95-110

Artificial Intelligence Based Predictive Analysis of Customer Churn

Vaani Gupta[1*], Aman Jatain[2]
Amity University Gurugram
Corresponding Author: Vaani Gupta vaanigupta01@gmail.com

A R T I C L E I N F O

Keywords: Artificial Intelligence Based Predictive Analysis of Customer Churn, Comparative Graphs, Deep Learning, Heroku, Machine Learning

Received : 05 February
Revised : 20 February
Accepted: 20 March

A B S T R A C T

Customer churn, also known as attrition, occurs when subscribers or customers stop doing business with an enterprise or organization by unsubscribing to a service, discontinuing membership or simply stopping payment. Churn is a critical metric because it is more cost-effective to retain existing customers than it is to acquire new ones. Since churning impedes growth, companies usually use a defined method for calculating customer churn in a given period. By monitoring churn rate and the various factors affecting it, organizations determine their customer retention success rates and identify strategies for improvement. The model, "Artificial Intelligence based Predictive Analysis of Customer Churn" is aimed at predicting a customer's likelihood of discontinuing the usage of the products or services extended by a company, using ML & DL algorithms. It is built for the example of a US based telephone service provider's services. This model uses machine learning's capability to identify patterns and build prediction algorithms using provided data, blended with the adoption of complex algorithms of deep neural networks, which form deep learning. Accuracy tests, comparative graphs and illustrations equip this project with higher efficiency and good performance by estimating which is the best model from the suite of models created

DOI: https://doi.org/10.55927/fjcis.v2i1.3926
ISSN-E: 2830-3040
https://journal.formosapublisher.org/index.php/fjcis

INTRODUCTION

From the direct broadcast satellite service provider customer who doesn't renew their subscription to the library member who cancels their membership; there is a great deal of customers who decide to stop continuing business with a company. This loss of business is called customer churn. Users can choose their best option from a wide array of providers because in today's world of digitalization and crowded marketplaces, it's easier than ever for customers to make comparisons. With factors like poor customer service & market fit, weak onboarding and involuntary churn, companies are leaning more and more towards investigating the reasons for customer churn, how much every factor contributes to this issue & are looking for methods to tackle this problem efficiently.

Machine learning in the form of Neural Networks, Bayes Network [1], [2], [5], Under Sampling [6], Rough Set classification [10], MTDF [11], and SVM [5] have successfully predicted churn rates with high accuracy in the past. This paper attempts to review these literature and expand further with newer and better performing methods, involving boosting algorithms and deep sequential modelling.

The customer churn predictive analysis application created in this project works on ML & DL models to render a solution to the above problem. On the application's webpage, the user must define values of parameters like,

1. Name of the state where the service is being used
2. The type of cell phone plan they have
3. The number, duration and charge of messages and calls made throughout the day
4. And importantly, the machine learning algorithm they would like to use to predict customer churn.

A dataset of 3333 instances with 20 attributes was used to train and test the models. 80% of this data was reserved for training the models and 20% was used to test the models hence prepared. Because of their applicability and diversity in this type of application, nine machine algorithms and one deep learning algorithm was chosen to build a model to help identify the customers who are going to migrate to other company services, by specifically pointing out the attributes and inputs of the company's services which make the customers switch service providers intentionally.

LITERATURE REVIEW

Most approaches applied to predict churn in telecom companies have used machine learning and data mining. Majority of related work focuses on applying only one data mining method to extract knowledge. The following research papers compare a maximum of four different models' performances and accuracies to determine customer churn rate in contrast to the ten machine learning or deep learning models prepared, studied and evaluated in this project.

Table 1. Approach and Techniques Used

	Approach and Techniques Used	Accuracy / Outcome
M.A.H. Farquad, 2014 [1]	Hybrid approach to overcome drawbacks of SVM (Black Box) model: – Phase I: SVM-RFE reduces feature set. – Phase II: SVM model obtained, support vectors extracted – Phase III: Rules Generated using Naïve Bayes Tree Highly unbalanced dataset - 93.24% loyal, 6.76% churned customers.	Model not scalable to large datasets
He et al., 2009 [2]	– Neural Network algorithm – Chinese telecom co. with 5.23 million customers	Overall accuracy rate: 91.1%
Huang et al., 2015 [3]	– Random Forest algo evaluated by Area under ROC Curve (AUC) – Big data enhances churn prediction depending on volume, variety, velocity of data – Data from Operation Support & Business Support Dept. at China's largest telecom. company	
Lee et al., 2011 [4]	– Partial Least Squares (PLS) method on highly correlated datasets among variables – Simple & implementable churn marketing program was employed – Marketing programs used to maintain churn levels	Optimal churner -level maintained effectively and efficiently
Gavril et al., 2016 [5]	– (Data Mining) Principal Component Analysis Algorithm, "PCA" to reduce data dimensions of the same data as used in this paper	AUC values: Bayes: 99.10% NN: 99.55% SVM: 99.70%

