Customer Churn Prediction Project Report

Author: [Devan]

Date: [06/09/2023]

Introduction

This report presents the Customer Churn Prediction project, which aims to predict customer churn for a telecom company using machine learning techniques. The project focuses on leveraging historical customer data to identify potential churners and take proactive measures to retain them.

Dataset Overview

- **Dataset Source:** The dataset used in this project is sourced from an Excel file, located at '/home/devan/Downloads/customer_churn_large_dataset.xlsx'.
- **Dataset Description:** The dataset contains customer information, including demographic data, subscription details, and historical churn status. It serves as a crucial resource for training and evaluating the machine learning model.

Data Preprocessing

Initial Data Exploration

- **Data Load:** The dataset was successfully loaded from the Excel file, and its basic structure was explored.
- **Missing Data:** No missing data was detected in the dataset, ensuring the quality of the data.
- **Outliers:** An initial analysis revealed no significant outliers.

Feature Engineering

- **Categorical Encoding:** Categorical variables, such as 'Gender' and 'Location,' were encoded using LabelEncoder to convert them into numerical format for model training.
- **Feature Exclusion:** Non-numeric columns, 'CustomerID' and 'Name,' were excluded from the dataset as they do not provide relevant information for churn prediction.

Data Splitting

- The dataset was split into training and testing sets using an 80-20 split ratio. The random_state parameter was set to 42 for reproducibility.

Feature Scaling

- StandardScaler was applied to scale numerical features, ensuring that all features have a similar scale, which is a requirement for many machine learning algorithms.

Model Building

- **Model Selection:** A RandomForestClassifier was chosen as the machine learning model due to its ability to handle complex relationships within the data.
- **Model Training:** The model was successfully trained on the scaled training data.
- **Model Evaluation:** Model performance was evaluated on the test data using accuracy as the primary metric.

Model Performance

```
) python main.py
Accuracy: 0.49
Precision: 0.49
Recall: 0.47
F1-Score: 0.48
AUC-ROC: 0.49
```

- **Accuracy:** The trained model achieved an accuracy of [0.49], indicating the percentage of correct predictions on the test data.
- **Precision, Recall, F1-score:** [Include precision [0.49], recall [0.47], and F1-score [0.48] AUC-ROC [0.49] metrics here.]

Model Deployment

- The trained model has been saved to a file named 'churn_prediction_model.pkl' using joblib. It can be deployed in a production-like environment to make real-time predictions.
- An example of providing new customer data for prediction was demonstrated in the code.

Conclusion

- In conclusion, the Customer Churn Prediction project successfully developed a machine learning model capable of predicting customer churn. The model achieved [Accuracy Score] accuracy on the test data, indicating its effectiveness in identifying potential churners.
- Future work could involve further model optimization, exploring additional machine learning algorithms, and deploying the model for practical use within the organization.