Decision making of churn rate in a company using a decision tree in R Studio

Name of Student

Institution Affiliation

Course Title

Introduction

This assignment aims to use a decision tree to predict the churn rate of BondTelco 'Company. The Company was concerned about the high churn rate of its customers. A marketing strategy has been developed by the sales team to try and lure the customers whose contracts are coming due. The customers were to be offered incentives to attract them. The main challenge about this strategy is that the Company will have to provide incentives for all the customers whether they are likely to leave or stay. To solve these issues, I have been employed, as a data scientist, to figure out the customers who are likely to leave so that they can be offered the incentives instead of all the customers. This will reduce the cost.

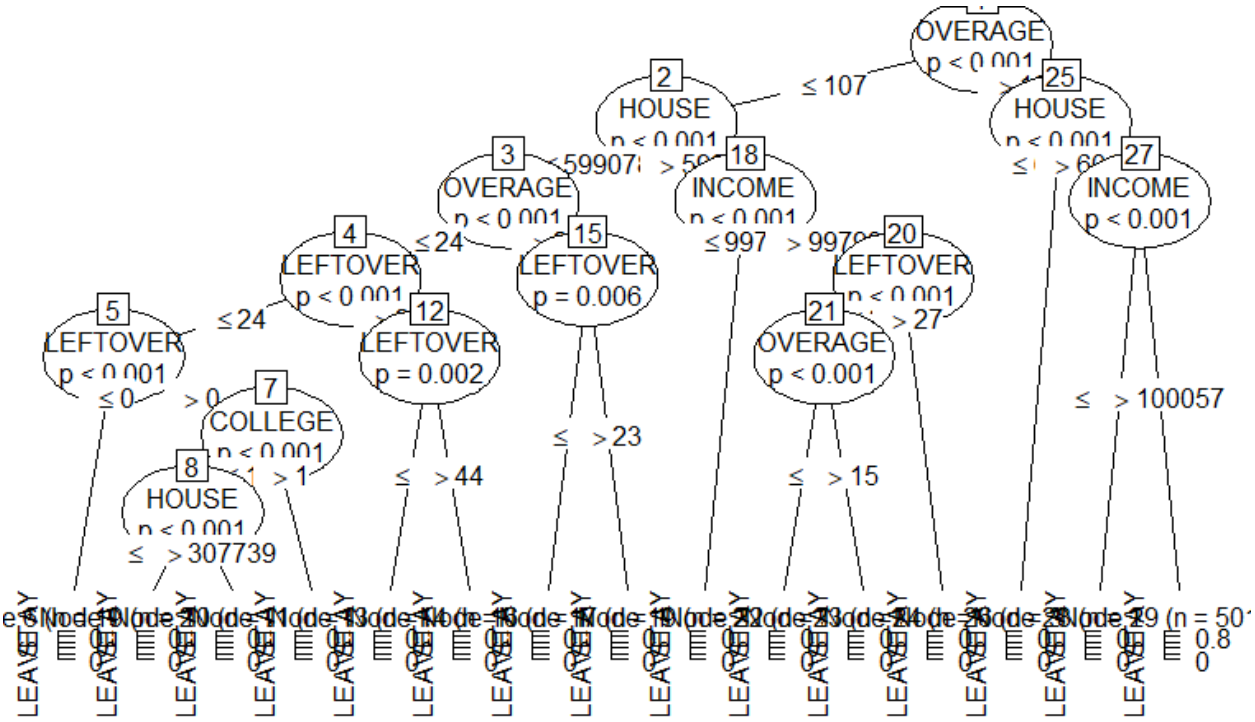
Method

To achieve this, the It staffs have provided data from the Company's database. The data contains information about 20,000 previous customers including all that left and those that stayed. The data will be analyzed using the decision tree in the R studio using the R programming language. R programming is a powerful statistical language that produces reliable outputs. The analysis was conducted in many ways. The data obtained contains 20,000 rows and 12 columns. To provide an accurate solution, few variables should be used (Bhargava et al. 2013). The variables should not be chosen at random, and instead, an analysis should be carried out to help select the variables that produce the most important prediction. The first decision tree will involve all the variables. From the output, one can choose the best variables for the model. Later on, a decision tree is conducted using the selected variables to produce an accurate prediction.

