**Predicting Customer Churn in a Telecommunications Company**  
 Varala Enosh Reddy

1. **Introduction**

Customer churn is a significant concern for businesses as it directly impacts revenue and growth. Predicting customer churn allows companies to proactively engage at-risk customers and implement retention strategies. This report details the process of building a churn prediction model, including data preprocessing, model training, evaluation, and analysis.

1. **Exploratory Data Analysis(EDA)**

Exploratory Data Analysis (EDA) is a critical step in understanding the dataset and uncovering patterns, anomalies, and insights. It involves summarizing the main characteristics of the data, often using visual methods.

* 1. Summary Statistics & Missing Values

We examine summary statistics to understand the central tendency, dispersion, and shape of the dataset’s distribution. Identifying and handling missing values is crucial. We check for any missing values in the dataset.

* 1. Column Wise Data Analysis to understand customer behavior and factors influencing churn.

We are splitting the columns into many different kinds for better visualization as Customer\_information column and done a visualization with a hue of churn

A graph of different colored bars

Description automatically generated with medium confidence

Numeric\_columns and done a visualization with a hue of churn

A graph of a graph

Description automatically generated with medium confidence

A graph of a graph

Description automatically generated with medium confidence

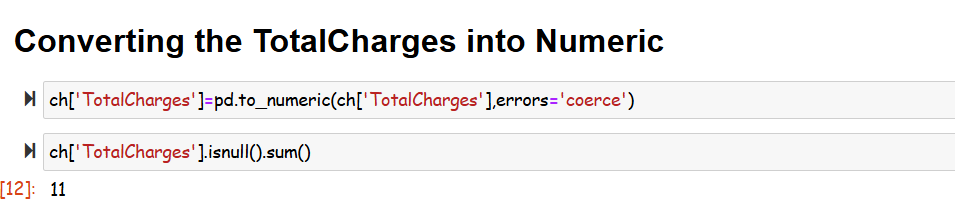
A detailed analysis of the tenure feature reveals a significant insight: customers with lower tenure are more likely to churn.

1. **Data Preprocessing**

Data preprocessing is a critical step in preparing the dataset for modeling. It involves handling missing values, encoding categorical variables, scaling numerical features, and splitting the data into training and test sets.

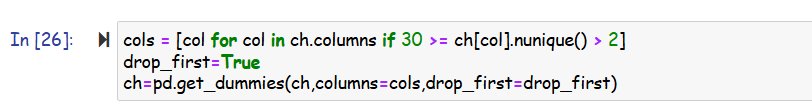
* 1. Handling Missing Values

The TotalCharges column had some missing values which were imputed with the column's mean to ensure no data is lost and to maintain consistency.

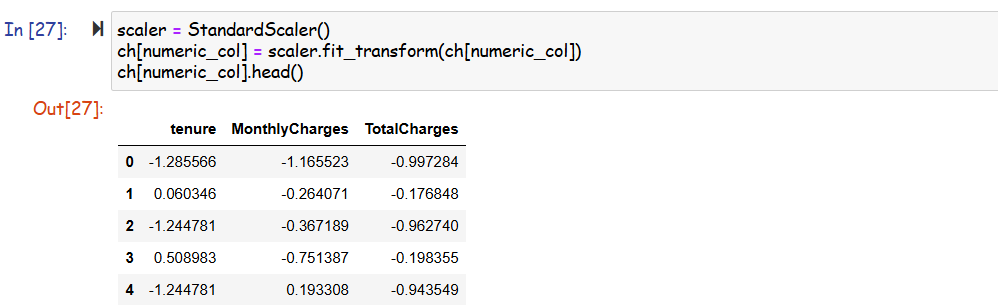


* 1. Encoding Categorical Variables

Categorical variables were converted into dummy variables using pd.get\_dummies to enable the model to interpret these features correctly.



* 1. Scaling Numerical Features

Numerical features were scaled using StandardScaler to standardize the data, ensuring each feature contributes equally to the model's performance.

* 1. Splitting the Data

The dataset was split into training and test sets to evaluate the model's performance on unseen data.

A computer code with black text

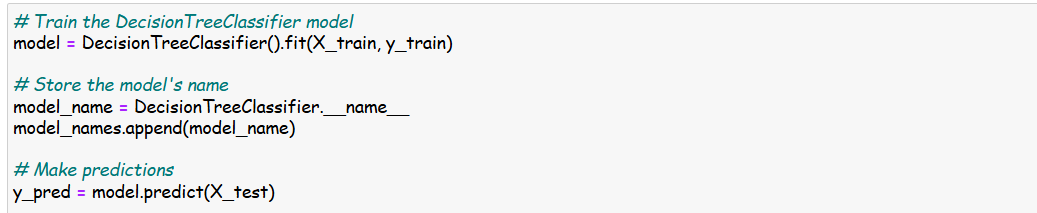
Description automatically generated

1. **Model Training and Evaluation**

DecisionTreeClassifier, LogisticRegression , GradientBoostingClassifier were trained and evaluated using accuracy, precision, recall, F1-score, and a classification report.

* 1. Model Training

The models were trained on the training dataset.



* 1. Model Evaluation

The model's performance was evaluated on the test dataset using several metrics:

A computer screen shot of a computer screen

Description automatically generated

THE END