

Predicting Customer Churn Using the Telco Customer Churn Dataset

Project Overview

This project focuses on predicting customer churn using the Telco Customer Churn dataset. The goal is to identify customers at risk of leaving and enable proactive retention strategies. The analysis involves data preprocessing, feature engineering, model training, and evaluation.

Dataset

The dataset used is `WA_Fn-UseC_-Telco-Customer-Churn.csv`, which contains 7043 rows and 21 columns. Key features include customer demographics, account information, and service usage details.

Key Steps

1. **Data Preprocessing:**
 - Handled missing values in `TotalCharges` by imputing the median.
 - Converted categorical features to numerical using one-hot encoding.
 - Scaled numerical features using `StandardScaler`.
2. **Feature Engineering:**
 - Created interaction features like `Tenure_MonthlyCharges`.
3. **Model Training:**
 - Trained Logistic Regression, Random Forest, and Gradient Boosting models.
 - Performed hyperparameter tuning using `GridSearchCV`.
4. **Evaluation:**
 - Identified Random Forest as the best model with an accuracy of 78.4% on the test set.
 - Evaluated metrics: Accuracy, Precision, Recall, F1-score, and AUC.

Results

- **Best Model:** Random Forest
- **Performance Metrics:**
 - Accuracy: 0.784
 - Precision: 0.777

- Recall: 0.784
- F1-score: 0.747
- AUC: 0.826
- **Feature Importance:**
 - Top features: `Tenure_MonthlyCharges`, `TotalCharges`, `MonthlyCharges`, and `InternetService_Fiber optic`.

Insights and Recommendations

1. **Target High-Risk Customers:** Focus on customers with shorter tenure and high monthly charges.
2. **Improve Fiber Optic Services:** Investigate the correlation between fiber optic internet and churn.
3. **Retention Strategies:** Offer incentives or improved services to reduce churn risk.

Next Steps

- Deploy the model for real-time churn prediction.
- Conduct further analysis on customer feedback and service quality.
- Explore additional features to enhance model performance.

Requirements

- Python 3.x
- Libraries: `pandas`, `matplotlib`, `seaborn`, `scikit-learn`, `numpy`

How to Run

1. Place the dataset in the project directory.
2. Run the script to preprocess data, train models, and evaluate results.
3. Review the output for insights and predictions.