Customer churn classification using machine learning techniques and comparing various classification methods

Adhithya Vattiam Ganesan, Kaushik Rajan, Sindhujan Dhayalan, Sushmit Kajalekar
X18105203, X17165849, X17170265, X18107001
School of computing
National college of Ireland

Abstract—**Abstract: The telecommunication sector is rapidly developing every day with the service providers offering a wide range of services and benefits to the customers for various reasons, but they also have a customer churn rate, i.e. the number of customers choosing to leave the services offered by one telecom service provider to another service provider. This directly affects the business of the company leading to huge loss. CRISP-Data Mining model is used to plan, understand and implement machine learning methods to predict the impact of the factors affecting customer retention and if a customer will stay or leave the company. To do the customer churn prediction, dataset from Kaggle is taken. Initially, the factors which has the highest correlation with the customer churn variable is determined. The customer dataset is taken and trained, then multiple algorithms such as Random Forest, Decision tree algorithms (CART & C5.0), Nave Bayes, KNN & ANN are applied to classify if the customer will leave or stay with the company. All the used models are evaluated and compared based on accuracy and ROC curve. the impact of various factors leading to customer churn rate and compared the models for churn prediction accuracy. C5.0 approach was found to be the best fit for churn prediction with easy implementation and best fit for the problem.**

Index Terms—**Random forest, C5.0, KNN, ANN, Naive Bayes, Decision tree, SVM.**

I. Introduction

Telecommunication sector is at its peak with the advancement of technologies, they offer a wide range of services from internet, mobile services and cable subscriptions all together as a bundle or one can select the services separately based on preference. This sector has a lot of players and each of them with the intention of trying to control the market share provide offers to lure the customers to change to their service providers for better services and offers. This movement of customers from one service provider to another service provider is known as customer churn rate.

This movement of customers affects the service providers in many ways, first this tarnishes the reputation of the company in the market creating a bad image of the service provided. Then secondly this leads to a business and financial loss. The service providers can cope up with business loss to some extent but once the reputation of the company is gone then it becomes irrevocable, which no organisation would want to face. Most of the telecommunication companies keep losing customers to competitors from time to time. The customers that stopped using the company's product or service during a certain time frame is known as customer churn [1]. On multiple instances it is observed that retaining present clients is a whole lot profitable and cheaper than attracting an entirely new client. Hence, customer churn prediction is turning into the top concerns that many organizations dedicate their time and assets to deal with it.

The service provider analyses the reason for a customer to discontinue their services to improve their business strategies, it is estimated that the cost of acquiring a new customer is 6-7 times more expensive than the cost of retaining an existing customer. Hence the service providers spend time analysing the reasons for churn, this may be due to dissatisfaction, cheaper or more affordable services provided by the competitors or better sales and marketing strategies employed by the competitors. Each of which indicating the areas where the company falls behind its competitor in the market [2].

In this paper we analyse the most commonly applied machine learning techniques for customer churn prediction by applying Decision Tree algorithms, K-Nearest Neighbour algorithm, Random Forest algorithm and Artificial neural network to predict the customer churn rate using telecommunication customer data and then compare the models based on their prediction accuracy, error rate and the ease of performing the classification. The objective of the paper is to find the best model for Telecom customer churn prediction.

II. Literature Review

The classification problem associated with customer churn prediction system can be fixed by using decision tree method. The information theory is used to create information nodes, branches as per the various values of different aspects[3]. It examines the statistics of capabilities from large instances, calculate the entropy of particular feature, and locate the maximum essential function for category. For over 13000 records in a customer churn data, the decision tree method showed 88.53% accuracy for prediction[4].

The imbalance between customer churn data and predicting applicable solutions on it often results in poor performance of learning and also it is difficult for machine learning algorithms[5]. The other classifier is c5.0 which gives good performance over the large data set. It performs constantly when variety of variables is large. The boosting technique is carried out in this set of rules to enhance the precision. The other purpose is that the C5.0 selection tree can overcome the scatter feature of the telecommunication information. This algorithm solves few problems such as low rate of churn, scatter problem. It provides sampling method which removes noise, redundancy and performs separate clustering. Precision ratio is 80% and prediction results are satisfactory enough to retain the potential churn customers[6].

In case of the company which is recently established or newly adopted some different technology or has lost the historical information associated with the customers, the traditional churn prediction systems wont be useful. In such cases JIT(Just-In-Time) approach is more practical alternative. This CCP process categories customers as churn and non-churn according to the probability of a customer activity in a nearer future. The four important aspect of this approach are behaviour of customer, the loyalty of customer with the company, customers personal details and some external factors. This model follows a Nave Bayes approach which calculates posteriori probabilities[7].

Another approach is based on information processing mechanism. It is primarily based on a large series of different neural networks. The structure of these network is determined by the inter connection between them. These networks are trained to achieve desirable outcome. Their maximum excellent characteristic is their capability to research automatically from the information so that one can provide a method for predictions. Hence these models are used to build efficient, reliable and high accuracy customer churn prediction system[8].

