

Telecom Customer Churn Prediction Project Documentation

1. Project Objective

The objective of this project is to predict customer churn in a telecom company using Machine Learning models such as Logistic Regression and Random Forest. The aim is to identify key factors influencing churn and provide business insights.

2. Dataset Overview

- Total Records: 2500
- Features: 7 Independent Variables + 1 Target Variable (Churn)
- Target Variable: Churn (Yes/No)
- Categorical Features: Contract Type, Internet Service, Payment Method
- Numerical Features: Tenure, Monthly Charges, Total Charges

3. Data Preprocessing

- Performed Label Encoding for categorical variables.
- Applied Standard Scaling to numerical features.
- Split dataset into Training and Testing sets.

4. Model Building

Model 1: Logistic Regression

- Trained using scaled data.
- Evaluated using Confusion Matrix and Classification Report.

Model 2: Random Forest Classifier

- Used to compare performance and feature importance.

5. Model Evaluation Metrics

- Confusion Matrix
- Accuracy Score
- Precision, Recall, F1-Score

6. Key Insight

From Logistic Regression coefficient analysis, Poor Internet Service was identified as the most influential factor affecting customer churn. Customers experiencing poor internet connectivity are more likely to leave the telecom service provider.

7. Business Recommendations

- Improve internet infrastructure quality.
- Provide compensation or discounts for affected customers.
- Offer loyalty programs for high-risk churn customers.
- Monitor service complaints proactively.

8. Conclusion

The model successfully predicts customer churn and highlights critical factors driving customer dissatisfaction. Logistic Regression provides clear interpretability through coefficients, while Random Forest improves prediction robustness.