

■ Telecom Customer Churn Prediction – Final Report

1. Introduction

Customer churn, the phenomenon where subscribers discontinue a service, is a critical challenge in the telecommunications industry. High churn rates lead to revenue losses and increased customer acquisition costs. This project focuses on predicting customer churn using machine learning techniques. By identifying the most important factors influencing churn, the company can design targeted retention strategies and reduce revenue leakage.

2. Methodology

2.1 Data Collection and Preparation

Dataset: Telecom customer dataset (Churning.csv) containing customer demographics, subscription details, billing information, and churn labels. Cleaning steps performed: - Removed irrelevant column (customerID). - Handled missing/invalid values in TotalCharges. - Normalized categorical responses (e.g., 'No internet service' → 'No'). - Encoded target variable (Churn) into binary (0 = No, 1 = Yes). - Converted categorical columns to category datatype. Final dataset was stored as Churning_cleaned.csv.

2.2 Exploratory Data Analysis (EDA)

- Distribution analysis revealed class imbalance: ~73% customers stayed, ~27% churned. - Key patterns: • Tenure: Customers with shorter tenure were more likely to churn. • Contract Type: Month-to-month contracts had higher churn rates compared to yearly contracts. • Monthly Charges: Higher charges were linked with increased churn.

2.3 Data Preprocessing

- Applied OneHotEncoding for categorical variables. - Standardized numerical variables (tenure, MonthlyCharges, TotalCharges). - Created train-test split (80% training, 20% testing).

2.4 Model Training and Evaluation

Models compared: - Logistic Regression - Decision Tree - Random Forest - Gradient Boosting
Evaluation metrics: Accuracy, Precision, Recall (Sensitivity), F1-Score, AUC-ROC.

3. Results

3.1 Best-Performing Model

Random Forest Classifier gave the best balance of performance: - Accuracy: ~82% - Precision: ~77% - Recall (Sensitivity): ~70% - F1-score: ~74% - AUC-ROC: ~0.84

3.2 Important Features

Top predictors of churn (from feature importance ranking): 1. Tenure – Customers with longer tenure are less likely to churn. 2. Contract Type – Month-to-month contracts strongly linked with churn. 3. Internet Service Type – Fiber optic users churn more than DSL users. 4. Monthly Charges – Higher charges increase likelihood of churn. 5. Online Security & Tech Support – Customers without these services are more likely to churn.

4. Discussion

- Contract type and tenure are the strongest predictors of churn, confirming business logic.
- Customers on month-to-month contracts face little switching cost, making them highly vulnerable.
- High monthly charges cause dissatisfaction, especially with limited value-added services.
- Customer support services (tech support, online security) strongly influence retention.

5. Conclusion and Recommendations

Conclusion: Machine learning can effectively predict customer churn, with Random Forest achieving strong performance. The analysis highlights key drivers of churn including contract type, tenure, and billing patterns. Recommendations: 1. Offer loyalty incentives to new customers to encourage long-term contracts. 2. Provide discounts or bundled offers to customers with high monthly charges. 3. Improve customer support services (tech support, online security). 4. Use the churn prediction model to flag high-risk customers for targeted retention campaigns.