	– 3 ML algos: Neural Networks, SVM, and Bayes Networks	
Burez & Van den Poel, 2019 [6]	– Studied unbalance datasets in churn prediction models – Compared Random Sampling, Advanced Under-Sampling, Gradient Boosting Model and Weighted Random Forests – Used AUC, Lift metrics to evaluate model	Under Sampling technique outperformed the others
Ver-braken et al., 2013 [7]	New performance measure - Expected Maximum Profit Criterion – aligned with the main objectives of the end users – assists companies with selecting the classifier that maximizes profit – provides info about the fraction of customer base to be included in the retention campaign	
Koen W. De Bock, 2012 [8]	– GAMensPlus - ensemble classifier based on generalized additive models (GAMs) - both performance and interpretability are reconciled and evaluated – GAMens, based on Bagging, the Random Subspace Method & semi-parametric GAMs as constituent classifiers, is extended to include 2 instruments for model interpretability: – the generalized feature importance scores – bootstrap confidence bands for smoothing splines – strong classification performance - at least as good as logistic regression, and GAM	
Idris, 2012 [9]	– genetic programming with AdaBoost – 2 datasets – Orange Telecom and cell2cell	63% & 89% respectively
Makhtar et al., 2017 [10]	Rough Set classification algorithm outperformed the other algorithms like Linear Regression, Decision Tree, and Voted Perception Neural Network	
Amin et al., 2016 [11]	– compared six different sampling techniques for over-sampling – MTDF and rules-generation based on genetic algorithms) outperformed the other compared oversampling algorithms	

Ning et al., 2014 [12]	– International Journal of Computer Applications (0975 – 8887) Volume 154 – No.10, November 2016 15 – separate customers into two clusters based on the weight assigned by the boosting algorithm	"Implementation Zone" - customers with highest churn propensity sought

No research could be found which applies deep learning on this kind of data used for predictive analysis of customer churn in a telecom company. Unlike in this project, there were few or no occurrences of the application of the XG boost algorithm or k-nearest neighbors. Most of the previous research papers have taken up SVM, Random Forests, Decision Trees or other Boosting Algorithms used as Ensemble Methods as their prediction modelling methods, which could be because the newer algorithms in this project's model, like the Sequential deep learning algorithm and XG Boost emerged popularly after these researches were conducted.

Churn prediction models [13] & [14], have low accuracy and prediction. Hence a good prediction model is required to avoid churn. This can be achieved by combining algorithms like KNN, SVM, Decision Tree, Logistic Regression and Sequential Deep Learning with boosting algorithms for higher accuracy and performance which can be served by the models put together in this project

METHODOLOGY

After transforming the collected data in a form suitable for building ML models, the right methods to train machines & fine-tune the models were found. The best machine learning performers were selected. Overall scope of the work involved in building DL / ML-powered systems capable of forecasting customer churn is as follows:

1. Understanding the final goal of the problem and dataset collection
2. Data preparation and pre-processing
3. Splitting the dataset to train, validate and test separate sections
4. Modelling and testing
5. Pickling, model deployment and monitoring

The variables defined in this project's dataset describe whether a customer has an international or voice mail plan, total minutes spent and charge for calls during day, evening and night. The last column states if a customer has churned. A large amount of information represented in graphic form is easier to understand and analyze. That is why data visualization is important when we are dealing with larger datasets.

Data Pre-Processing

Dataset Encoding

Label Encoding refers to converting strings or labels into numeric or machine-readable form. It is an important pre-processing step for a structured dataset in supervised learning. The cat.codes method is used for encoding the training data & the label encoding method for testing data. After converting labels into categorical values, the order of columns of the dataset needs to be restored for correct functioning and operation of the dataset.

Scaling

Data Fitting and Transformation

fit_transform() function is used on the training data to scale it and learn the scaling parameters of that data. Here, the model will learn the mean and variance of the features of the training set. These learned parameters are then used for transformation or scaling the test data.

Data Normalization

Scaling is about converting numeric attributes that span different ranges so that they have same scale, such as between 0 and 1 for the smallest and biggest value for an attribute. Data scaling is applied on the churn datasets for better performance of the models and to reduce complexity.

Model Training and Evaluation

Numerous models are trained to define which one of them provides the most accurate predictions by running suitable accuracy and precision tests. To develop the simplest model able to formulate a target value fast and well enough, model tuning has been used.