Support Vector Machines are used for minimizing the risks. It uses the kernel method to draft the data to a 2D or 3D space. The data is divided in linear manner so that further classification can be done. In this paper, an approach for churn prediction by using SVM is discussed. To gain more accurate results and performing, a grid search manner along with the SVM was used. The SVM for customer churn prediction method by using SVM provided 98.7% accuracy[9].

The analysis of the imbalanced data for the churn prediction based on various predictive model with respect to KNN classifier. The ratio of non-churn customers compared to churn clients is only 2 percent, therefore lowering the imbalance of the instructions before enforcing the models may be very critical. Another important issue raised due to ratio imbalance is over sampling resulting in redundancy. This is handled by the algorithm SMOTE (Synthetic Minority Oversampling Technique). In the crucial aspect of model selection, the implementation of various model like AdaBoost, Extra Trees, kNN, Neural Network and XGBoost[10] is compared. The study is carried out based on the different resources provided by one of the internet service providers of leading Telecom Company. Usage based on numerous statistical assessments; 121 most relevant functions had been decided on to use for training. these attributes were accumulated over 365 days and examined the seasonal effect. The non-parametric lazy learning algorithm used to carry out the classification is KNN known as k-Nearest Neighbour. The algorithm takes all the nearer cases in training set into the consideration using measures like Euclidian distance. This classifier is to understand statistical analysis of customer churn and pattern recognition. XGBoost feature importance score is used to understand the top number of customer variable and best fit model. In the similar way, maximum appropriate top functions for AdaBoost, more trees, KNN and Neural network version are determined by evaluating their performances. The important aspect of this study is feature engineering on modelling phase and providing methodology which can handle imbalanced information in churn prediction system[11].

Customer churn analysis, is an important factor to be considered while analysing company profits and revenues. The only challenge in case of churn analysis is low rate of churned customers in entire data which causes imbalance data information and data sets. The paper put forth two approaches to analyse the imbalanced data. One is sampling approach and another one is known as cost-sensitive approach. The motive of this analysis is to address the imbalanced records in the dataset used in this study, so that it can produce greater performance in churn prediction model. The dataset is processed by the sampling approach and WRF classifier classification used on the data produces inadequate results hence sampling techniques are used to minimize the unbalancing between the classes churn and non-churn. SMOTE[12] is used to increase the probability of fetching the data from churn class. The two methods mentioned earlier to reduce imbalanced data between the classes, uses many classifiers to solve this issue such as, logistic regression, linear classification, nave Bayes, decision tree, multi-layer perception neural networks, support vector machine, data mining evolutionary algorithm and random forest. The paper proposes the classification algorithm along with WRF and sampling which is used to generate multiple decision trees from the training data and provide the churn result with precision. The paper tests proposed model and data with 10 cross validations to obtain result with maximum precision. Which proves this model provides the churn analysis with reliable accuracy and can assist in reducing potential customer churn[5].

III. Methodology

CRISP DM Model has been used for this Data Mining project. Cross-Industry Standard Process for Data Mining is a 6-phase constructed approach used for data mining project.

A. Business Understanding

This is the initial stage of the Data Mining project. Information regarding the customer churn and details on the determined outputs are decided.

Business Questions:

1) Which variables have the highest impact on the customer retention?

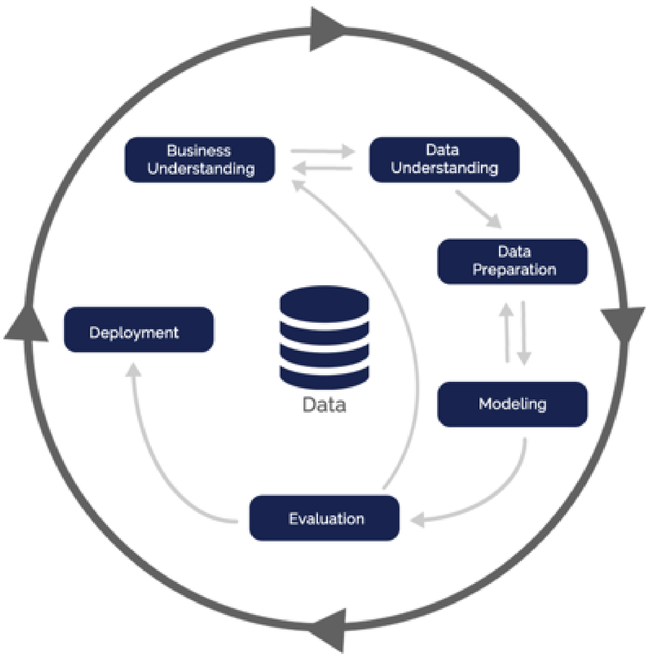

Fig. 1. CRISP DM Model

2) Which model performs the best at classifying if the customer will leave or stay?

