**Project Name: Customer Churn Analysis for Telecom Industry**

**Introduction**

In the rapidly expanding and extremely competitive telecom industry, already keeping a customer is top priority. An important issue is customer churn, that is, when users stop using a service. This project aims to mine customer behaviours for significant churn-relevant factors. The ultimate objective is to create a predictive model which enables the company to know customers at high risk of churn and therefore prevent the churn.

**Abstract**

The aim of this project is to understand the following with the help of the telecom data: 1. To predict the churn in a functional telecom. Key objectives included:

* Conducting a SQL-based data aggregation in order to aggregate churn behaviour by contract type, payment method, and internet use.
* Constructing a binary classification model for customers' churn prediction.
* Interpreting Model Decisions with ELI5 To Find Key Churn Drivers.
* Grouping customers in At Risk, Loyal, Dormant or other such categories for enhanced strategy planning.

The project was able to identify the main causes of churn and created Random Forest model with 80% accuracy to help with the proactive retention.

**Tools Used**

* **Python** (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, ELI5)
* **SQL-style data manipulation** (via Pandas)
* **Google Colab** for analysis, modelling, presentation and code execution

**Steps Involved in Building the Project**

1. **Data Cleaning & Preprocessing**
   * Removed null values and irrelevant columns (e.g., customerID)
   * Applied one-hot encoding to handle categorical variables
2. **Exploratory Data Analysis (EDA)**
   * Month-to-month customers churned the most (42.7%)
   * Electronic check users had the highest churn rate (45.3%)
   * Customers with high monthly charges and short tenure showed higher churn
3. **SQL-Based Aggregations**
   * Aggregated churn count by contract type and payment method
   * Identified Fiber Optic users as high-paying and high-churn customers
   * Found that automatic payments reduce churn significantly
4. **Churn Prediction Model**
   * Used **Random Forest Classifier**
   * Train-test split (80-20), model accuracy = **~80%**
   * Confusion Matrix showed:
     + True Positives: 173
     + True Negatives: 951
     + Model was better at detecting non-churners than churners
5. **Model Explainability (ELI5)**
   * Top contributing features: **Tenure**, **Contract Type**, **Monthly Charges**, **Electronic Check**
   * Helped interpret "why" a customer might churn
6. **Customer Segmentation**
   * **At Risk**: Month-to-month, electronic check users, high charges
   * **Loyal**: Long tenure, two-year contract, auto-pay
   * **Dormant**: Low activity, low billing customers

**Conclusion**

This project gave clear insights into which customers are likely to churn and why. By combining data analysis with machine learning and explainability tools, we not only predicted churn but also uncovered practical reasons behind it. Companies can now use this knowledge to offer better plans, promote automatic payments, or target at-risk customers with special offers. The approach can also be applied to other businesses that deal with subscription services or customer retention challenges.