CUSTOMER CHURN PREDICTION USING MACHINE LEARNING

1. Introduction
Customer churn refers to the loss of customers when they stop using a company's service. In subscription-based businesses, retaining customers is critical for long-term profitability. Predicting customer churn using machine learning helps organizations identify at-risk customers and take proactive steps to improve retention.

2. Objective
The objective of this project is to develop a machine learning model that predicts whether a customer will churn based on historical customer data such as usage behavior and demographics.

3. Dataset Description
The dataset contains customer information including demographic details, service usage, and account-related attributes. The target column is **Churn**, where: 0 → Customer stays 1 → Customer churns The dataset includes both numerical and categorical features.

4. Data Preprocessing
The following preprocessing steps were performed: Removal of irrelevant and duplicate records Handling missing values Encoding categorical variables into numerical form Separating features and target variable The dataset was split into training and testing sets for model evaluation.

5. Model Selection
Logistic Regression and Random Forest classifiers were considered for churn prediction. Logistic Regression was chosen as the baseline model due to its simplicity and interpretability.

6. Model Training
The selected model was trained using the training dataset. It learned patterns related to customer behavior that indicate the likelihood of churn.

7. Model Evaluation
The model performance was evaluated using: Accuracy Precision Recall F1-score Confusion Matrix Recall was prioritized to ensure customers likely to churn were correctly identified.

8. Results
The model achieved satisfactory accuracy and effectively identified customers at risk of churn. Evaluation metrics showed balanced performance between precision and recall.

9. Conclusion
This project demonstrates the effectiveness of machine learning techniques in predicting customer churn. Such models enable businesses to design targeted retention strategies and reduce customer loss.

10. Future Enhancements
Use advanced models like Gradient Boosting or XGBoost Apply feature scaling and feature selection Handle class imbalance using resampling techniques Deploy the model for real-time churn prediction

11. Tools & Technologies Used
Python, Jupyter Notebook, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn.

12. Final Outcome
The project successfully predicts customer churn and fulfills the requirements of Task 3 of the CODSOFT Machine Learning Internship.