B. Data Understanding

For this project, Customer churn dataset created on 2018-02-23 is sourced Kaggle[1]. The dataset consists of 21 attributes and details with the datatype of each variable is given in figure2.

Attribute name	Description	Datatype
customerID	Unique ID of the customer	String
gender	Gender of the customer (Male/Female)	String
SeniorCitizen	The customer is senior citizen or not (0/1)	Numeric
Partner	The customer has a partner or not (Yes/No)	String
Dependents	The customer has dependents of not (Yes/No)	String
tenure	Number of months with the company	Numeric
PhoneService	The customer has phone service or not (Yes/No)	String
MultipleLines	The customer has multiple services or not (Yes/No/No Phone)	String
InternetService	Customer's internet service provider (DSL/Fiber optic/No)	String
OnlineSecurity	The customer has online security or not (Yes/No/No internet service)	String
OnlineBackup	The customer has online backup or not (Yes/No/No internet service)	String
DeviceProtection	The customer has device protection or not (Yes/No/No internet service)	String
TechSupport	The customer has tech support or not (Yes/No/No internet service)	String
StreamingTV	The customer has streaming service or not (Yes/No/No internet service)	String
StreamingMovies	The customer has streaming movie services or not (Yes/No/No internet service)	String
Contract	Customer contract type (Month-to-month, One year, Two year)	String
PaperlessBilling	The customer has paperless billing or not (Yes/No)	String
PaymentMethod	Payment method type (Bank transfer, Electronic check, Credit card, Mailed check)	String
MonthlyCharges	Monthly charges for the customer	Numeric
TotalCharges	Total charges for the customer	Numeric
Churn	Did the customer stay or leave? (Yes/No)	String

Fig. 2. Dataset information

C. Data Preparation

Data Preparation and pre-processing is the most important stage of any data mining project. In this stage, the dataset is checked for outliers, missing values and also if the attributes are in proper datatype. Following steps were done in order to convert the data to the stage where it is suitable for model implementation:

1) The datatypes of few attributes were changed. The categorical variables were converted into factors.
2) The dataset was checked for outliers. Interquartile range rule was used and the range value is set based on the Q1 and Q3. The IQR rule was used on the tenure attribute.
3) Missing values were replaced with the mean of the attribute instead of removing the row.
4) CustomerID was removed as it was not used for prediction.
5) The variables were normalized using Z-score normalization. The normalization was done using scale() function in R.

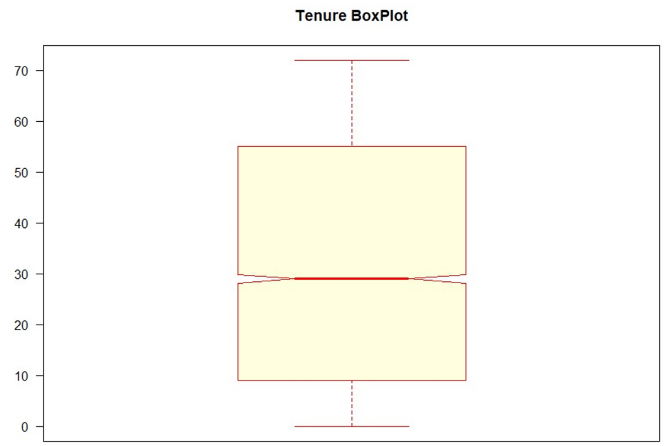

Fig. 3. Boxplot of Tenure

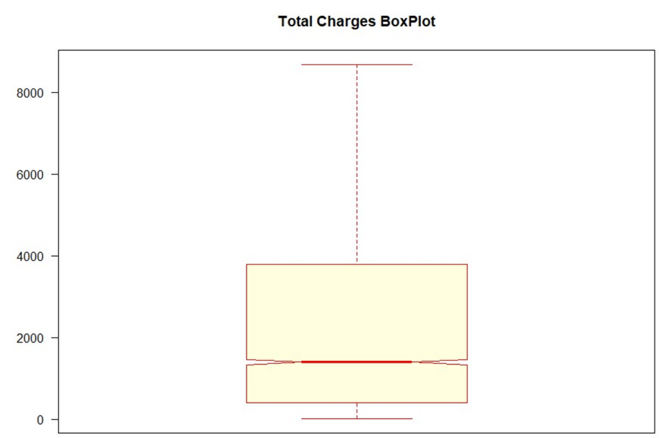

Fig. 4. Boxplot of Totalcharges

D. Modelling and Evaluation of the models

After the data was cleaned and converted as per the requirement, we split the dataset into training and test data in order to implement the models. The models learn from the training set and implement on the test set. The models that are implemented are Random forest, C5.0, K-Nearest Neighbour, Decision Tree and Artificial neural network.

