Customer Churn Analysis and Prediction Report

Objective

The objective of this analysis is to understand factors contributing to customer churn in a telecom company and build a predictive model to identify customers likely to churn. By analyzing the data, we aim to provide actionable insights to improve customer retention.

Dataset Overview

The dataset contains information on telecom customers, including demographic details, services used, billing information, and churn status. Key details:

- **Rows (Records)**: 7,043
- **Columns (Features)**: 21
- **Target Variable**: Churn (Yes/No)

Data Cleaning

1. **Conversion of TotalCharges**:
    - TotalCharges had non-numeric values like blank entries. These were replaced with 0 and converted to numeric.
2. **Missing Values**:
    - No missing values were found after cleaning.

Exploratory Data Analysis (EDA)

1. Churn Distribution

- **Churned Customers**: 26.54%
- **Non-Churned Customers**: 73.46%

- **Key Insight**: The dataset is imbalanced, requiring techniques like oversampling or class weighting during model training.

2. Tenure Distribution by Churn

- **Churned Customers**:
  - Tend to have shorter tenures.
  - Many churned customers are new or recent users.
- **Non-Churned Customers**:
  - Have a wider range of tenures, with many being long-term users.
- **Key Insight**: Retention strategies should focus on customers with shorter tenures.

3. Monthly Charges by Churn

- **Churned Customers**:
  - Tend to have higher monthly charges.
- **Non-Churned Customers**:
  - Include a wider spread, with many paying lower charges.
- **Key Insight**: High charges may lead to dissatisfaction and churn. Offering competitive pricing or value-added services might reduce churn.

Feature Engineering

1. **Categorical Encoding**:
   - Categorical variables like gender, Contract, PaymentMethod were encoded using Label Encoding for machine learning compatibility.
2. **Exclusion of customerID**:
   - customerID was excluded from features as it has no predictive value.

Model Development

Random Forest Classifier

- **Model Selection**:
  - Random Forest was chosen for its robustness and ability to handle imbalanced datasets.
- **Training and Testing Split**:
  - **80% Training** and **20% Testing**

Evaluation Metrics

Confusion Matrix

- The confusion matrix highlights the performance in predicting churn:
  - True Positives (Correctly predicted churn): High
  - True Negatives (Correctly predicted no churn): High

Classification Report

- **Precision**: The proportion of correct positive predictions out of all predicted positives.
- **Recall**: The proportion of actual positives correctly predicted.
- **F1-Score**: Harmonic mean of precision and recall.

ROC-AUC Score:

- The model achieved a strong **ROC-AUC Score**, indicating good separation between churn and non-churn predictions.

Findings

1. **Churn Factors**:
   - Customers with higher monthly charges are more likely to churn.
   - Short-tenured customers are at higher risk of churn.
2. **Actionable Insights**:
   - **Retention Strategies**:
     - Focus on reducing churn for new customers (short tenure) with personalized engagement.
     - Consider revising pricing for high-paying customers or offering loyalty rewards.
   - **Targeted Marketing**:
     - Develop tailored campaigns for customers identified as high risk by the predictive model.

Conclusion

This analysis highlights the key drivers of customer churn and demonstrates the utility of a predictive model in identifying at-risk customers. By addressing factors like high monthly charges and lack of retention strategies for new customers, the telecom company can significantly reduce churn and improve customer satisfaction.