Results

The first model was carried out with the variable 'LEAVE' as the outcome variable and the remaining variables as the predictor variables. The outcome variables were initially converted to factor variables, and all the predictor variables were converted to numeric variables (Yu, Asche and Fairchild, 2011). The plot for the first model was as shown below

Fig 1: Decision Tree 1



From fig 1, one can notice that the most accurate variable that should be used in the model is OVERAGE and HOUSE. However, it is very difficult to decide on the tree. Therefore, the data is subsetted into three variables with LEAVE as the outcome variable and OVERAGE and HOUSE as the independent variables. The results of the 2nd model are summarized by the figure below:

Fig 2: Decision Tree 2

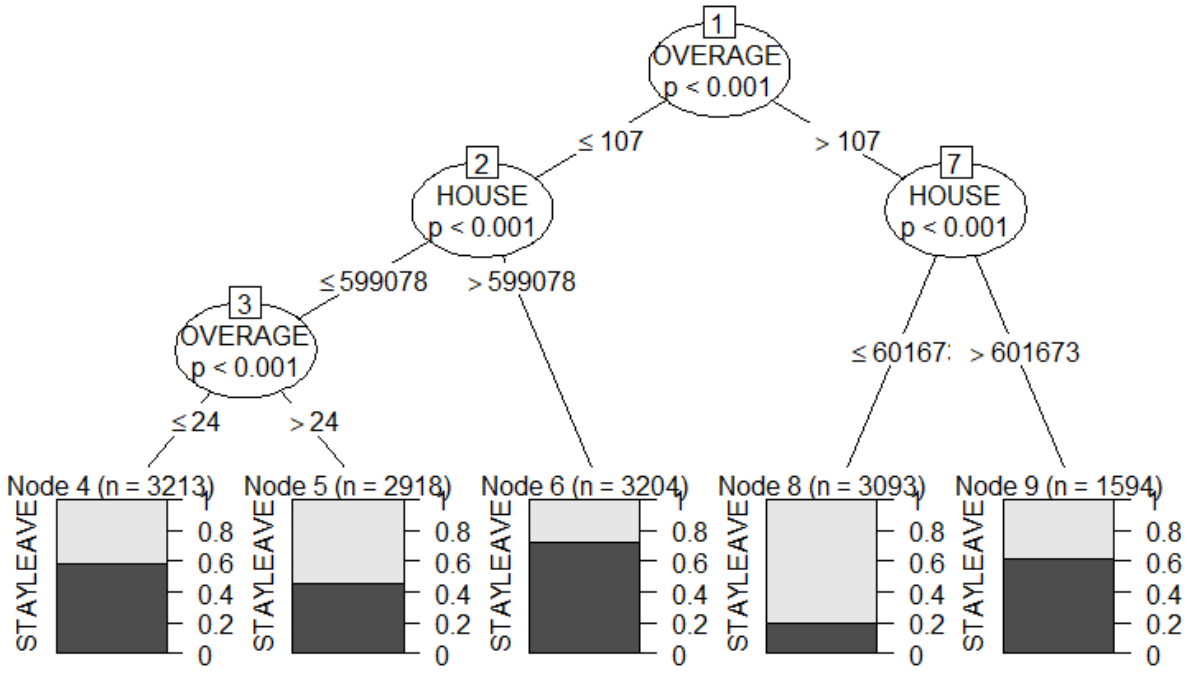


Figure 2 shows the prediction of the two variables. It can be seen that a 40 % probability of the customers who had Overage less than 107 and the value of the house less than 599078 with Overage value below 24 were likely to leave the Company. About 60 % probability of the customers who had an Overage less than 107, the value of the house less than 599078 with the Overage above 24 were likely to leave the company. Only 30 % probability of the customers who has an overage of less than 107 and the value of house greater than 599078 were likely to leave the company. About 80 % probability of the customers who had an Overage of above 107 and the value of the house less than 60167 were likely to leave the company, and finally, about 40 % of the customers who had an Overage above 107 and the value of the house greater than 601673 were likely to leave.

Conclusion and Recommendation

From the above analysis, one can observe that the majority of the customers who were likely to leave the Company had an overage of above 107 and the value of house below 60,167 followed by those who had an Overage less than 107, the value of the house less than 599078 with the Overage above 24. Therefore the Company should provide incentives to mainly these two groups to reduce the level of churn rate and also reduce the cost

References

Bhargava, N., Sharma, G., Bhargava, R. and Mathuria, M., 2013. Decision tree analysis on j48 algorithm for data mining. *Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering*, *3*(6).

Yu, J., Asche, C.V. and Fairchild, C.J., 2011. The economic burden of dry eye disease in the United States: a decision tree analysis. *Cornea*, *30*(4), pp.379-387.