DEVELOPMENT OF A CLASSIFICATION MODEL FOR THE PREDICTION OF CHURN AMONG CUSTOMERS USING DECISION TREES ALGORITHM

Presented by

BALOGUN, Jeremiah Ademola

Department of Computer Science and Mathematics

MOUNTAIN TOP UNIVERSITY, OGUN STATE, NIGERIA

@

NCS INTERNATIONAL CONFERENCE ON

INNOVATIVE SYSTEMS FOR DIGITAL ECONOMY (ISDE 2021)

August 5, 2021

INTRODUCTION

- According to Machowska (2018), customer churn or attrition occurs whenever a customer switches from the services or goods of one company to another or leaves a given market altogether.

- It is worthy of note that the loss of customers can lead to substantial economic losses in business because a strong relationship between organizations and customers is required for generating long-term profit (Hamilton, Rust, & Dev, 2017).
  - Every business is expected to have a very good understanding of its customer's needs.

- Customers are one of the company's most valuable assets, hence:
  - Businesses need to provide customers incentives in order to retain and continue to grow the business (Abbasi, Tarhini, Hassouna, & Shah, 2015).
  - However, any negative interaction with a customer is an indication that a customer may likely churn (Amin, Khan, Ali, & Anwar, 2014).

INTRODUCTION…

- The ability to effectively manage customer churn makes it possible to acquire precise and real-time information about customers who are likely to quit using a company's service (Hu, Shu, & Qiao, 2014).

- Predicting a customer's potential to churn in advance, provides valuable insight needed for retaining and increasing customer base (Ballings & Van, 2012).
  - As such, churn prediction models are required for monitoring customer relationship management in order to preserve the customers that intend to quit a service.

- Searching and identifying customers who show an inclination to abandon the services of a company is of crucial importance to customer-oriented retention strategy (Lessman & Vob, 2009; Blattberg, Kim, & Neslin, 2010).
  - Customers can churn for various reasons, and churn can happen at any time.
  - Hence, it is important that these decisions are data-driven whenever it is available.

INTRODUCTION…

- In customer retention management, it is important to know why customers are leaving the company to better act on the drivers of churn which leads to an understanding of the comprehensibility of the model (De Bock & Van, 2012).

- The size of the data and the classification models are of great importance to the assessment of the performance of algorithms adopted for churn prediction (Martens, Vanthienen, Verbeke, & Baesens, 2011).
  - Classification model types define the format in which the results of the model are returned.

- Common classification models include: linear models, non-linear models, rule-based models, tree-based models and more generally, machine learning algorithms.
  - Among the machine learning algorithms adopted for classification modeling, the decision trees algorithm possesses attributes similar to both a tree-based and rule-based model.

INTRODUCTION…

- The decision trees algorithm implements the solution for a classification model as a top-down hierarchical tree using a divide and conquer approach (Alwis, Kumara, & Hapuarachchi, 2018).

- The tree adopts the input variables as its nodes; starting from the root node at the top to successive parent and child nodes while using their label values as edges used to connect successive nodes.
  - This interconnection of nodes continues to the bottom of the tree where the terminal node (or leaf) is used to describe the target class.

- The decision trees uses this hierarchical representation to describe the importance or relevance of variables with respect to the target class of interest.
  - The tree can also be interpreted as a set of IF-THEN rules which are proportional in number to the number of leaves (or terminal nodes of the tree).

INTRODUCTION…

Rule 1:
If (outlook=sunny)
AND (humidity=high)
THEN (play=no)

