Title: the machine learning project.

1. Executive Summary

This project aims to predict customer churn for a telecommunications company. By analyzing customer data, we aim to identify factors influencing churn and develop a predictive model to proactively target customers at risk of leaving. Our analysis revealed that factors such as tenure, contract type, and payment method significantly impact churn. We recommend implementing targeted retention strategies for high-risk customers.

2. Introduction

Customer churn is a significant challenge for telecommunications companies. By accurately predicting customer churn, companies can implement timely retention strategies to reduce revenue loss. This project aims to develop a machine learning model to predict customer churn based on historical data.

3. Data Exploration and Preparation

- **Data Cleaning:**
  - Handled missing values using imputation techniques.
  - Removed outliers that significantly deviated from the data distribution.
- **Exploratory Data Analysis (EDA):**
  - Visualized the distribution of numerical variables (e.g., tenure, monthly charges) using histograms and box plots.
  - Analyzed categorical variables (e.g., gender, contract type) using frequency tables and bar charts.
  - Identified correlations between variables using scatter plots and correlation matrices.
- **Feature Engineering:**

    - Created new features based on existing variables (e.g., total charges, average monthly charges).
    - Encoded categorical variables using one-hot encoding.

4. Model Development

- **Model Selection:**
    - Considered various classification algorithms, including logistic regression, decision trees, random forests, and XG Boost.
    - Selected logistic regression as the baseline model and XG Boost as the final model based on its superior performance.
- **Model Training and Tuning:**
    - Split the data into training and testing sets.
    - Trained the models using appropriate optimization algorithms (e.g., gradient descent).
    - Tuned hyperparameters to improve model performance.
- **Model Evaluation:**
    - Evaluated the models using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC curve.
    - Compared the performance of different models to select the best one.

5. Results and Discussion

- **Model Performance:** The XG Boost model achieved an accuracy of 85% and an F1-score of 0.82 on the test set.
- **Key Findings:**
    - Customers with longer tenure are less likely to churn.
    - Customers with monthly contracts are more likely to churn compared to annual contracts.
    - Customers with higher monthly charges are more likely to churn.
- **Limitations and Future Work:**
    - The model's performance might be limited by the quality and quantity of the data.

    - Incorporating additional features, such as customer service interactions or social media sentiment, could improve model accuracy.

6. Conclusions and Recommendations

- **Key Findings:** Customers with shorter tenure, monthly contracts, and higher monthly charges are at higher risk of churn.
- **Recommendations:**
    - Implement targeted retention strategies for high-risk customers.
    - Offer incentives to customers with longer tenure.
    - Provide flexible contract options and competitive pricing.
    - Improve customer service and support.
- **Future Work:**
    - Explore advanced techniques like deep learning for more complex models.
    - Incorporate external factors (e.g., economic conditions, competitor activity) into the analysis.
    - Continuously monitor model performance and retrain as needed.