Random Forest: Random forest is an ensemble of decision tree which can be used for regression and classification. The

[1]https://www.kaggle.com/blastchar/telco-customer-churn/metadata

main advantage of random forest is that it does not produce overfit results and also produces great accuracy. We have used random forest to classify if the customer will churn or not. The confusion matrix is given in figure5.

```
> table(Real=test[,20],Predict)
    Predict
Real    0    1
   0   10  457
   1   17 1277
> # Accuracy rate - Random Forest
> ACC <- (test[,20]== Predict) *100
> RF_accuracy <- sum(ACC)/length(ACC)
> RF_accuracy
[1] 73.08348
> # Error rate - Random Forest
> error<- (test[,20]!=Predict )
> errorRate<-sum(error)/length(error)
> errorRate
[1] 0.2691652
> #Precision, Recall & F1 values
> precision <- posPredValue(Predict, test[,20], positive="1")
> precision
[1] 0.7364475
> recall <- sensitivity(Predict, test[,20], positive="1")
> recall
[1] 0.9868624
> F1 <- (2 * precision * recall) / (precision + recall)
> F1
[1] 0.843461
```

Fig. 5. Random forest model Results- Rcode

The Random forest produced results with 73.08% accuracy and with the error rate of 0.27. The area under the curve value is 0.63 which falls in the D category,The ROC of the model is displayed in figure6.

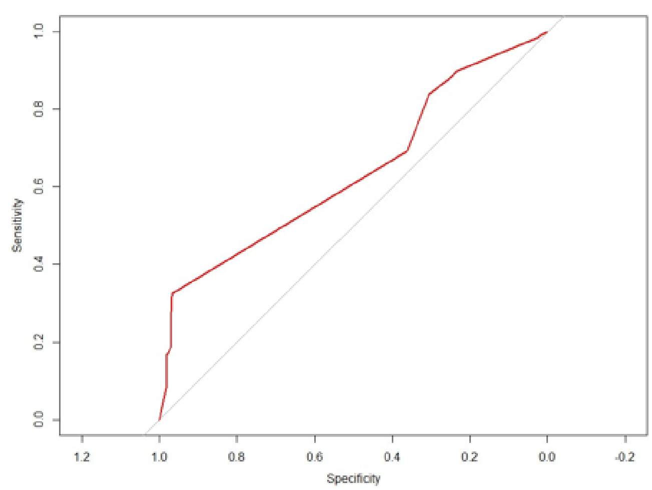

Fig. 6. ROC of Random forest model

Business question 1:

Random forest to find the importance the variables in classifying the customer churn variable:
We use random forest to determine the importance of each of the variable and to find which variable has the highest correlation with the customer churn variable.

From the above plot we can infer that the variables Tenure, Total Charges, contract and monthly charges are the most important variables which are used to classify the customer churn.

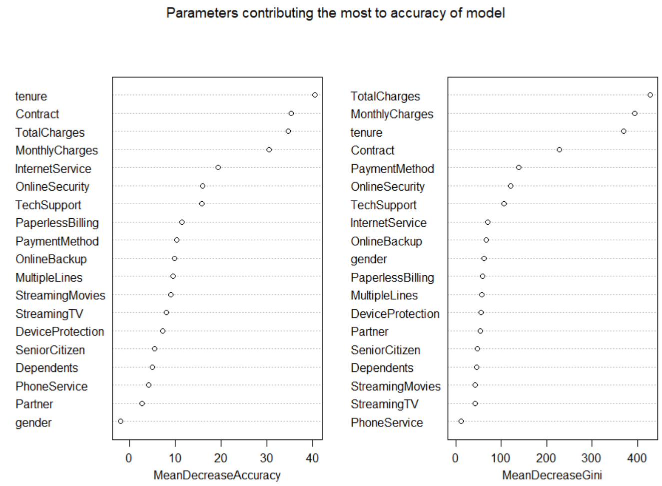

Fig. 7. Parameters contributing the most to accuracy of the model

C5.0: C5.0 is an updated version of C4.5 algorithm which can produce 2 kinds of models, namely, decision tree and rule set. The C5.0 can also find the attribute relevance. We have implemented the C5.0 algorithm by using Tenure, Total Charges, Contract and monthly charges as independent variables. The model is evaluated based on the value of accuracy, error rate, precision, recall and AUC value.

```
> table(actual=test[,20],c50=c50Predict)
      c50
actual    0    1
     0  213  254
     1  150 1144
> # Accuracy rate - c5.0
> ACC <- (test[,20]== c50Predict) *100
> c50_accuracy <- sum(ACC)/length(ACC)
> c50_accuracy
[1] 77.05849
> plot(c50)
> # Error rate - c5.0
> error<- (test[,20]!=c50Predict)
> errorRate<-sum(error)/length(error)
> errorRate
[1] 0.2294151
> #Calculating Precision, Recall $ F1
> precision <- posPredValue(c50Predict, test[,20], positive="1")
> recall <- sensitivity(c50Predict, test[,20], positive="1")
> F1 <- (2 * precision * recall) / (precision + recall)
> precision
[1] 0.8183119
> recall
[1] 0.8840804
> F1
[1] 0.8499257
```

Fig. 8. C5.0 Results - Rcode

The accuracy was found to be 77.05% with error rate of 0.22. The values of precision and recall were found to be 0.81, 0.88 and the F1 value is 0.84.AUC value was found to be 0.79 which falls under C category.Model of C5.0 and ROC curve is displayed in figure9 and figure10 respectively.