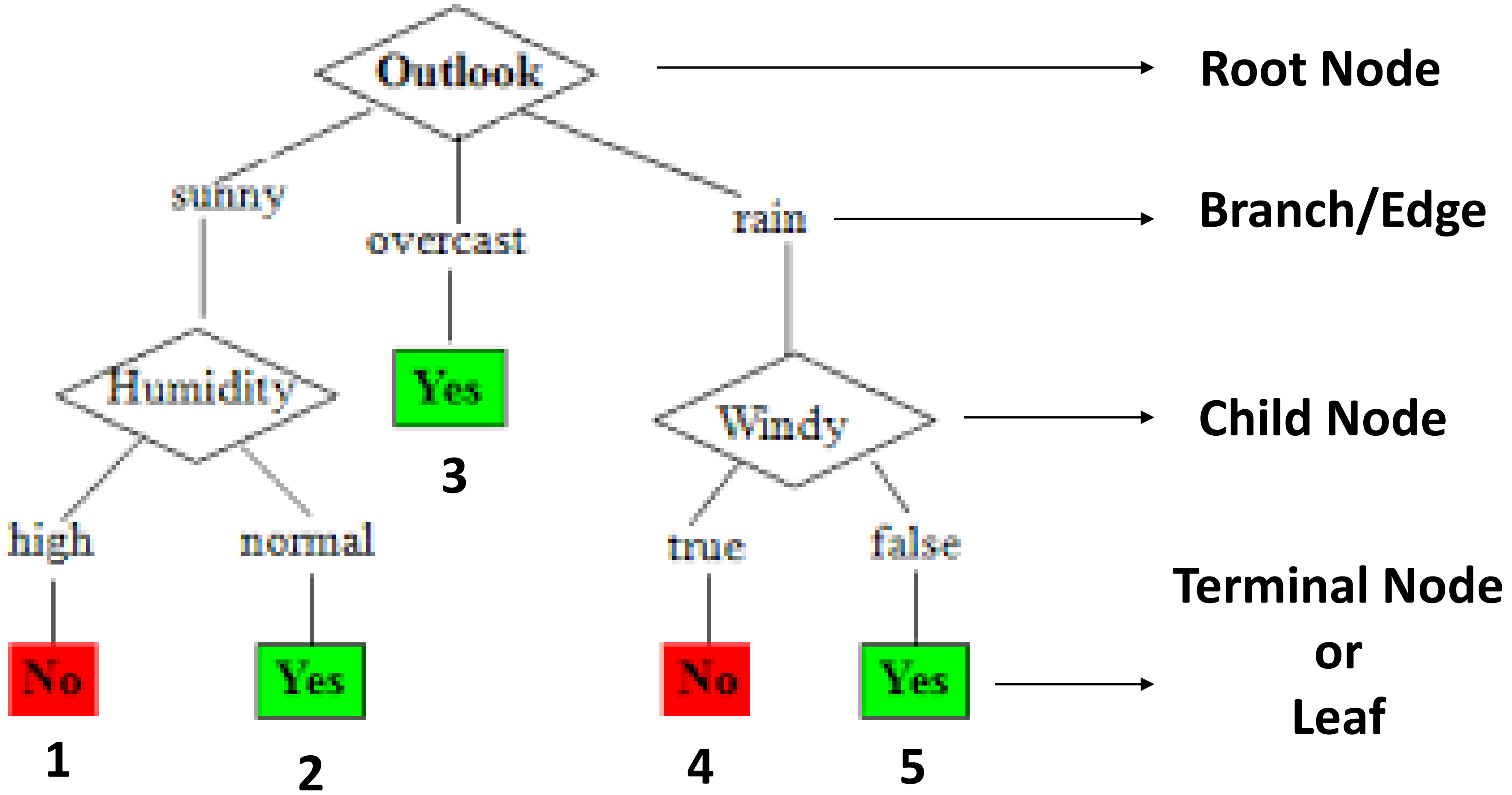

Figure 1: A Sample Decision Trees Model

STATEMENT OF PROBLEM

- According to a review of literature, a number of work has been conducted in the application of statistical analysis to customer churn in telecommunication (Alwis, Kumara &Hapuarachchi, 2018), the application of machine learning in electronic banking (Keramati, Ghaneei, Mirmohammadi & Seyed, 2016), electronic mail (Singh & Sharma, 2017) and logistic service (Pradeep, Sushmitha, Swati, & Akshay, 2017).
  - These studies emphasized the capacity of decision trees in providing effective classification models for customer churn by identifying relevant variables.
  - However, none of these studies provided information on the structure of the hierarchical tree that was used to implement the customer churn prediction problem.
- Related studies failed to identify the variables that are relevant to the classification of customer churn by the decision trees thereby revealing variables of interest required for improving the study of potential customer churn relative to their respective dataset.
- The results of related study were dependent on single simulations runs and failed to perform repetitive trails using varying proportion of training and testing datasets in order to measure their impact of model performance.

AIM AND OBJECTIVES OF STUDY

- This study is aimed at the application of decision trees algorithm to the classification of customer churn based on information collected about related variables with the purpose of identifying important variables relevant for improving customer churn prediction.