Preprocessing, scaling, training, and all other experimentation was performed on the dataset with the help of several Numpy, Pandas and Matplotlib tools which facilitated a smooth remodeling and good comparative study between the ML models that were finalized after much trial testing and fine tuning.

	State	Account length	Area code	International plan	Voice mail plan	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Total eve calls
0	KS	128	415	No	Yes	25	265.1	110	45.07	197.4	99
1	OH	107	415	No	Yes	26	161.6	123	27.47	195.5	103
2	NJ	137	415	No	No	0	243.4	114	41.38	121.2	110
3	OH	84	408	Yes	No	0	299.4	71	50.90	61.9	88
4	OK	75	415	Yes	No	0	166.7	113	28.34	148.3	122

Figure 1. Snippet of the First Five Rows of 11 Columns of the Training Data

As seen in Figure 3.1, there are 20 attributes in the dataset, 16 of which are of the datatype Number - 8 Integer and 8 Float datatype; 3 attributes are String type and the last attribute, Churn is of Boolean datatype. The string and Boolean

attributes were categorized to yield a dataset suitable for application of ML & DL algorithms.

The models were tried for higher accuracy and better set of parameters to enhance their performance. Comparisons and tallies were made through heat maps, scatter plots, bar graphs; confusion matrices and classification reports were used to check the models' performance and improve predictive outcomes.

The KNeighborsClassifier was initially defined with n_neighbors = 1. After fitting the model and applying the predict() function, a classification report was prepared as follows:

	precision	recall	f1-score	support
0	0.90	0.93	0.91	572
1	0.46	0.38	0.42	95
accuracy			0.85	667
macro avg	0.68	0.65	0.66	667
weighted avg	0.84	0.85	0.84	667

Figure 2. Classification Report of KNN with n_neighbours = 1

A confusion matrix is a summary of prediction results on a classification problem. The number of correct and incorrect predictions are summarized with count values and broken down by each class. This is the key to the confusion matrix. It is confused when it makes predictions. A confusion matrix, heat map was created for the current KNN model, with actual and predicted labels on the two axes of the graph as follows.

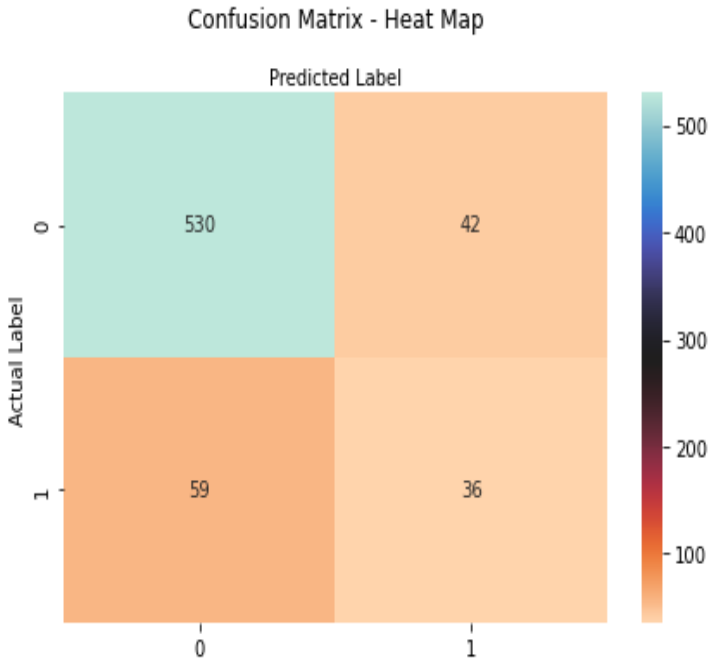

Figure 3. Confusion Matrix – Heat Map of KNN with n_neighbours = 1

Plotting a training and testing Accuracy vs n_neighbors plot (Figure 3.4) & checking for the training and testing accuracy for the algorithm with a new n_neighbors value everytime shows that the KNN Model gives maximum accuracy when the number of neighbors is 9.

