#### **MIS-64036: Business Analytics**

#### **(Fall 2022)**

**Project - Customer Churn Prediction**

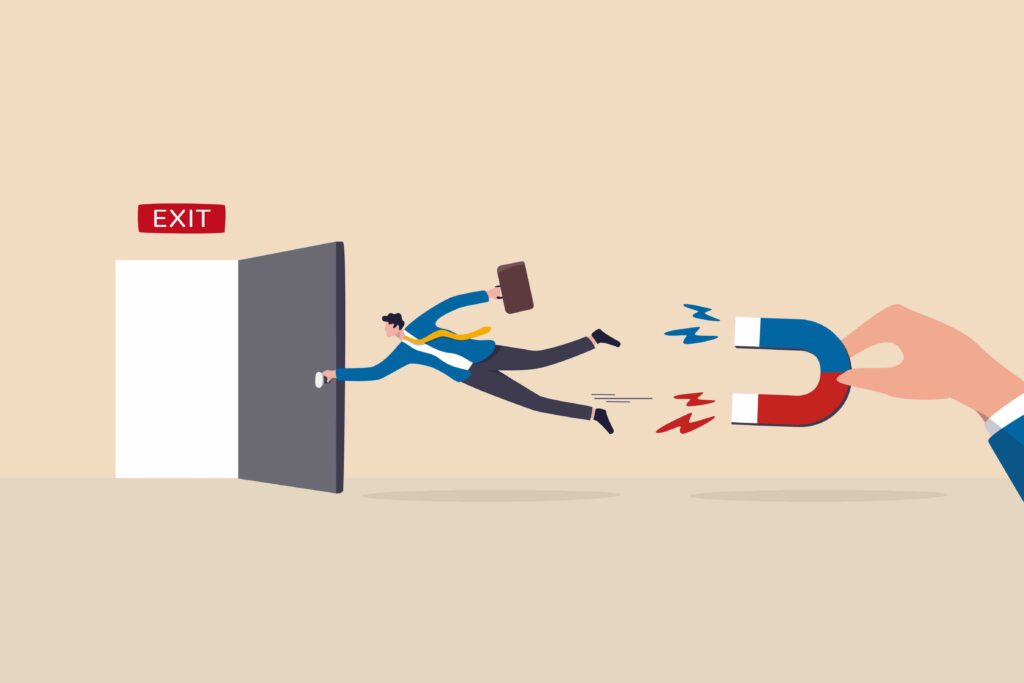
**Submitted By: Group 7**

Niharika Kolliboyina- Modelling Strategy & Model Performance, PPT

Snehitha Anpur – Introduction & Data Exploration, PPT, Insights & Conclusions

Avinash Ravipudi - Modelling Strategy & Model Performance, PPT

Sony Kiran - Modelling Strategy & Model Performance, PPT



**Table of Contents**

[Project Goal/Abstract:](#_TOC_250014)

[Introduction:](#_TOC_250013)

[Data Exploration:](#_TOC_250012)

[Descriptive Statistics:](#_TOC_250011)

Data Cleaning (Missing Values):

[Modelling Strategy:](#_TOC_250009)

[Model Performance:](#_TOC_250008)

For Logistic Regression Model

For Decision Tree

[Compare the Confusion Matrix for 2 Models:](#_TOC_250006)

[Decision-Tree: Classification Model](#_TOC_250005)

[Build the Decision Tree model:](#_TOC_250004)

[Visualization of Decision Tree graph:](#_TOC_250003)

[Prediction of Test Dataset](#_TOC_250002)

[Insights and Conclusions:](#_TOC_250001)

[Recommendations to Curtail the Churn Rate:](#_TOC_250000)

## **Abstract:**

One of the most difficult problems facing large firms is customer churn, which is a major worry. Due to the direct influence that customer churn has on businesses' earnings, particularly in the telecom sector, businesses are attempting to develop a variety of approaches to predict potential client attrition. As a result, to prevent customer turnover, it is essential to identify the elements that lead to it. Prediction is vital for businesses in today's economic environment since acquiring a new client cost more than maintaining old ones. This project's main accomplishment is the creation of a churn prediction model that enables ABC Wireless Inc. to identify the customers who are most likely to leave and lower the churn rate by keeping those customers. Offering them unique promos or incentives is part of this retention strategy. The model created in this project employs machine learning approaches. The model's performance may be measured using the Decision Tree is used, and the derived value is 91.8. The model was developed and validated using a huge dataset generated by converting enormous amounts of raw data given by ABC Telecom. The dataset includes all customer information collected and was used to train, test, and assess the system at ABC Wireless Inc. Two methods were tested in the model: Decision Tree and Logistic Regression. However, the Decision Tree algorithm generated the best results. In this churn predictive model, this algorithm was employed for categorization.

## **Introduction:**

One of the major sectors in developed nations is now the telecommunications industry. The level of competition has increased due to technical developments and an increase in operators. Using a variety of techniques, businesses are fighting hard to survive in this cutthroat environment. To enhance revenue, three key techniques have been put forth: acquiring new clients, upselling current clients, and lengthening client retention. The third strategy, however, is the most profitable one, demonstrating that keeping an existing customer cost much less than acquiring a new one and is also thought to be much simpler than the upselling strategy, when these strategies are compared and the value of return on investment (Roi) of each is considered. The third tactic, when used, churn, also referred to as "the client transfer from one provider to another," must be reduced by businesses.

