Customer Churn Analysis - Project Report

Introduction

This project focuses on analyzing and predicting customer churn using machine learning techniques. The aim is to identify patterns in the data that indicate whether a customer is likely to discontinue services, helping businesses take proactive measures to retain them.

Objective

To predict customer churn based on historical data and identify the key factors contributing to it.

Tools & Technologies Used

- Python

- Jupyter Notebook

- Pandas, NumPy

- Matplotlib, Seaborn

- Scikit-learn (Logistic Regression)

Dataset Description

The dataset includes various customer-related features such as demographics, account information, service usage, and churn status. Key columns include:

- CustomerID
- Gender
- SeniorCitizen
- Partner
- Dependents
- Tenure
- PhoneService
- InternetService
- MonthlyCharges
- TotalCharges
- Churn (Target)

Project Workflow (Step-by-Step Explanation)

1. Data Cleaning

  - Removed any unnecessary whitespace from column names.
  - Ensured there were no missing or null values in key fields.

2. Data Preprocessing

  - Converted string-based numerical columns (like TotalCharges) to actual numeric types.
  - Categorical variables were encoded using one-hot encoding or binary encoding.

3. Feature Selection

  - Selected relevant features that had the most impact on churn.
  - Removed unnecessary columns that did not contribute to the model.

4. Model Building

  - Split the dataset into training and testing sets.
  - Used logistic regression as the classification model.

5. Model Evaluation

  - Evaluated the model using accuracy score and classification report.
  - Interpreted precision, recall, and F1-score to assess performance.

✔Outcome

The model was able to predict churn with reasonable accuracy. Key features like Tenure, InternetService, and MonthlyCharges were found to have a significant impact on churn. Insights from the model can help businesses take strategic actions to improve customer retention.