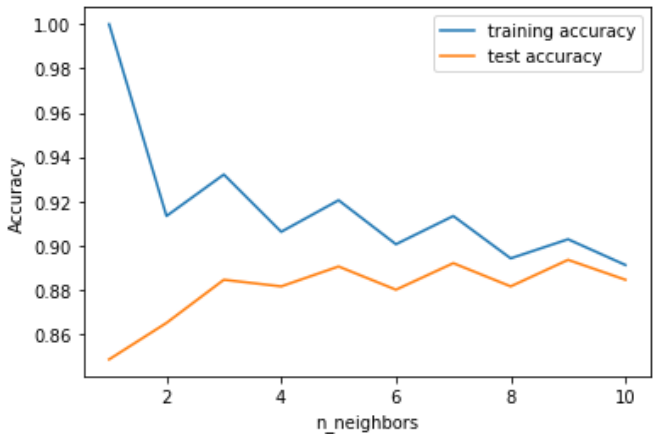

Figure 4. Accuracy vs. n_neighbor Line Plot in KNN

Corr() function was used to list out the correlation of all attributes to the Churn attribute. While the highest degree of correlation was that of International plan, 0.277489, lowest degrees of correlation went down to negative values, signifying the diversity of the collected data and the attributes recorded. corr() function was used to list out the correlation of all attributes to the Churn attribute. While the highest degree of correlation was that of International plan, 0.277489, lowest degrees of correlation went down to negative values, signifying the diversity of the collected data and the attributes recorded.

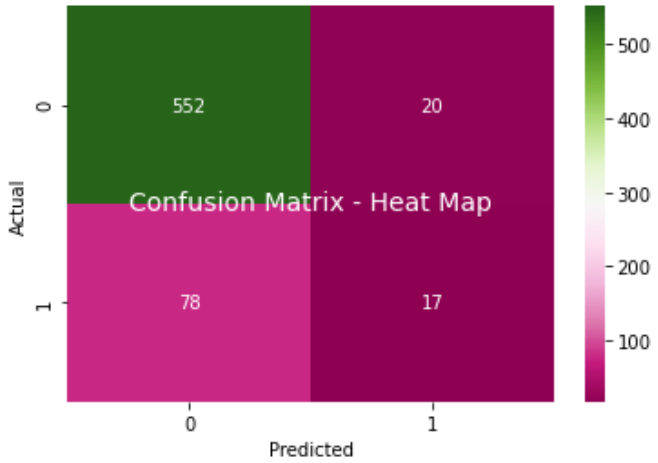

Figure 5. Confusion Matrix – Heat Map of Logistic Regression Model

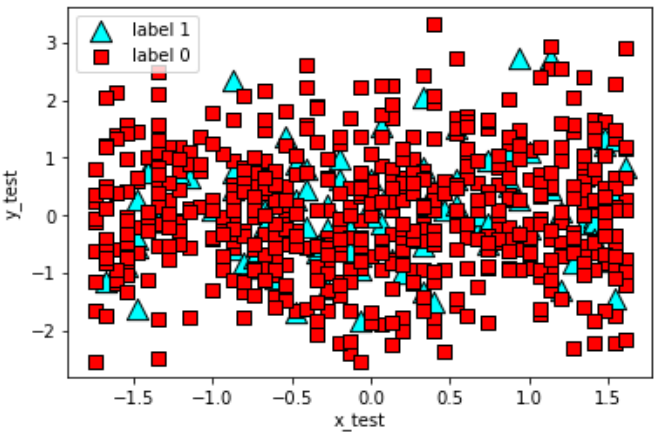

Figure 6. Scatter Plot on x_test and y_test Values

Figure 3.7 shows a horizontal bar graph depicting the importance of all the features. Certain parameters like n_estimators, random_state and probability of models like SVM and Random Forest Classifier were experimented with to bring out the highest accuracy of the model. The five previously created ML models were passed as estimators into the Voting Classifier with voting chosen as soft.