In service industries where services are fiercely competitive, customer churn is a major problem. On the other side, identifying consumers who are likely to abandon the business early on could represent a significant new revenue source.

The primary goal of this project is to analyze the data and find various patterns using predictive modeling to ascertain whether an existing customer of ABC Wireless telecom company will switch the service provider. Many researchers agree that machine learning technology is highly efficient to predict this situation.

## **Data Exploration:**

To conduct the study and create the model for this project, we needed the historical data(Churn Train.csv). The services that the clients of the ABC Wireless Telecom firm utilize are detailed in this data. This dataset has 20 characteristics, 3333 Customers, and both category and numerical variables. The dataset's attributes include:

* state (categorical)
* account\_length
* area\_code
* international\_plan (yes/no)
* voice\_mail\_plan (yes/no)
* number\_vmail\_messages
* total\_day\_minutes
* total\_day\_calls
* total\_day\_charge
* total\_eve\_minutes
* total\_eve\_calls
* total\_eve\_charge
* total\_night\_minutes
* total\_night\_calls
* total\_night\_charge
* total\_intl\_minutes
* total\_intl\_calls
* total\_intl\_charge
* number\_customer\_service\_calls
* Churn- (Target variable) which takes two values ‘no’ and ‘yes’.

Let’s run some exploratory data analysis on the dataset through data visualization

#### **Descriptive Statistics:**

Descriptive Statistics can reveal very useful insights about data. Here we can observe the Mean, Median, Max and Min, 1st and 3rd Quartile information of the variables in the dataset in the below picture

Text

Description automatically generated with medium confidence

## 

## **Modelling Strategy:**

The three situations covered in the prior part on logistic regression and decision tree models were put into practice. We divided the "Churn Train" dataset into training and validation sets by 70% and 30%, respectively, to evaluate the models' performance.

According to our estimate, the Decision Tree model on a set of data where the NAs have been imputed using the Mice approach has the greatest Accuracy at 91.80. It's also intriguing to observe that even Specificity is the highest of all, demonstrating that it could forecast genuine negatives more correctly than other models and methodologies used.

As a result, we made the decision to construct our algorithm for forecasting customer turnover using the Decision Tree model. After applying the model to the test data, it finally projected that 156 customers would have left or switched companies.

PS: The values given in the table may vary depending on system session info such as package versions etc.

## **Model Selection:**

#### **Compare the Confusion Matrix for 2 Models:**

* **Confusion Matrix for Logistic Regression**

#### **Graphical user interface Description automatically generated with medium confidence**

#### Text Description automatically generated with medium confidence

* **Confusion Matrix for Decision Tree Model**

#### **Text, letter Description automatically generated**

#### **Text Description automatically generated**

Accuracy and Specificity are higher for Decision Tree Model compared to Logistic Regression Model.

Having considered all the performance metrics and other data outlined under Modeling Strategy

, it would be prudent to go with Decision tree Model for the case at hand.

For this dataset, sensitivity or recall should be high because we shouldn’t predict the churn as ‘no’ but the actual churn is as Yes. That is false negatives. From the above model, recall value is 0.53

### **Decision-Tree: Classification Model**

The main goal of the decision tree model is to Classify or predict an outcome based on a set of predictors.

#### **Build the Decision Tree model:**

#### **Graphical user interface, text Description automatically generated**

#### **Diagram Description automatically generated**

#### 

#### **Visualization of Decision Tree graph:**

#### **A picture containing indoor Description automatically generated**

The decision tree's nodes, which are represented by the rectangles in the image, have leaves and branches. The model's branches represent the criteria requested, while the leaves represent the class level. We have a "yes" and a "no" on either side that determines if the data point meets the requirement. For instance, the first "activity" in the root node (total day charge) is "No" because the majority of the data points there fulfill the total day charge less than 45 or being less than 45 is the predominate activity. The likelihood that a new data point will select "Yes" to that action is 94%, while the remaining 6% will select "No." Each node is subject to the same reasoning.

As a result, the Total Day Charge, Number of Client Service Calls, and other factors must be taken into consideration when determining if a customer is likely to leave. Based on that, ABC Wireless Inc. might get crucial information to improve the way it makes business decisions.

# **Prediction of Test Dataset**

#### Graphical user interface, text, application, email Description automatically generated

We will use the provided test dataset (Customers\_To\_Predict) to predict the customer churn using the Decision Tree Model.

This output shows that the model is predicting the churn

#### 

The decision tree model's predicted retention rate of 1444 customers out of 1600 customers is shown in the summary table, which is fantastic news for ABC Wireless Inc.

##### ***Recommendations to Curtail the Churn Rate:***

1. Enhance and deliver superior customer service
2. Maintain or raise general quality
3. Reduce the daytime and international daytime fees or keep the fees competitive for these two groups.
4. Maintain or improve overall quality (Bandwidth in highly dense areas etc.)
5. Reward loyal consumers to keep them coming back.