- The specific research objectives are to
  i. elicit knowledge on the variables associated with customer churn prediction;

  ii. formulate the classification model for customer churn;

  iii. simulate the model; and

  iv. validate the model.

RELATED WORKS 1 OF 3

Keramati, Ghaneei, Mirmohammadi, and Seyed (2016), applied predictive data mining to the development of a predictive model for customer churn in electronic banking service.

- **Methods**: Data for this study was collected from information about electronic banking services provided to customers following which decision trees (DT) algorithm was used to develop the predictive model using a 10-fold cross validation method.

- **Results**: The results of the study showed that the use of the DT algorithm revealed a limited number of variables which were more relevant for predicting churn among customers provided electronic banking service.

- **Remarks**: The study failed to provide details on the structure of the decision tree model used to classify customer churn nor considered the effect of varying folds on the performance of decision trees algorithm used.

RELATED WORKS 2 OF 3

Sabbeh (2018), performed a comparative analysis of the performance of machine learning algorithms for the classification of customer churn.

- **Methods**: The study adopted the use of filter-based feature selection for the identification of the most relevant among the explanatory variables collected. Also, the study adopted the use of eight machine learning algorithms for the classification of customer churn based on the collected dataset based on a 10-fold cross validation.

- **Results**: The results of feature selection revealed that 13 features among the identified variables were deemed relevant using the filter-based algorithms with decision trees having the best performance.

- **Remarks**: The study failed to provide details of the effect of varying folds on the performance on the machine learning algorithm used.

RELATED WORKS 3 OF 3

Ahmad, Jafar, and Aljoumaa (2019), performed a comparative analysis of the application of machine learning to the prediction of customer churn in the telecommunication sector.

- **Methods**: The study adopted the use of feature ranking for ranking the variables according to their effect on improving the performance of machine learning algorithm on churn prediction.
- **Results**: The results of feature selection revealed that 13 features among the identified variables were deemed relevant using the filter-based algorithms with decision trees having the best performance.
- **Remarks**: The study failed to identify the performance of the relevant variables selected on the performance of the prediction of customer churn.

METHOD I – DATA IDENTIFICATION AND COLLECTION

- The data that was adopted for this study contained various information about customers of Telco communication Company; a publicly available dataset.
  - The dataset was downloaded as a.csv file from the Kaggle online repository via the URL https://www.kaggle.com/blastchar/telco-customer-churn.
  - The datasets consist of 7043 records with a binary classification of the target class, such that 1870 were customers who churned and 5173 were customers who did not churn.

- The dataset was composed of information about 20 variables broadly classified into three (3) groups, namely:
  - **demographic data of customers** – gender, senior citizen status, marital status and dependents
  - **subscribed service** – phone service, multiple line, internet service, online security and backup, device protection, technical support, streaming TV and movies.
  - **customer account information** – customer ID, contract terms, paperless billing, payment method, monthly charges, total charges, and tenure

