Name : Youssef Ahmed Mohammed Salama.

DEPI Assignment 1: Customer Churn Prediction for Tele Company.

Instructor: Dr. Abdallah Abouzeid.

Date: 30/10/2024.

# Data Science Workflow for Customer Churn

* **Problem Understanding**
  + **Goal:** Develop a model that flags customers likely to churn so the company can take steps to keep them on board.
* **Data Collection**
  + **Dataset:** This has customer details like age, services used, contract type, and churn status.
  + **Approach:** Start by skimming the data to understand what’s there and make note of anything we might want to adjust or add.
* **Data Preprocessing**
  + **Data Cleaning:**
    - **Missing Values:** Find and handle any missing data, either by filling gaps (mean/mode imputation) or dropping rows if they’re not essential.
    - **Outliers:** Identify extreme values that could distort the model.
  + **Normalization and Encoding:**
    - **Encoding:** Turn categorical fields (like contract types) into numerical values.
* **Exploratory Data Analysis (EDA)**
  + **Initial Exploration:**
    - Look at each feature and its relationship to churn through summary stats and plots.
  + **Patterns and Relationships:**
    - Use graphs (histograms, bar charts) and a correlation matrix to spot patterns. Example: Are short-term contract users more likely to leave?
  + **Insights:**
    - Pinpoint features strongly associated with churn, like high monthly charges.
* **Modeling**
  + **Model Selection:**
    - Try out models like logistic regression (simple and interpretable) and random forest or gradient boosting (for complex patterns).
  + **Model Training:**
    - Train and test models using part of the data to find the best approach.
  + **Feature Importance:**
    - In tree models, check feature importance to see what drives churn most.
* **Evaluation and Tuning**
  + **Metrics:**
    - Use metrics like precision and recall, which are useful for spotting churn in imbalanced data.
  + **Cross-Validation:**
    - Use cross-validation to make sure the model performs well on new data.
* **Deployment and Monitoring**
  + **Making Predictions:**
    - Once tuned, use the model to flag at-risk customers so the company can offer them incentives.
  + **Tracking Model Performance:**
    - Regularly update the model with new data for better accuracy.
  + **Insights to Action:**
    - For example, if short contracts correlate with churn, the company could offer discounts for longer contracts.