Telco Customer Churn Prediction Project Report

1. Objective

The primary goal of this project is to **predict customer churn** — i.e., whether a customer is likely to discontinue services — using historical customer data from a telecom company. Early identification of churn helps businesses design **targeted retention strategies** and improve customer satisfaction.

---

2. Dataset Overview

- **Dataset:** Telco Customer Churn
- **Total Records:** 7,043
- **Key Features:**
    - **Demographic details:** Gender, SeniorCitizen, Partner, Dependents
    - **Account information:** Tenure, Contract, PaymentMethod, PaperlessBilling
    - **Service usage:** InternetService, PhoneService, StreamingTV, StreamingMovies
    - **Financials:** MonthlyCharges, TotalCharges
    - **Target variable:** Churn (Yes/No)

---

3. Data Preprocessing

To prepare the dataset for model training, several preprocessing steps were performed:

- **Handling missing values:** Missing entries in TotalCharges were replaced with the median value.
- **Encoding categorical variables:** Label Encoding was applied to convert text-based features into numerical values.
- **Feature scaling:** Numerical variables were standardized to ensure uniform scale for the logistic regression model.
- **Train-test split:** The dataset was split into **80% training** and **20% testing** data.

---

4. Exploratory Data Analysis (EDA)

EDA revealed the following insights:

- **Average Tenure:** 32.37 months
- **Average Monthly Charges:** ₹64.76
- **Most common Internet Service:** Fiber optic (43.96%)
- **Top Payment Method:** Electronic check
- Customers with **month-to-month contracts** and **electronic payment methods** showed higher churn rates.

These insights highlight behavioral and contractual factors influencing churn.

---

5. Model Used

Model: Logistic Regression

Logistic Regression was chosen for its simplicity, interpretability, and effectiveness in binary classification problems like churn prediction. It predicts the probability of churn (Yes/No) based on feature relationships.

---

6. Model Performance

Metric	Score
Accuracy	~0.8147
ROC-AUC Score	0.861

The ROC-AUC score indicates the model's ability to distinguish between churners and non-churners.
A value of **0.861** shows the model performs well, though there is room for improvement through feature engineering or resampling.

---

7. Insights from Model

- Customers with **month-to-month contracts**, **higher monthly charges**, and **fiber optic internet** are more likely to churn.
- Long-term contract customers (one or two years) tend to stay longer.
- Customers using **electronic checks** have a higher churn probability compared to automatic payments.

---

8. Dashboard Overview

The **Telco Customer Churn Dashboard** (see image) provides a **visual summary** of key insights:

- **Total Records:** 7,043 customers
- **Average Tenure:** 32.37 months
- **Average Monthly Charges:** ₹64.76
- **Gender Comparison:** Male and female customers show similar average spending patterns.
- **Service Distribution:** 44% use Fiber optic, followed by DSL and No Internet service.
- **Payment Preferences:** Majority use **electronic check**, indicating potential link with churn.
- **Contract Insights:** Customers with month-to-month contracts pay slightly higher average charges, showing possible dissatisfaction or flexibility preference.

The dashboard allows management to:

- Quickly analyze churn-related patterns
- Filter by demographic or service categories
- Identify customer segments most at risk

---

9. Conclusion

The **Logistic Regression model** successfully predicts customer churn with good accuracy and discrimination ability.

Key takeaways:

- **Churn is influenced** by contract type, monthly charges, and payment method.
- **Longer contracts and auto-payments** reduce churn risk.
- The **dashboard visualization** aids in strategic decision-making by simplifying data interpretation.