METHOD I – DATA IDENTIFICATION AND COLLECTION…

```
@relation customerChurnData

@attribute gender {Female,Male}
@attribute SeniorCitizen {Yes,No}
@attribute Partner {Yes,No}
@attribute Dependents {Yes,No}
@attribute tenure numeric
@attribute PhoneService {Yes,No}
@attribute MultipleLines {Yes,No,No-phone-service}
@attribute InternetService {DSL,Fiber-optic,No}
@attribute OnlineSecurity {No,Yes,No-internet-service}
@attribute OnlineBackup {No,Yes,No-internet-service}
@attribute DeviceProtection {No,Yes,No-internet-service}
@attribute TechSupport {No,Yes,No-internet-service}
@attribute StreamingTV {No,Yes,No-internet-service}
@attribute StreamingMovies {No,Yes,No-internet-service}
@attribute Contract {Month-to-month,One-year, Two-year}
@attribute PaperlessBilling {Yes,No}
@attribute PaymentMethod {Electronic-check,Mailed-check,Bank-transfer,Credit-card}
@attribute MonthlyCharges numeric
@attribute TotalCharges numeric
@attribute Churn {Yes,No}

@data
Female,No,Yes,No,1,No,No-phone-service,DSL,No,Yes,No,No,No,No,Month-to-month,Yes,Electronic-check,29.85,29.85,No
Male,No,No,No,34,Yes,No,DSL,Yes,No,Yes,No,No,No,One-year,No,Mailed-check,56.95,1889.5,No
Male,No,No,No,2,Yes,No,DSL,Yes,Yes,No,No,No,No,Month-to-month,Yes,Mailed-check,53.85,108.15,Yes
Male,No,No,No,45,No,No-phone-service,DSL,Yes,No,Yes,Yes,No,No,One-year,No,Bank-transfer,42.3,1840.75,No
Female,No,No,No,2,Yes,No,Fiber-optic,No,No,No,No,No,No,Month-to-month,Yes,Electronic-check,70.7,151.65,Yes
Female,No,No,No,8,Yes,Yes,Fiber-optic,No,No,Yes,No,Yes,Yes,Month-to-month,Yes,Electronic-check,99.65,820.5,Yes
Male,No,No,Yes,22,Yes,Yes,Fiber-optic,No,Yes,No,No,Yes,No,Month-to-month,Yes,Credit-card,89.1,1949.4,No
Female,No,No,No,10,No,No-phone-service,DSL,Yes,No,No,No,No,No,Month-to-month,No,Mailed-check,29.75,301.9,No
Female,No,Yes,No,28,Yes,Yes,Fiber-optic,No,No,Yes,Yes,Yes,Yes,Month-to-month,Yes,Electronic-check,104.8,3046.05,Yes
Male,No,No,Yes,62,Yes,No,DSL,Yes,Yes,No,No,No,No,One-year,No,Bank-transfer,56.15,3487.95,No
Male,No,Yes,Yes,13,Yes,No,DSL,Yes,No,No,No,No,No,Month-to-month,Yes,Mailed-check,49.95,587.45,No
Male,No,No,No,16,Yes,No,No,No-internet-service,No-internet-service,No-internet-service,No-internet-service,No-internet-service,No-internet-service,Two-year,No,Credit-card,18.95,326.8,No
Male,No,Yes,No,58,Yes,Yes,Fiber-optic,No,No,Yes,No,Yes,Yes,One-year,No,Credit-card,100.35,5681.1,No
Male,No,No,No,49,Yes,Yes,Fiber-optic,No,Yes,Yes,No,Yes,Yes,Month-to-month,Yes,Bank-transfer,103.7,5036.3,Yes
Male,No,No,No,25,Yes,No,Fiber-optic,Yes,No,Yes,Yes,Yes,Yes,Month-to-month,Yes,Electronic-check,105.5,2686.05,No
Female,No,Yes,Yes,69,Yes,Yes,Fiber-optic,Yes,Yes,Yes,Yes,Yes,Yes,Two-year,No,Credit-card,113.25,7895.15,No
Female,No,No,No,52,Yes,No,No,No-internet-service,No-internet-service,No-internet-service,No-internet-service,No-internet-service,No-internet-service,One-year,No,Mailed-check,20.65,1022.
Male,No,No,Yes,71,Yes,Yes,Fiber-optic,Yes,No,Yes,No,Yes,Yes,Two-year,No,Bank-transfer,106.7,7382.25,No
Female,No,Yes,Yes,10,Yes,No,DSL,No,No,Yes,Yes,No,No,Month-to-month,No,Credit-card,55.2,528.35,Yes
Female,No,No,No,21,Yes,No,Fiber-optic,No,Yes,Yes,No,No,Yes,Month-to-month,Yes,Electronic-check,90.05,1862.9,No
Male,Yes,No,No,1,No,No-phone-service,DSL,No,No,Yes,No,No,Yes,Month-to-month,Yes,Electronic-check,39.65,39.65,Yes
```

Normal text file | length : 872,095 lines : 7,067 | Ln : 1 Col : 1 Sel : 0 | 0 | Windows (CR LF) | UTF-8 | INS

Figure 2: Dataset collected for the Classification of Customer Churn

METHOD II – CLASSIFICATION MODEL FORMULATION

- Consider the historical dataset *S* which was collected and consists of *j* customer records containing *i* attributes (or variables) which represent the set of classification factors, X_{ij} in addition to their respective target variable (presence of customer churn), Y_j.

