📊 Customer Churn Prediction – Telecom Industry

📌 Project Overview

Customer churn is one of the biggest challenges in the telecom industry. This project focuses on analyzing customer data and building a Machine Learning model to predict whether a customer is likely to leave (churn) or stay with the company.

The objective of this project is to:

- Perform **data cleaning and preprocessing**
- Conduct **Exploratory Data Analysis (EDA)**
- Build and train a **predictive model**
- Evaluate model performance
- Provide actionable business insights

📁 Dataset Description

The dataset used in this project contains telecom customer information such as:

- Customer demographics
- Account information
- Service subscriptions
- Monthly and total charges
- Tenure
- Churn status (Target Variable)

Target Variable:

- **Churn**
  - Yes → Customer left the company
  - No → Customer stayed

🛠️ Technologies Used

- **Python**
- **Pandas** – Data manipulation
- **NumPy** – Numerical operations
- **Matplotlib** – Data visualization
- **Scikit-learn** – Machine Learning
- **Jupyter Notebook** – Development environment

🔎 Project Workflow

1️⃣ Data Loading

- Imported dataset using Pandas
- Inspected data using .head(), .sample(), .shape(), .dtypes()

2️⃣ Data Cleaning & Preprocessing

- Removed unnecessary column: customerID
- Converted TotalCharges from object to numeric
- Handled missing values using:

pd.to_numeric(df.TotalCharges, errors='coerce')

- Removed rows with blank TotalCharges
- Checked for null values
- Converted categorical variables into numerical form (Encoding)

3️⃣ Exploratory Data Analysis (EDA)

Key visualizations performed:

- 📊 Tenure vs Churn
- 📊 Monthly Charges vs Churn
- 📊 Histogram comparison for churned vs non-churned customers

Insights:

- Customers with **low tenure** are more likely to churn.
- Higher monthly charges slightly increase churn probability.
- Long-term customers are less likely to leave.

4️⃣ Feature Engineering

- Converted categorical features using:
  - Label Encoding
  - One-Hot Encoding
- Scaled numerical features for better model performance

- Separated:
    - **X (Independent variables)**
    - **y (Target variable)**

5️⃣ Model Building

Applied Machine Learning algorithms to predict churn.

Typical models used:

- Logistic Regression
- Artificial Neural Network (if implemented)
- Other classification models (optional)

Model training steps:

- Train-test split (80%-20%)
- Model fitting
- Prediction on test data

6️⃣ Model Evaluation

Evaluated using:

- Accuracy Score
- Confusion Matrix
- Classification Report
- Precision
- Recall
- F1 Score

Evaluation helps determine:

- How well the model predicts churn
- Whether the model is biased toward one class

📈 Key Findings

- Customers with short tenure are more likely to churn.
- High monthly charges increase churn risk.

- Customers with long-term contracts are more stable.
- Proper preprocessing significantly improves model accuracy.

🧠 Business Impact

This model can help telecom companies:

- Identify high-risk customers
- Implement retention strategies
- Offer targeted promotions
- Reduce revenue loss
- Improve customer lifetime value

📁 Project Structure

```
Customer-Churn-Prediction/
│
├── Customer_churn.csv
├── churn_prediction.ipynb
├── README.md
```

▶️ How to Run the Project

1. Clone the repository:

git clone <repository-link>

2. Install required libraries:

pip install pandas numpy matplotlib scikit-learn

3. Open Jupyter Notebook:

jupyter notebook

4. Run all cells in sequence.

🚀 Future Improvements

- Apply advanced models (Random Forest, XGBoost)
- Hyperparameter tuning
- Cross-validation

- Deploy model using Flask/Streamlit
- Build a dashboard for visualization
- Implement real-time churn prediction system

🎯 Conclusion

This project demonstrates an end-to-end data analysis and machine learning workflow for churn prediction. It highlights how data preprocessing, EDA, and model building together help solve real-world business problems in the telecom industry.