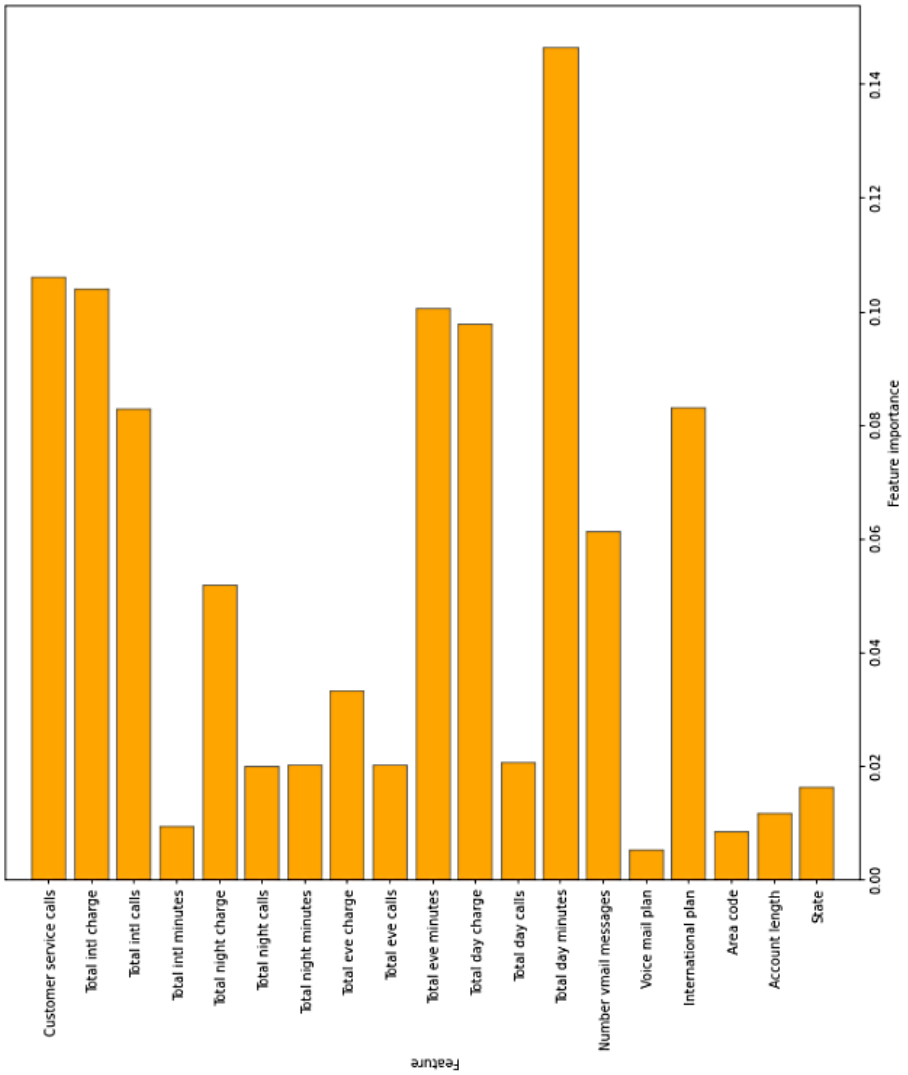

Figure 7. Feature Importance vs. Feature Bar Graph

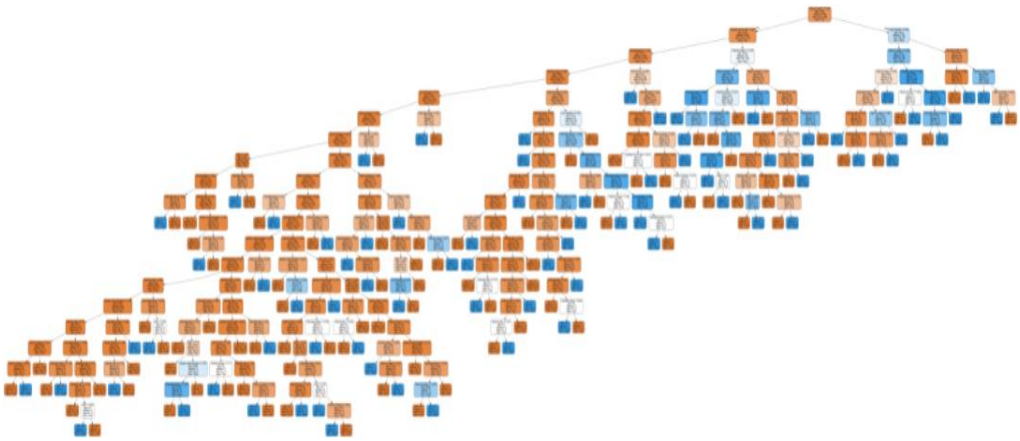

Figure 8. Full Decision Tree of the DT Classifier

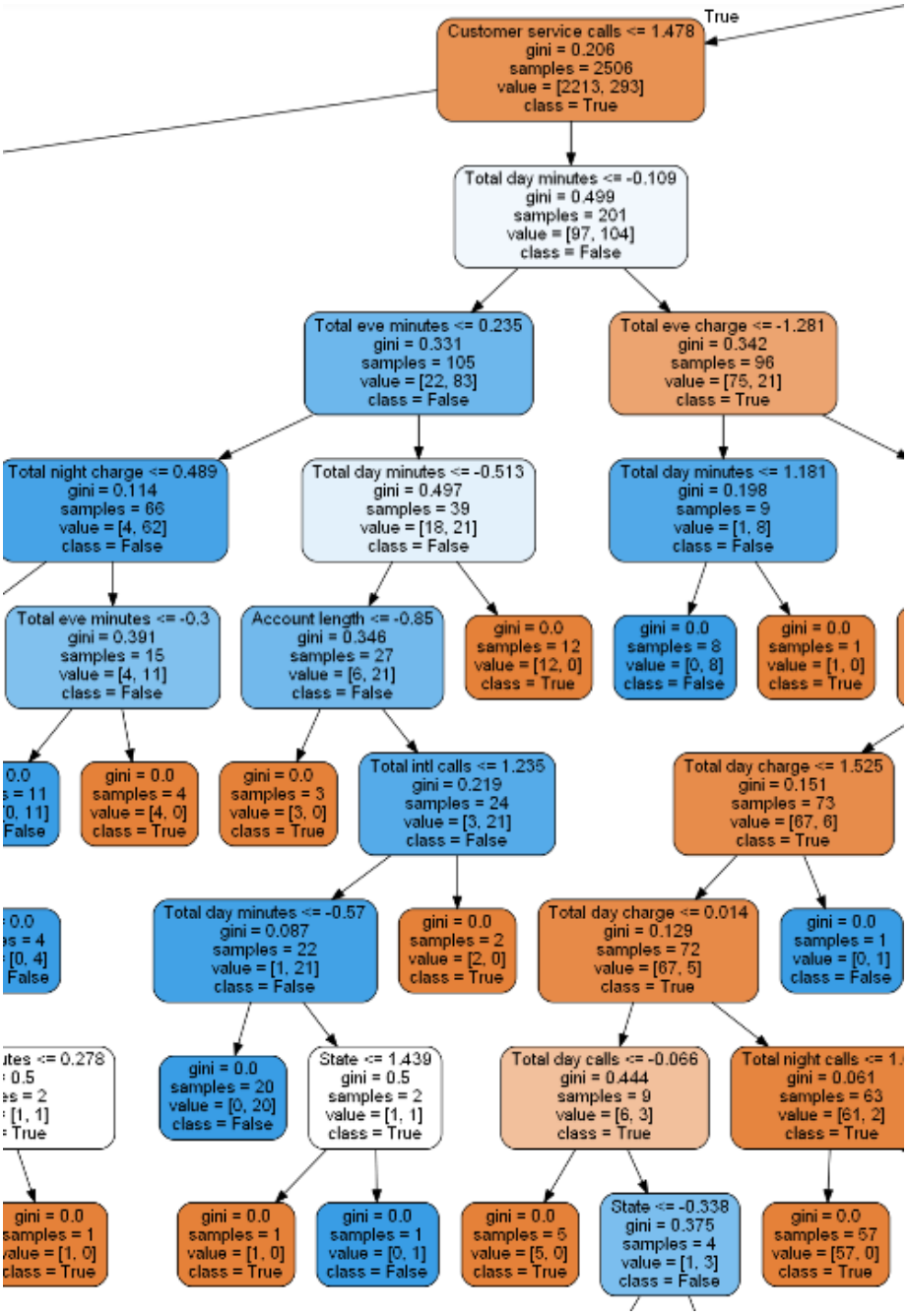

Figure 9. A Segment of the Decision Tree Produced by DT Classifier

The Decision Tree Classifier model (Fig 8 and Fig 9) was used to determine the Feature Importance of every attribute of the dataset.

	precision	recall	f1-score	support
0	0.96	0.99	0.97	572
1	0.91	0.76	0.83	95
accuracy			0.96	667
macro avg	0.94	0.87	0.90	667
weighted avg	0.95	0.96	0.95	667

```
[[565   7]
 [ 23  72]]
```