- Therefore, the purpose of the decision trees algorithm is to infer a mapping φ with the lowest error required for determining the customer churn *Y* from information about associated variables *X* according to equation (1).

$$\varphi: X \rightarrow Y \qquad (1)$$

$$defined\ by\ \varphi(X) = \begin{cases} Churn\ (Yes) \\ No\ Churn\ (No) \end{cases}$$

- The decision trees algorithm was used to split the customer churn dataset into subsets by recursive partitioning of the parent nodes (variables) into child nodes based on the homogeneity of the within–node instances or separation of between-node instances with respect to customer churn information.

- Thus, at each node, attributes were examined using a root test and the splitter is chosen to be the attribute which after producing child nodes according to the value of the attribute variable provides a good classification of customer churn.

METHOD II – DECISION TREES MODELING

- **CART Decision Trees Algorithm**
  - The CART implements its decision trees as nodes with binary edges.
  - CART was used to estimate the Gini index of each variable by estimating the conditional probability of the occurrence of the value *v* of a variable *X* with respect to each customer churn classification $c \in C$.
  - The closer the Gini index is to 1 then it is associated with either class C and thus used to split the dataset.

$$Gini\ Index(X) = 1 - \sum_{c=1}^{C} (p(v|c))^2 \qquad (2)$$

- **C4.5 Decision Trees Algorithm**
  - The C4.5 implements its decision trees as nodes with multiple edges proportional to the variable's values.
  - The C4.5 decision trees algorithm adopts the use of entropy *H(X)* to measure the level of impurity or randomness existing within a dataset.

$$Entropy, H(X) = -\sum_{i=1}^{n} p(v_i) \log_2 p(v_i) \qquad (3)$$

  - In addition, a mutual entropy *H(C|X)* was determined in terms of entropy as a function of the conditional probability of the value *v* of an attribute *X* with respect to the values *c* of the target class *C*.

$$Mutual\ Entropy, H(C|X) = -\sum_{c=1}^{C} p(v|c) \log_2 p(v|c) \qquad (4)$$

  - The value of the information gain (IG) was determined by subtracting the entropy from the mutual entropy.
  - The closer the IG is to 1 then, the more effect of information about *X* on the determination of the target class *C*.

METHOD III – MODEL SIMULATION

- This study adopted the use of the percentage split for the simulation of the classification model for the prediction of customer churn using the Waikato Environment for Knowledge Analysis (WEKA) software; a Java-based machine learning simulation platform.
  - This percentage split technique adopted a proportion of the dataset for training (or building) the model while the remaining proportion was used to validate (or test) the model.

- This study used 10 runs of the percentage split method for simulating and validating the model based on the number of records shown below:
  - Run 1 – 50% used for training while 50% was used for testing.
  - Run 2 - 55% used for training while 45% was used for testing.
  - Run 3 - 60% used for training while 40% was used for testing.
  - Run 4 - 65% used for training while 35% was used for testing.
  - Run 5 - 70% used for training while 30% was used for testing.
  - Run 6 - 75% used for training while 25% was used for testing.
  - Run 7 - 80% used for training while 20% was used for testing.
  - Run 8 - 85% used for training while 15% was used for testing.
  - Run 9 - 90% used for training while 10% was used for testing.
  - Run 10 - 95% used for training while 5% was used for testing.

METHOD IV – MODEL VALIDATION

- Following the simulation of the predictive model for customer churn prediction, there was a need for the interpretation of the simulation of test dataset which was required for the validation of the predictive model.
  - This was done by representing the correct and incorrect classifications made by the decision trees algorithm on a 2 by 2 square confusion matrix as shown in the figure below.

Churn	No Churn	
A	B	Churn
C	D	No Churn

- The sum of the horizontal cells provide the number of actual records, such that A+B and C+D are the total actual churn and no churn records respectively.
- The sum of the vertical cells provide the number of predicted records, such that A+C and B+D are the total predicted churn and no churn records respectively.
- A and D are correct predictions of the churn and no churn records respectively while B and C are incorrect predictions of churn as no churn and no churn as churn respectively.