NAIVE BAYES Naive Bayes algorithm is based on the Bayes theorem of probability distribution. Nave Bayes was implemented and the accuracy was found to be 69.90. The confusion matrix is displayed in figure11.

The error rate was 0.30 and precision, recall and F1 values are 0.88, 0.67 and 0.76 respectively.The value of area under

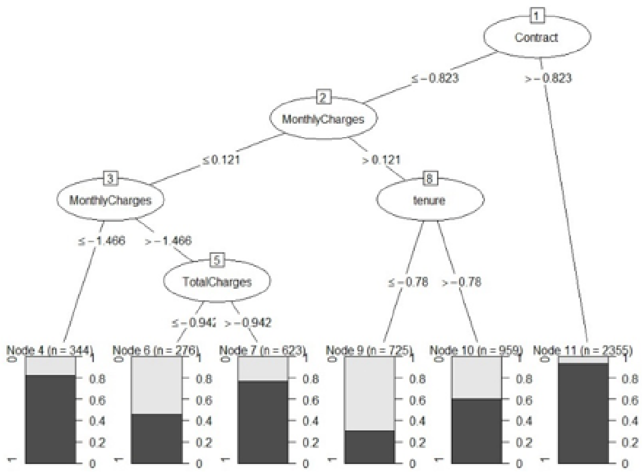

Fig. 9. C5.0 Model

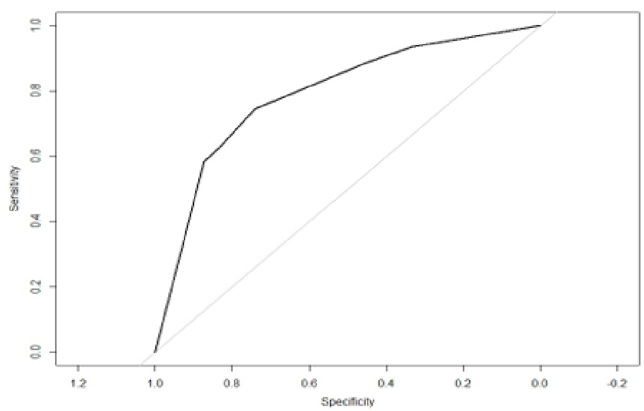

Fig. 10. C5.0 Model ROC curve

```
> table(NB=NBPredict,Class=test$Churn)
   Class
NB    0   1
  0 359 422
  1 108 872
> # Accuracy rate - NaiveBayes
> ACC <- (test[,20]== NBPredict) *100
> NBaccuracy <- sum(ACC)/length(ACC)
> NBaccuracy
[1] 69.90346
> # Error rate - NaiveBayes
> error<- (test[,20]!=NBPredict)
> errorRate<-sum(error)/length(error)
> errorRate
[1] 0.3009654
> #Calculating Precision, Recall $ F1
> precision <- posPredValue(NBPredict, test[,20], positive="1")
> precision
[1] 0.8897959
> recall <- sensitivity(NBPredict, test[,20], positive="1")
> recall
[1] 0.6738794
> F1 <- (2 * precision * recall) / (precision + recall)
> F1
[1] 0.7669305
```

Fig. 11. Naive Bayes Results RCode

the curve was found to be 0.79 which falls under C category. The ROC curve of the model is displayed in figure12.

Decision Tree: Decision tree is a graphical representation of all the possible solution to a decision. The main advantage