Figure 10. Classification Report and Confusion Matrix of XGB Classifier

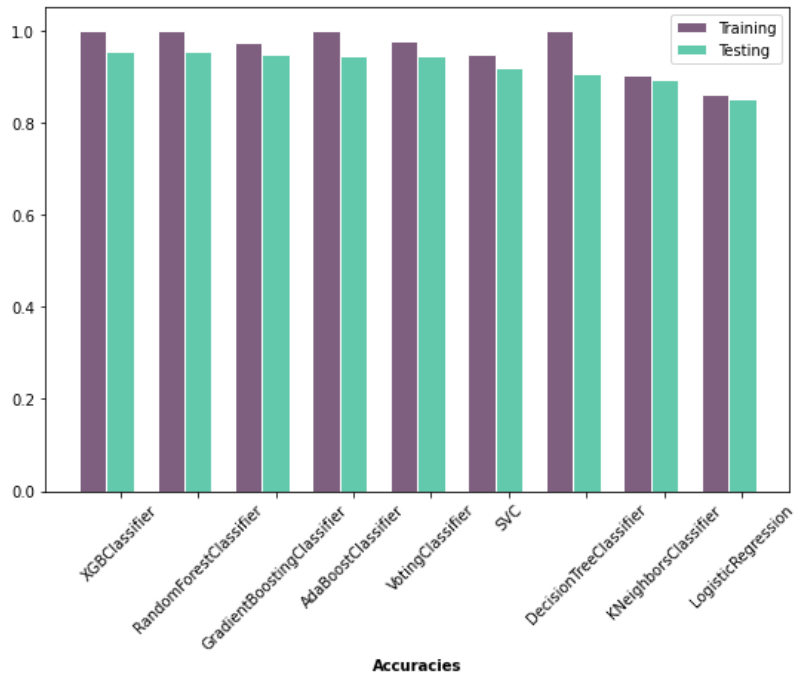

Figure 11. Bar Graph – Training and Testing Accuracies of the ML Models

Two layers and an output layer have been added in the deep learning sequential model. Standard layer type, Dense has been used, in which all nodes in the previous layer connect to the nodes in the current layer. After much experimentation, 20 neurons or nodes and 19 input parameters were passed in the first input layer of the model. The activation function used for the first two layers is ReLU. Sigmoid Activation Function has been used in the last layer. After changes in optimizers and loss functions, and many reruns, a suitable DL model was tailored for higher accuracy scores. Its training accuracy is 0.773 and the testing accuracy is 0.482.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 20)                400
_________________________________________________________________
dense_1 (Dense)              (None, 14)                294
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 15
=================================================================
Total params: 709
Trainable params: 709
Non-trainable params: 0
_________________________________________________________________
```

Figure 12. Sequential Deep Learning Model Designed for Churn Dataset

Feature importance determined the significance of each attribute in deciding whether churning would occur. Multiple sets of sample values were passed into the 10 models to ensure correct functioning. The various statistical results were analyzed to compare the performance of the models in predicting churn.

Pickle module of python and model.save of keras library were used to save the serialized ML & DL models to disk respectively for further loading. A webpage application layout was designed using Streamlit.
Network URL: http://192.168.1.104:8501
Webpage Link: https://customer-churn-pred-ml.herokuapp.com/

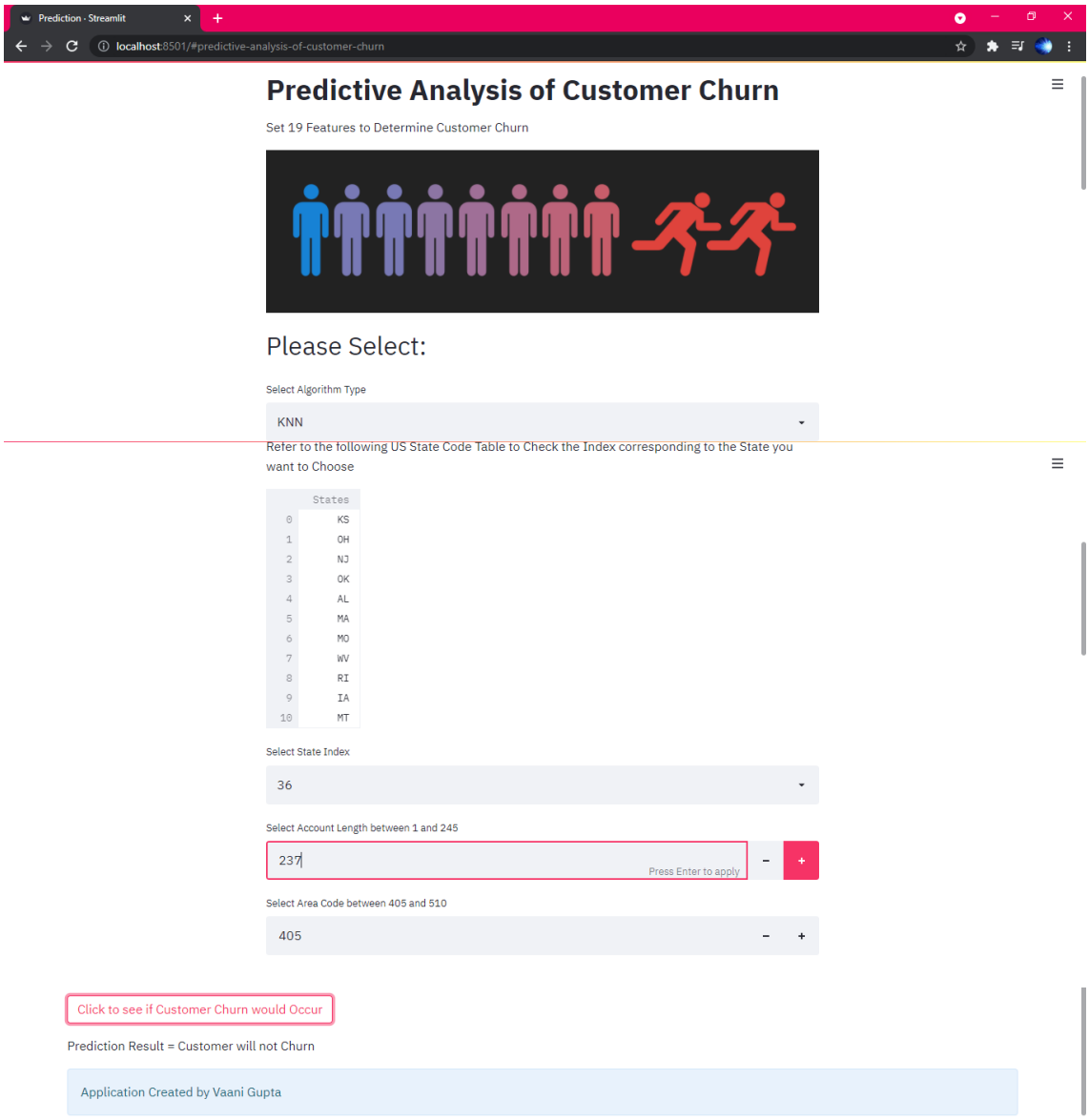

Figure 13. Heroku App Webpage for ML Models

CONCLUSION AND FUTURE SCOPE

Customer churn, an important factor to be considered for a company to make good business, isn't something many people who are not in the business sector realise the significance of. The attrition predictive models created in this project can help companies, mainly in the telecom sector, to have a profitable income and achieve good revenue consequently. The telecom company enjoying the benefits of this model can have a clear view of the problem and provide them exciting offers to stay in that service.

The obtained results show that the proposed churn model produces better results using the suite of machine learning techniques applied in this project instead of just using 3-4 algorithms as has been seen in previous works in this field. The XGB Classifier, Random Forests, Gradient Boosting and Ada Boost algorithms have proven to yield the best results due to maximum accuracy (95.5% & 95.35% respectively for the former two).

