Customer Churn Prediction - Project Report

1. Introduction

This project focuses on predicting customer churn using machine learning. The dataset includes various features about customer behavior and subscription patterns. A Random Forest Classifier is used to perform the classification task.

2. Data Preprocessing

The dataset is loaded from a CSV file. One of the critical preprocessing steps includes converting the 'TotalCharges' column to a numeric data type. Null values are visualized using a heatmap and dropped accordingly.

3. SQL Data Aggregation

Using SQLite, the data is loaded into an in-memory database to perform SQL queries. This includes computing contract-wise customer counts and average monthly charges.

4. Encoding and Splitting

Categorical features are encoded using LabelEncoder. The data is then split into training and test sets for model evaluation.

5. Model Training

A Random Forest Classifier is trained on the preprocessed dataset. Model evaluation is done using a classification report and a confusion matrix.

6. Explainability using SHAP

SHAP (SHapley Additive exPlanations) is used to interpret the model's predictions. This allows better insights into feature importance and model decisions.

7. Visualization

The project uses Seaborn and Matplotlib for visualizing null values, feature importance, and SHAP values.