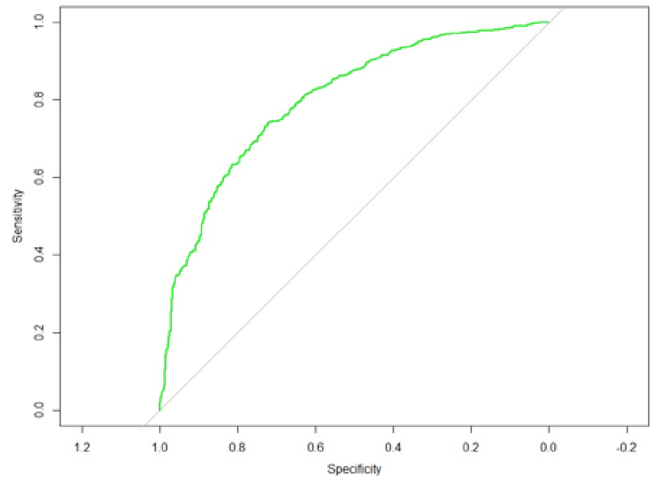

Fig. 12. ROC Curve of Naive Bayes model

of a decision tree is that it can be easily explained. We have implemented the Decision tree algorithm and got the following results. The confusion matrix is given below in the figure13. The accuracy, error rate, precision, recall and F value were

```
> table(actual=test[,20],DT=DTPredict)
      DT
actual    0    1
     0  215  252
     1  136 1158
> # Accuracy rate - Decsion Tree
> ACC <- (test[,20]== DTPredict) *100
> accuracy <- sum(ACC)/length(ACC)
> accuracy
[1] 77.96706
> # Error rate - Desicion Tree
> error<- (test[,20]!=DTPredict)
> errorRate<-sum(error)/length(error)
> errorRate
[1] 0.2203294
> #Calculating Precision, Recall $ F1
> precision <- posPredValue(DTPredict, test[,20], positive="1")
> precision
[1] 0.8212766
> recall <- sensitivity(DTPredict, test[,20], positive="1")
> recall
[1] 0.8948995
> F1 <- (2 * precision * recall) / (precision + recall)
> F1
[1] 0.8565089
```

Fig. 13. Decision Tree Results Rcode

found to be 77.96, 0.22, 0.82, 0.89 and 0.85 respectively.The area under the curve value was found to be 0.79 which falls under the C category.The Model of decision tree and ROC curve is displayed in figure14 and figure15.

K-Nearest Neighbour: It is a supervised machine learning classification model. The main advantages if KNN are it is simple and effective and does not have any assumption. The disadvantages of KNN are that it requires large amount of memory and does not work well if the dataset has missing values. We have implemented the KNN algorithm with K = 1,5,10 and 30. The results obtained show the accuracy increases with increase in K value. Figure16, Figure17 and Figure18 show Rcode implementation of KNN model for K value= 1, 5 and 30 respectively.

We can see that the KNN produces highest accuracy of 78.30% when the K value is 30.Comparison of Performance

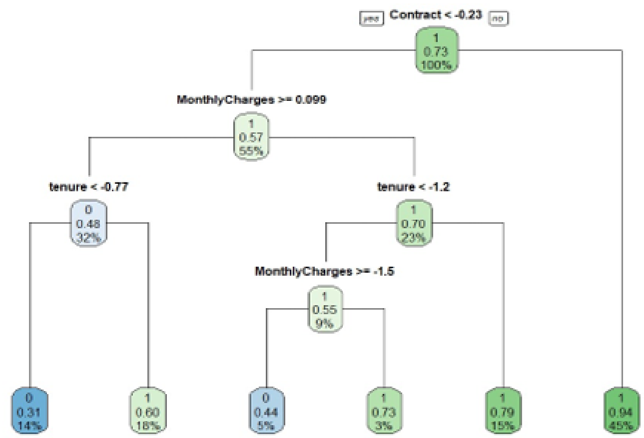

Fig. 14. Decision Tree Model

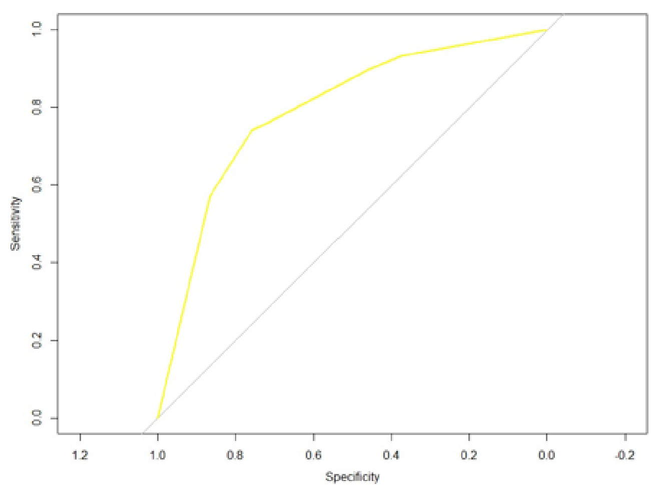

Fig. 15. Decision Tree ROC curve

K = 1

```
> confusionMatrix
          Actual
Prediction    0    1
         0  233  223
         1  234 1071
> ACC <- sum(test[,20]==KNN1)/nrow(test)*100
> ACC
[1] 74.04884
> #Calculating Precision, Recall $ F1
> precision <- posPredValue(KNN1, test[,20], positive="1")
> precision
[1] 0.8206897
> recall <- sensitivity(KNN1, test[,20], positive="1")
> recall
[1] 0.8276662
> F1 <- (2 * precision * recall) / (precision + recall)
> F1
[1] 0.8241631
```

Fig. 16. KNN Model with K = 1 - Rcode

measure when values of K are 1,5,10 and 30 are displayed in figure19.