METHOD IV – MODEL VALIDATION…

- **Accuracy** - was used to measure the proportion of the total correctly classified records for both churn class (Yes and No) expressed as a percentage according to equation (5).

$$Acccuracy = \frac{A + D}{A + B + C + D} \times 100\% \qquad (5)$$

- **True Positive (TP) rate/Sensitivity/Recall** - was used to assess the proportion of actual records that were correctly classified according to equations (6a) and (6b).

$$TP\ rate_{Churn} = \frac{A}{A + B} \qquad (6a)$$

$$TP\ rate_{No\ Churn} = \frac{D}{C + D} \qquad (6b)$$

- **False Positive (FP)/False Alarm rate** - was used to assess the proportion of actual records of a target class which were misclassified as the other class according to equations (7a) and (7b).

$$FP\ rate_{Churn} = \frac{C}{C + D} \qquad (7a)$$

$$FP\ rate_{No\ Churn} = \frac{B}{A + B} \qquad (7b)$$

- **Precision** - was used to assess the proportion of predicted records of a target class which were correctly classified according to equations (8a) and (8b).

$$Precision_{Churn} = \frac{A}{A + C} \qquad (9a)$$

$$Precision_{No\ Churn} = \frac{D}{B + D} \qquad (9b)$$

RESULTS

Table 1: Validation of Classification Model using C4.5 (left) and CART (right)

Training/Testing proportion	C4.5 Decision Trees Algorithm Performance Evaluation					
	Build Time (s)	Test Time (s)	Accuracy (%)	TP rate	FP rate	Precision
50/50	0.36	0.28	77.76	0.778	0.389	0.769
55/45	0.35	0.42	77.25	0.772	0.398	0.764
60/40	0.36	0.12	77.28	0.773	0.385	0.767
65/35	0.31	0.04	77.69	0.777	0.406	0.767
70/30	0.30	0.03	78.61	0.786	0.376	0.780
75/25	0.36	0.07	77.46	0.775	0.379	0.769
80/20	0.32	0.00	77.64	0.776	0.361	0.773
85/15	0.31	0.04	78.22	0.782	0.387	0.773
90/10	0.35	0.00	77.84	0.778	0.383	0.771
95/5	0.30	0.00	80.68	0.807	0.326	0.807

Training/Testing proportion	CART Decision Trees Algorithm Performance Evaluation					
	Build Time (s)	Test Time (s)	Accuracy (%)	TP rate	FP rate	Precision
50/50	8.44	0.01	78.84	0.790	0.348	0.785
55/45	17.19	0.01	78.48	0.785	0.334	0.785
60/40	8.28	0.01	79.09	0.791	0.341	0.788
65/35	8.21	0.01	78.90	0.789	0.358	0.784
70/30	8.29	0.01	79.74	0.797	0.384	0.788
75/25	8.41	0.01	79.61	0.796	0.392	0.785
80/20	8.52	0.00	79.70	0.797	0.355	0.790
85/15	8.32	0.00	79.35	0.794	0.352	0.788
90/10	8.73	0.00	79.69	0.797	0.362	0.789
95/5	9.25	0.00	81.82	0.818	0.307	0.818

