**📄 Customer Churn Analysis and Prediction**

**Task 4: Churn Prediction Model**

**Objective**

The objective of this task is to develop a churn prediction model using machine learning techniques to help the company identify customers who are at risk of leaving. This enables proactive retention strategies and better resource allocation.

**Description**

In this task, a **logistic regression model** was implemented to predict whether a customer will churn based on their attributes and service usage patterns. The model was trained on the **Telco Customer Churn** dataset and evaluated using standard performance metrics.

**Methodology**

1️⃣ **Data Source:**

* Dataset: *Telco Customer Churn*
* URL: https://raw.githubusercontent.com/IBM/telco-customer-churn-on-icp4d/master/data/Telco-Customer-Churn.csv

2️⃣ **Data Preprocessing:**

* Removed customerID column (not useful for prediction).
* Converted TotalCharges to numeric and handled missing values.
* Encoded categorical features using label encoding.
* Scaled numerical features using **StandardScaler**.

3️⃣ **Model Development:**

* Used **logistic regression** as the baseline machine learning model.
* Split the dataset into training (70%) and testing (30%) sets using stratified sampling.
* Trained the model and generated predictions on the test set.

4️⃣ **Model Evaluation:**  
The model's performance was evaluated using:

* **Accuracy** – proportion of correct predictions
* **Precision** – proportion of true positives among predicted positives
* **Recall** – proportion of true positives among actual positives
* **F1 Score** – harmonic mean of precision and recall
* **ROC-AUC Score** – measure of the model’s ability to distinguish between classes

**Skills Applied**

* **Machine learning expertise** — applied logistic regression for binary classification.
* **Programming skills** — implemented the model using Python, with libraries like Pandas, Scikit-learn, Matplotlib, and Seaborn.
* **Model evaluation** — analyzed accuracy, precision, recall, F1 score, and ROC-AUC to assess the model’s performance.

**Results**

* **Accuracy:** *0.8005*
* **Precision:** *0.6429*
* **Recall:** *0.5615*
* **F1 Score:** *0.5994*
* **ROC-AUC Score:** *0.8355*

The confusion matrix and ROC-AUC indicated that the model performed well in identifying customers at risk of churn, with balanced precision and recall.

**Conclusion**

The logistic regression model provides a reliable baseline for churn prediction. The model’s predictions can help the company target at-risk customers more effectively. Future enhancements could include trying more complex algorithms (e.g., Random Forest, XGBoost) and tuning hyperparameters for improved performance.