**USECASE 1.2**

**TELE-COMMUNICATION CUSTOMER CHURN**

## **Problem Statement:**

Build a predictive model to estimate the likelihood of customer churn for a telecommunication company. The model should consider factors such as customer demographics, usage patterns, billing history, and customer service interactions.

## **Sample Dataset:**

Customer Churn Prediction: https://www.kaggle.com/datasets/blastchar/telco-customer-churn

### DATA ANALYSIS OF NUMERICAL AND CATEGORICAL COLUMNS:

Started by analysing numerical columns, calculating descriptive statistics (count, min, max, mean, median, std, skewness, kurtosis) to understand their distribution and characteristics. A correlation matrix and heatmap were then used to visualize relationships between numerical features, identifying potential multicollinearity.

Turning to categorical columns, computed value counts and percentages to examine their distribution. Bivariate analysis was conducted to investigate how categorical variables relate to churn, revealing significant differences in churn rates across categories. Visualizations such as count plots and stacked bar plots were used to illustrate these relationships, uncovering patterns that inform further analysis.

### BIVARIATE ANALYSIS

Conducted bivariate analysis to examine how churn varies with important numerical variables: tenure, monthly charges, and total charges. Using Kernel Density Estimate (KDE) plots, we compared the distributions of 'Tenure', 'Monthly Charges', and 'Total Charges' between churned and non-churned customers. These plots provide insights into the differences in distribution between customers who churned and those who did not, highlighting potential patterns and relationships that could influence churn.

**INSIGHTS:**

* **High Churn for:**
* Customers who opted for 'Phone Service' and Internet service as 'Fibre Optics'.
* Customers who have less 'Contract Period' and payment method as 'Electronic Check'.
* Customers with lower tenure, lower total charges, and higher monthly charges.

**RECOMMENDATIONS:**

* Combining phone and fibre optic internet services can often result in cost savings compared to having separate plans.
* Provide real-time payment confirmation, faster check clearing, and e-check verification services for electronic check payments.
* Consider offering flexible contract plans, exceptional customer support, competitive pricing, and promotions.

### MACHINE LEARNING MODEL BUILDING APPROACH

To predict customer churn in a telecom dataset, several machine learning models were implemented and evaluated. Here's a theoretical overview of the approach

**DATA PREPARATION AND TRANSFORMATION:**

The data was split into input features (X) and the target variable (y) with a 75/25 train-test split. Categorical features were one-hot encoded to convert them into a numerical format. Numerical features were standardized using StandardScaler to ensure all features contribute equally to model training. This approach prepares the data for training machine learning models to predict customer churn.

**MODELS ACCURACY’S:**

* **KNN: 0.79**
* **Decision Tree: 0.79**
* **Logistic Regression: 0.81**
* **Random Forest: 0.80**

**MODEL COMPARISION AND SELECTION:**

After evaluating each model, Logistic Regression achieved the highest test accuracy of 0.81. This indicates that it predicts customer churn with the highest accuracy among the models tested. Logistic Regression's regularization penalty parameter was set to 'l2' for optimal performance.

**CONCULSION:**

Building a machine learning model involves several critical steps, including data preparation, transformation, model training, hyperparameter tuning, and evaluation. Each step plays a crucial role in ensuring the model's accuracy and reliability. In this case, Logistic Regression emerged as the best model for predicting customer churn, achieving an accuracy of 81%. This model can now be deployed to predict and mitigate customer churn, thereby improving customer retention strategies.