RESULTS

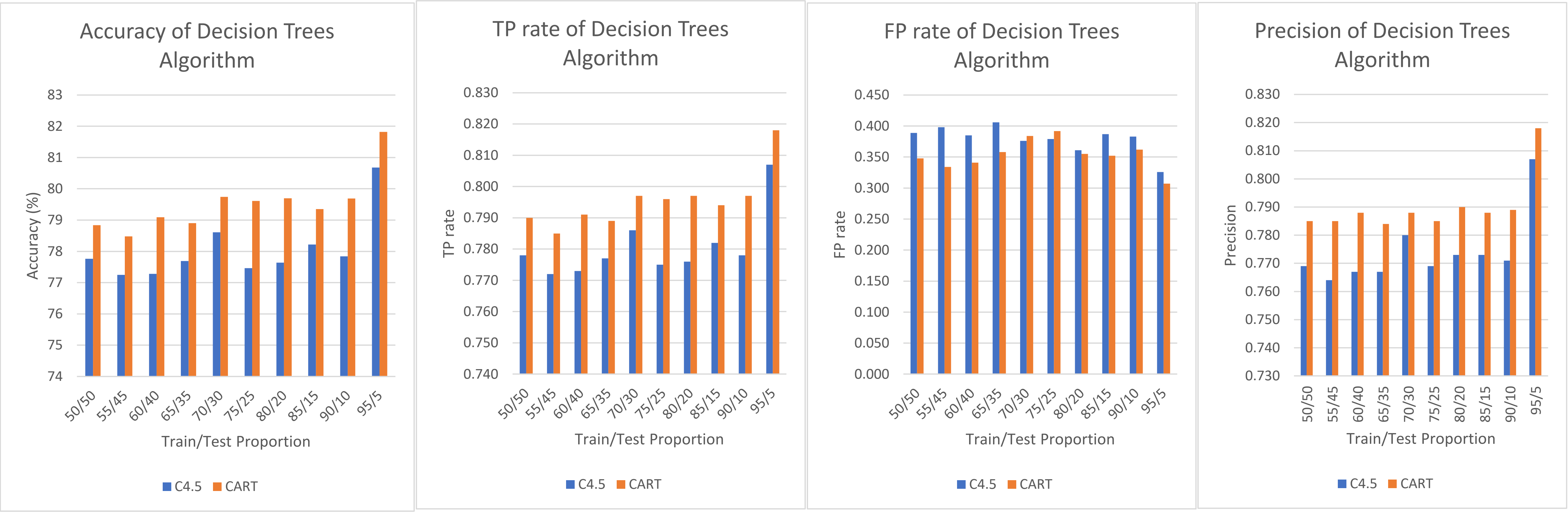

Figure 3: Graphical Plot of Results of Accuracy, TP rate, FP rate and Precision (from left to right) for Decision Trees Algorithms

