■ Project 2 – Telco Customer Churn Prediction

Objective

To analyze telecom customer data and predict churn (whether a customer will leave).
Identify key factors that drive customer churn.

Dataset

Source: Telco Customer Churn dataset (Kaggle).
Rows: ~7,000 customers.
Key features: tenure, contract type, payment method, monthly charges, total charges, internet services, customer demographics.

Tools & Technologies

Programming Language: Python
Libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn
Environment: Kaggle Notebooks

Workflow

1. Data Preprocessing
- Dropped unnecessary columns (e.g., customerID).
- Converted TotalCharges to numeric & filled missing values.
- One-hot encoded categorical variables.
2. Exploratory Data Analysis (EDA)
- Churn distribution (Yes/No).
- Churn by contract type, tenure, internet service, monthly charges.
- Insights into customer behavior patterns.
3. Model Building
- Logistic Regression model (with balanced classes).
- Train-test split (80-20).
4. Model Evaluation
- Metrics: Accuracy, Precision, Recall, F1-score, ROC-AUC.
- Visualized results with confusion matrix.
5. Insights
- Customers with month-to-month contracts, short tenure, and high charges are more likely to churn.
- Long-term contracts and automatic payment methods reduce churn risk.

Status

■ Completed – Dataset cleaned, analyzed, model trained, evaluation done, insights documented.