Support vector machine (SVM): SVM is a supervised method which can be used for classification and regression.

K = 5

```
> confusionMatrix5
          Actual
Prediction    0    1
         0  240  191
         1  227 1103
> ACC5 <- sum(test[,20]==KNN5)/nrow(test)*100
> ACC5
[1] 76.26349
> #Calculating Precision, Recall $ F1
> precision <- posPredValue(KNN5, test[,20], positive="1")
> precision
[1] 0.8293233
> recall <- sensitivity(KNN5, test[,20], positive="1")
> recall
[1] 0.8523957
> F1 <- (2 * precision * recall) / (precision + recall)
> F1
[1] 0.8407012
```

Fig. 17. KNN Model with K = 5 - Rcode

K = 30

```
> confusionMatrix30
          Actual
Prediction    0    1
         0  217  132
         1  250 1162
> ACC30 <- sum(test[,20]==KNN30)/nrow(test)*100
> ACC30
[1] 78.30778
> #Calculating Precision, Recall $ F1
> precision <- posPredValue(KNN30, test[,20], positive="1")
> precision
[1] 0.8229462
> recall <- sensitivity(KNN30, test[,20], positive="1")
> recall
[1] 0.8979907
> F1 <- (2 * precision * recall) / (precision + recall)
> F1
[1] 0.8588322
```

Fig. 18. KNN Model with K = 30 - Rcode

K Model Value	Accuracy	Error	Precision	Recall	F1
KNN1	74.0488358886996	0.223736513344691	0.820689655172414	0.827666151468315	0.824163139669104
KNN5	76.2634866553095	0.223736513344691	0.829323308270677	0.852395672333848	0.840701219512195
KNN10	78.0238500851789	0.223736513344691	0.828861493836113	0.883307573415765	0.855218855218855
KNN30	78.3077796706417	0.223736513344691	0.822946175637394	0.897990726429675	0.858832224685883

Fig. 19. Comparison of KNN model with different K values

SVM is good at handling non-linear and complex dataset. The only disadvantage of SVM is that it is a Blackbox method. We have implemented SVM and obtained the following results. The Confusion matrix is displayed in figure20.

The accuracy, error rate, precision, recall and F value were found to be 78.13%, 0.21, 0.83, 0.87 and 0.85 respectively.

Artificial Neural Network: ANN is a supervised algorithm which is typically used for regression and classification. The main disadvantages are it is difficult to understand as it is a black box method and also takes long time for processing. ANN was implemented and the results displayed in figure21 obtained. Accuracy of ANN was found to be 79%

The Model of ANN and ROC Curve is displayed in figure22 and figure23 respectively.

```
> table(actual=test[,20],Predict)
      Predict
actual    0    1
     0  240  227
     1  158 1136
> # Accuracy rate - SVM
> ACC <- (test[,20]== Predict) *100
> accuracy <- sum(ACC)/length(ACC)
> accuracy
[1] 78.13742
> # Error rate - SVM
> error<- (test[,20]!=Predict )
> errorRate<-sum(error)/length(error)
> errorRate
[1] 0.2186258
> #Calculating Precision, Recall $ F1
> precision <- posPredValue(Predict, test[,20], positive="1")
> precision
[1] 0.8334556
> recall <- sensitivity(Predict, test[,20], positive="1")
> recall
[1] 0.877898
> F1 <- (2 * precision * recall) / (precision + recall)
> F1
[1] 0.8550997
```

Fig. 20. SVM model Results - RCode

```
> ErrorRate
[1] 0.2084043157
> accuracy <- 1 - ErrorRate
> accuracy
[1] 0.7915956843
```

Fig. 21. ANN - Results

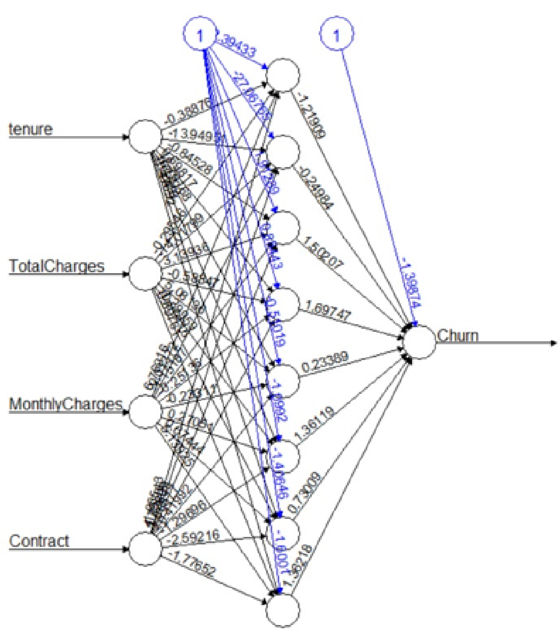

Fig. 22. ANN - Model

Business question 2:
Which model produce better results?