RESULTS – DECISION TREE MODEL

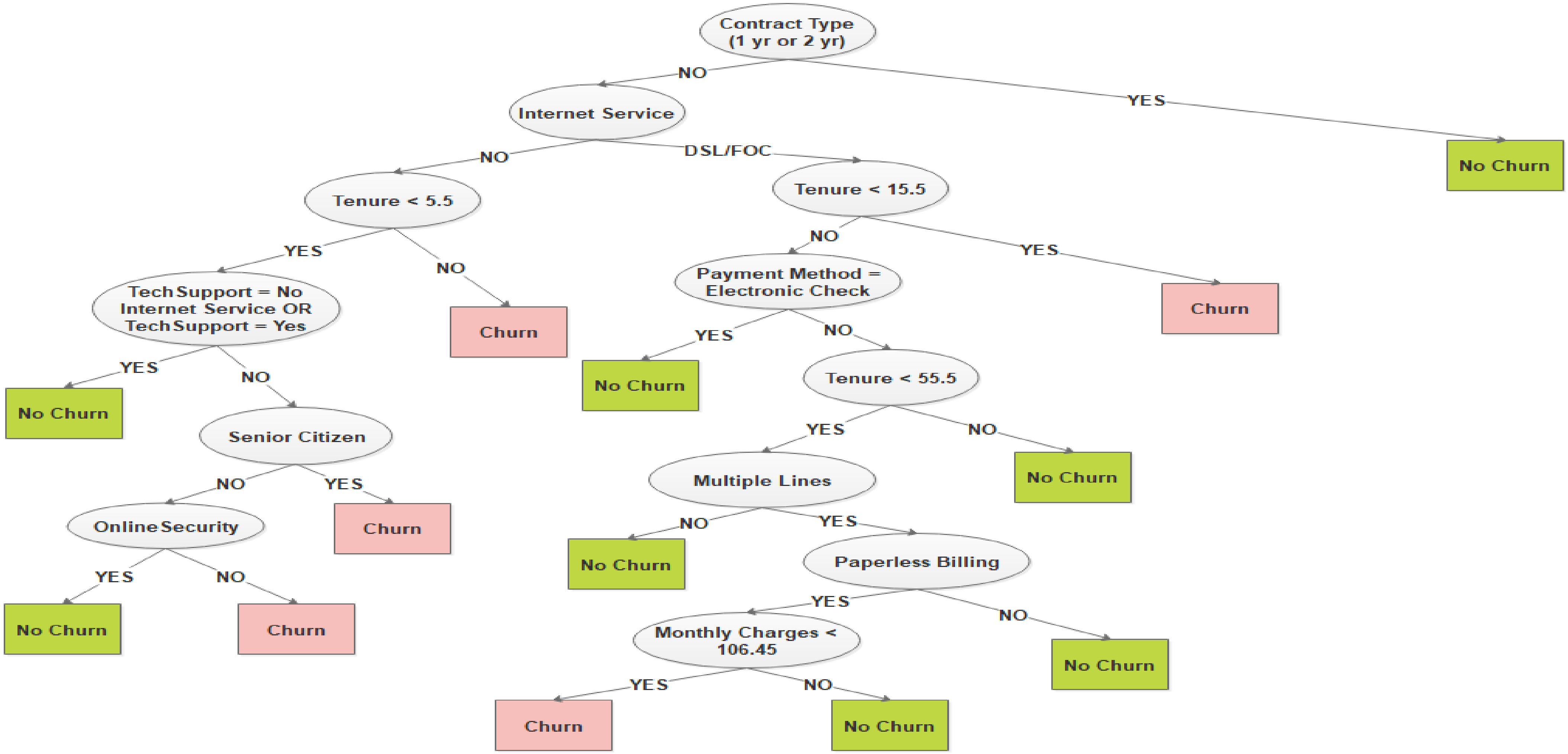

Figure 4: Decision Tree Model for Customer Churn Prediction

RESULTS – IDENTIFICATION OF RELEVANT VARIABLES

- The following variables were the most important and relevant for the prediction of customer churn relative to the telecommunication data adopted in this study:
  1. The type of contract
  2. Internet Service
  3. Tenure of Service
  4. Technical Support
  5. Payment Method used
  6. Senior Citizen Status
  7. Online Security
  8. Multiple Lines
  9. Paperless Billing
  10. Monthly Charges

CONCLUSION

- The study concluded that information assessed about customer demographics, services subscribed to by customer and customer account information can be used to assess customer churn prediction using machine learning algorithm.

- Adopting a decision tree for the modeling of customer churn prediction behaviour provides a structural and easy to interpret model for non-experts also thereby improving decision-support.

- The structural model developed by the decision trees can be converted to IF-THEN statement containing the combination of the values of variables (parent and child nodes) as antecedent statements while the value of the target class (leaf or terminal node) becomes the consequent part.

- Also, among customer account information the most relevant variables are contract type, tenure and monthly charges. The study identified 10 relevant variables from the initial 19 variables collected for customer churn prediction.

- The study concluded that model performance increased with increasing size of training dataset.

REFERENCES

Ahmad, A., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in Big data platform. *Journal of Big Data, 6*(28), 1-24.

Alwis, P., Kumara, B., & Hapuarachchi, H. (2018). Customer Churn Analysis and Prediction in Telecommunication for Decision Making. *2018 International Conference on Business Innovation (ICOBI)*, (pp. 40-45). Colombo.

Ballings, M., & Van, D. (2012). Customer Event History for Churn Prediction: How Long is Long Enough. Expert Systems with Applications, 39, 13517-13522.

Blattberg, R., Kim, B., & Neslin, S. (2010). Database Marketing: Analyzing and Managing Customers. New York, NY: Springer.

De Bock, K., & Van, d.-P. (2012). Reconciling Performance and Interpretability in Customer Churn Prediction using Ensemble Learning based on Generalized Additive Models. Expert System with Applications, 39, 6816-6826.

Hu, C., Shu, H., & Qiao, X. (2014). Customer Segmentation Model Research based on Organizational Customer Life Cycle in Telecom Operator. *International Conference on Education and Social Science.* Atlantis Press. Retrieved from http://dx.doi.org/10.2991/icetss-14.2014.88

Keramati, A., Ghaneei, H., Mirmohammadi, H., & Seyed, M. (2016). Developing a prediction model for customer churn from electronic banking services using data mining. Financial Innovation, 2(10), 1-13.

Lessman, S., & Vob, S. (2009). A Reference Model for Customer-Centric Data Mining with Support Vector Machines. European Journal of Operational research, 199, 520-530.

Martens, D., Vanthienen, J., Verbeke, W., & Baesens, B. (2011). Performance of Classification Models from a User-Perspective. Decision Support Systems, 51(4), 782-793.

Pradeep, B., Sushmitha, V., Swati, M., & Akshay, H. (2017). Analysis of Customer Churn prediction in Logistic Industry using machine learning. International Journal of Scientific and Research Publications, 7(11), 401-403.

Sabbeh, A. (2018). Machine Learning Techniques for Customer Retention: A Comparative Study. International Journal of Advanced Computer Science and Applications, 9(2), 273-281.

Singh, D., & Sharma, V. (2017, June). Predicting Churn in E-Mail using Decision Trees. International Journal for Research in Applied Science and Engineering (IJRASET), 5(6), 288-298.