Customer Churn Prediction using Machine Learning

1. Introduction

Customer churn (or customer attrition) is when clients stop using a company's products or services. Predicting churn allows businesses to take proactive measures to retain customers and reduce revenue loss.

This project applies **machine learning models** to predict whether a customer is likely to churn based on historical data and customer attributes.

---

2. Objectives

1. To analyze customer data and identify key factors contributing to churn.
2. To build and evaluate multiple machine learning models for churn prediction.
3. To provide actionable insights that help improve customer retention strategies.

---

3. Dataset

The dataset contains customer demographic, account, and usage details.

1. Target variable: **Churn** (Yes/No).
2. Features include: customer demographics, subscription details, contract type, tenure, monthly charges, total charges, etc.

---

4. Methodology

1. Data Preprocessing

- Handled missing values.
- Converted categorical variables into numerical form (encoding).
- Scaled numerical features for uniformity.

2. Exploratory Data Analysis (EDA)

- Visualized churn distribution.

- Examined relationships between churn and features such as tenure, contract type, and charges.

3. Model Development

- Trained multiple ML models:
- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine (SVM)
- Gradient Boosting / XGBoost
- Compared accuracy, precision, recall, F1-score, and ROC-AUC.

4. Model Evaluation

The best-performing model was selected based on recall and ROC-AUC (important for churn detection).

---

5. Results

Key churn indicators identified: short tenure, high monthly charges, and month-to-month contracts.

Among models tested, **Random Forest / XGBoost** gave the highest performance with strong recall (catching most churn cases) and good overall accuracy.

Visualizations such as feature importance plots highlighted the most influential factors in churn prediction.

---

6. Conclusion

Machine learning can effectively predict customer churn. Retention strategies should target customers with month-to-month contracts, high charges, or short tenure.

Future work could include:

1. Hyperparameter tuning with GridSearch/RandomizedSearch.
2. Deploying the model as a web app (using Flask/Streamlit).
3. Using deep learning techniques for improved predictions.

7. Tools & Libraries

- Python
- Pandas, NumPy (data handling)
- Matplotlib, Seaborn (visualization)
- Scikit-learn (ML models & evaluation)
- XGBoost (boosted trees model)

---

8. References

Scikit-learn Documentation

XGBoost Documentation

Telecom churn datasets (Kaggle / IBM sample datasets)