**Problem statement**

Customer churn is a significant issue for telecommunications companies, as losing customers negatively impacts revenue and growth. In this analysis, we aim to identify the factors influencing customer churn for a telecommunications company using a dataset (Churn.txt) that contains information on 1,477 customers. The customers are categorized into three groups: current customers, involuntary leavers, and voluntary leavers. The objective is to build a decision tree model that predicts whether a customer will churn and, if so, whether the churn is voluntary or involuntary. By understanding the drivers of churn, management can implement targeted strategies to reduce customer attrition and improve customer retention.

**Brief observations from a Data Audit**

The data audit of the dataset, consists of 1,477 observations. Continuous variables such as LONGDIST, International, LOCAL, DROPPED, AGE, CHILDREN, and Est\_Income exhibit a wide range of values, indicating variability among customers. Notably, the AGE variable ranges from 18 to 97 years, with an average age of approximately 57.6, suggesting a diverse customer age profile. Several nominal and flag variables, including PAY\_MTHD, LocalBillType, LongDistanceBillType, SEX, STATUS, Car\_Owner, and CHURNED, are marked as having "too many values," which may complicate analysis and signal a need for further categorization or simplification. The Est\_Income variable shows a broad range from about 110 to over 99,000, with a mean income of approximately 50,290, highlighting significant income diversity that could impact credit risk assessments

**What was done and analysis based on screenshots**

The dataset churn2.xlsx was imported into SPSS Modeler, and key variables such as CHURNED and Est Income were prepared for analysis. A decision tree was built using the C5.0 algorithm, with CHURNED as the main decision node to predict customer churn based on various attributes. The tree helped identify patterns and relationships between customer characteristics and their likelihood of churning. The analysis of the decision tree provided insights into key factors driving churn, offering valuable information for improving customer retention strategies.

Its performance was evaluated by comparing the model's predicted outcomes ($C-CHURNED) with the actual churn outcomes (CHURNED). Key results from the analysis show that the model accurately predicted 1,279 cases, achieving an overall accuracy of 86.59%. However, there were 198 incorrect predictions, accounting for 13.41% of the total. The model's evaluation was based on a total of 1,477 cases. These results suggest a relatively strong model performance in predicting customer churn.

A screenshot of a computer

Description automatically generated

**How can management make use of the model**

Based on the analysis, the decision tree model offers several application in customer retention classification. The model demonstrates high accuracy, successfully predicting customer churn in about 87% of cases, indicating its reliability in identifying customers likely to leave. It highlights key factors influencing churn, such as age and income, which helps management understand the demographic and behavioral characteristics of at-risk customers.

Practically, this model can guide management in implementing targeted retention strategies. For instance, personalized offers and improved customer service can be directed at identified at-risk customers to encourage their continued patronage. Additionally, understanding the factors contributing to churn allows for more effective resource allocation, enabling marketing efforts to focus on demographics that show a higher likelihood of leaving. The model can also inform product and service improvements, ensuring that offerings align better with customer needs and expectations. Customer segmentation can be refined based on churn likelihood, allowing for more tailored communication strategies that resonate with specific groups. Overall, these applications help retain customers while fostering loyalty and enhancing overall customer satisfaction.

**Appendix**

A screenshot of a computer

Description automatically generated

|  |
| --- |
|  |

A screenshot of a calculator

Description automatically generated

**Predicted values table**

A screenshot of a computer

Description automatically generated

**Decision Tree visualization**

A computer screen shot of a computer generated diagram

Description automatically generated