- Accuracy of Random Forest - 73.08%
- Accuracy of C5.0 - 77.05%

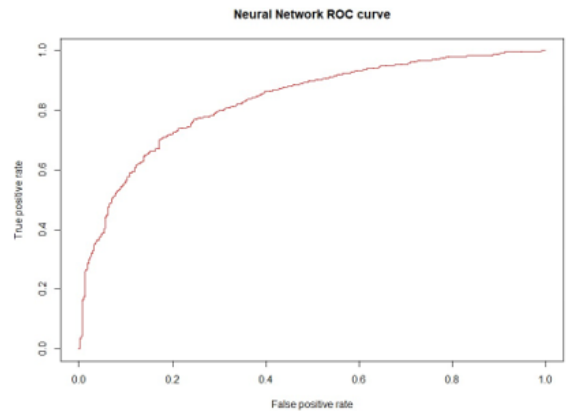

Fig. 23. ANN - ROC Curve

- Accuracy of Decision Tree 77.96%
- Accuracy of Naive Bayes - 69.90%
- Accuracy of KNN - 78.30%
- Accuracy of SVM - 78.13%
- Accuracy of ANN - 79.15%

ANN produces better accuracy compared to other models. Model comparison of Random Forest,Decision tree(CART),C5.0 and Naive Bayes is displayed in figure24

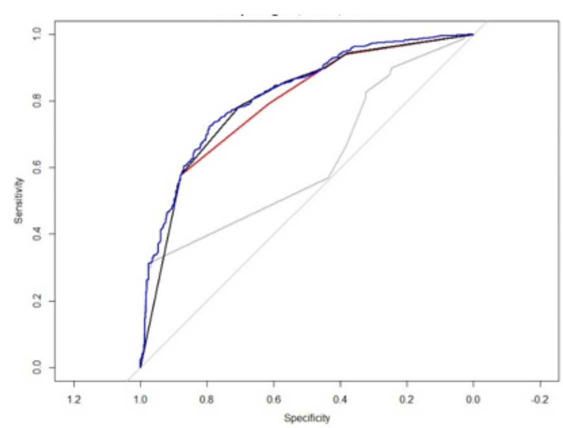

Fig. 24. ROC Curve comparison of RF,CART,C5.0 and NB

E. Deployment

This is the final phase of a CRISP DM Model where the models are deployed in real life scenarios.

IV. Conclusion and Future work

In this project, we have classified the customer churn by using machine learning algorithms such as random forest, C5.0, Decision tree, KNN, ANN and SVM. We have also found the importance of each variable in contribution to the classification of customer churn variable. In future, we can work with a larger dataset and also use various other attributes to do the same task of classifying the customer churn.

References

[1] S. Amaresan, "What is customer churn?" 2019. [Online]. Available: https://blog.hubspot.com/service/what-is-customer-churn

[2] M. Karanovic, M. Popovac, S. Sladojevic, M. Arsenovic, and D. Stefanovic, "Telecommunication services churn prediction-deep learning approach," in *2018 26th Telecommunications Forum (TELFOR)*. IEEE, 2018, pp. 420–425.

[3] H. Ma, M. Qin, and J. Wang, "Analysis of the business customer churn based on decision tree method," in *2009 9th International Conference on Electronic Measurement & Instruments*. IEEE, 2009, pp. 4–818.

[4] F. Guo and H.-L. Qin, "The analysis of customer churns in e-commerce based on decision tree," in *2015 International Conference on Computer Science and Applications (CSA)*. IEEE, 2015, pp. 199–203.

[5] V. Effendy, Z. A. Baizal *et al.*, "Handling imbalanced data in customer churn prediction using combined sampling and weighted random forest," in *2014 2nd International Conference on Information and Communication Technology (ICoICT)*. IEEE, 2014, pp. 325–330.

[6] H. Li, D. Yang, L. Yang, X. Lin *et al.*, "Supervised massive data analysis for telecommunication customer churn prediction," in *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom)*. IEEE, 2016, pp. 163–169.

[7] A. Amin, B. Shah, A. M. Khattak, T. Baker, S. Anwar *et al.*, "Just-in-time customer churn prediction: With and without data transformation," in *2018 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2018, pp. 1–6.

[8] S. H. Dolatabadi and F. Keynia, "Designing of customer and employee churn prediction model based on data mining method and neural predictor," in *2017 2nd International Conference on Computer and Communication Systems (ICCCS)*. IEEE, 2017, pp. 74–77.

[9] A. Rodan, H. Faris, J. Alsakran, and O. Al-Kadi, "A support vector machine approach for churn prediction in telecom industry," *International journal on information*, vol. 17, no. 8, pp. 3961–3970, 2014.

[10] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, 2016, pp. 785–794.

[11] D. Do, P. Huynh, P. Vo, and T. Vu, "Customer churn prediction in an internet service provider," in *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 2017, pp. 3928–3933.

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.