REFERENCES

A.H. Farquad, Vadlamani Ravi, and S. Bapi Raju 2014, Churn prediction using comprehensible support vector machine: An analytical CRM application. Applied Soft Computing 19: 31–40.

He Y, He Z, and Zhang D. 2009, A study on prediction of customer churn in fixed communication network based on data mining. Sixth international conference on fuzzy systems and knowledge discovery, vol. 1: 92–4.

Huang F, Zhu M, Yuan K, and Deng EO. 2015, Telco churn prediction with big data. ACM SIGMOD international conference on management of data: 607–18.

H. Lee, Y. Lee, H. Cho, K. Im, and Y.S. Kim 2011, Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model. Decision Support System 52: 207–216.

Brandusoiu I, Toderean G, and Ha B. 2016, Methods for churn prediction in the prepaid mobile telecommunications industry. International conference on communications: 97–100.

Burez D, and den Poel V. 2019, Handling class imbalance in customer churn prediction. Expert Syst Appl. 36(3): 4626–36.

T. Verbraken, W. Verbeke, and B. Baesens 2013, A novel profit maximizing metric for measuring classification performance of customer churn prediction models. IEEE Transaction on Knowledge and Data Engineering 25: 961–973.

Koen W. De Bock, and Dirk Van den Poel 2012, Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models. Expert Systems with Applications 39: 6816–6826.

Idris A, Khan A, and Lee YS. 2012, Genetic programming and adaboosting based churn prediction for telecom. IEEE international conference on systems, man, and cybernetics (SMC): 1328–32.

Makhtar M, Nafis S, Mohamed M, Awang M, Rahman M, and Deris M. 2017, Churn classification model for local telecommunication company based on rough set theory. J Fundam Appl Sci. 9(6): 854–68.

Amin A, Anwar S, Adnan A, Nawaz M, Howard N, Qadir J, Hawalah A, and Hussain A. 2016, Comparing oversampling techniques to handle the class imbalance problem: a customer churn prediction case study. IEEE Access 4: 7940–57.

L. Ning, L. Hua, L. Jie, and Z. Guangquan 2014, A customer churn prediction model in telecom industry using boosting. IEEE Trans. Ind. Inform. 10: 1659–1665.

Ahmad A.K., Jafar A., and Aljoumaa K. 2019, Customer churn prediction in telecom using machine learning in big data platform. J Big Data. 6, 28.

A. Saran and D. Chandrakala. 2016, A Survey on Customer Churn Prediction using Machine Learning Techniques. International Journal of Computer Applications. 154(10): 13-16.

Vadakattu, B. Panda, S. Narayan, and H. Godhia Nov 2015, Enterprise subscription churn prediction. Proc. IEEE Int. Conf. Big Data: 1317-1321.

Ascarza E, Iyengar R, and Schleicher M. 2016, The perils of proactive churn prevention using plan recommendations: evidence from a field experiment. J Market Res. 53(1): 46–60.

Chih-Fong Tsai, and Yu-Hsin Lu 2009, Customer churn prediction by hybrid neural networks. Expert Systems with Applications 36: 12547–12553.