**CUSTOMER CHURN ANALYSIS AND PREDICTION**

RD INFRO TECHNOLOGY – Internship Project

TASK 1 : Data Cleaning and Preprocessing

NAME : S.SYLVIASHARON

PROJECT OVERVIEW

The aim of this project is to analyze customer churn in a telecommunications company and develop predictive models to identify at-risk customers. The insights generated from this analysis will help the company reduce churn and improve customer retention.

**Task 1: Data Cleaning and Preprocessing**

**Description:**

* **Handle missing values** in the dataset to ensure no data integrity issues affect the model.
* **Convert categorical variables** into numerical formats using encoding techniques like One-Hot Encoding.

### ****Dataset Overview****

* Dataset Name: Telco Customer Churn
* Total Records: 7,043
* Attributes: Customer ID, Gender, Senior Citizen, Partner, Tenure, Monthly Charges, Total Charges, etc.
* Target Variable: Churn

**Steps Performed**

**1. Loading and Initial Inspection**

* Used Pandas to load and view dataset structure using .info(), .head(), .isnull().sum().

**2. Handling Missing Values**

* Identified missing values in the TotalCharges column due to blank spaces.
* Converted TotalCharges to numeric.
* Applied **median imputation** using SimpleImputer from Scikit-learn to fill missing values.

**3. Dropping Irrelevant Columns**

* Dropped the customerID column as it does not contribute to prediction.

**4. Encoding Categorical Features**

* **Label Encoding** was applied to binary columns like gender, Partner, Dependents, Churn, etc.
* **One-Hot Encoding** was applied to multi-category columns like Contract, InternetService, PaymentMethod, etc., using pd.get\_dummies().

**Cleaned Data Summary**

* All missing values handled.
* All features converted to numerical format.
* Final shape: 7043 rows × 30+ columns (after encoding)
* Data is now ready for visualization and modeling.

### ****Tools & Technologies****

* Language: Python
* Libraries: Pandas, NumPy, Scikit-learn
* IDE: Vs Code

### ****Skills Demonstrated****

* Data Cleaning and Inspection
* Handling Missing Values
* Feature Engineering
* Encoding Techniques (Label Encoding, One-Hot Encoding

### ****Conclusion****

The dataset has been successfully cleaned and preprocessed as per the task requirements. It is now suitable for applying machine learning algorithms in the next steps of the project. The transformations applied ensure improved model performance and accuracy.