Name:

Muhammad Awais

Roll no:

114

Report:

Customer Churn Prediction Project

1. Project Overview

The primary goal of this project is to develop a machine learning model that predicts customer churn. Churn refers to customers discontinuing their relationship with a business, which can lead to significant revenue loss. By identifying customers likely to churn, companies can take proactive steps to retain them and improve overall business performance.

---

2. Dataset Description

The dataset used for this project is the **Customer Churn Dataset**. It contains information about customer demographics, account details, and service usage.

Dataset Overview

- **Total Records**: 5,880

- **Total Features**: 21 (19 predictors + 1 target + 1 ID column)
- **Target Variable**: Churn (Yes/No)
- **Feature Types**:
    - **Categorical**: gender, Partner, Dependents, Contract, etc.
    - **Numerical**: tenure, MonthlyCharges, TotalCharges.

Key Features

- **Tenure**: How long the customer has been with the company.
- **Monthly Charges**: The monthly bill amount.
- **Contract Type**: Month-to-month, one year, or two-year contracts.
- **Internet Service**: Type of internet service used.

---

3. Data Preprocessing

Steps Taken

1. **Dropped Irrelevant Columns**: Removed the customerID column as it is non-informative.
2. **Encoding Categorical Variables**: Converted categorical variables into numerical form using label encoding.
3. **Handled Missing Values**: Checked for missing values; none were found in this dataset.
4. **Feature Scaling**: Not applied as the dataset already contained scaled features like MonthlyCharges.

Preprocessed Dataset

After preprocessing, the dataset included 19 independent features and 1 target variable (Churn).

---

5. Model Development

Model Used

A **Logistic Regression** model was chosen for its simplicity and interpretability. The model was trained to classify customers as either likely to churn (1) or not churn (0).

Train-Test Split

- **Training Set**: 80% of the data (4,704 records)
- **Testing Set**: 20% of the data (1,176 records)

Evaluation Metrics

- **Accuracy**: The percentage of correctly classified predictions.
- **Confusion Matrix**: Evaluates false positives and negatives.
- **Classification Report**: Provides precision, recall, and F1-score.

---

6. Results and Evaluation

Model Performance

- **Accuracy**: 79%
- **Confusion Matrix**:

lua

Copy code

[[820, 102],

[144, 110]]

    - **True Positives (Churn Predicted Correctly)**: 110
    - **True Negatives (No Churn Predicted Correctly)**: 820

- **False Positives**: 102
- **False Negatives**: 144

Classification Report:

Metric	Precision	Recall	F1-Score
Not Churn (0)	85%	89%	87%
Churn (1)	52%	43%	47%

Insights

- The model performs well in identifying customers who are unlikely to churn.
- Improvement is needed in predicting churn customers, possibly through better feature engineering or using more advanced algorithms.

---

7. Application Phase

An interactive web application was developed using **Flask**. The app allows users to input customer details and predict whether they are likely to churn.

App Features

- User-friendly interface to input customer details.
- Real-time churn prediction with probabilities for churn and retention.
- Suggestions for actions to reduce churn risk (future enhancement).

---

7. Conclusion

The project successfully developed a churn prediction model that achieves an accuracy of 79%. While the model works well for identifying non-churning customers, future work should focus on:

1. **Improving Churn Prediction**: Explore advanced algorithms like Random Forest, Gradient Boosting, or Neural Networks.
2. **Feature Engineering**: Investigate additional features that might influence churn, such as customer feedback or competitor offers.
3. **Deployment**: Enhance the web application with interactive dashboards and automated retention strategies.