Training and Testing Sets Documentation

Split Ratio

The original preprocessed dataset was split using an 80/20 ratio:

- Training set: 80% of the data
- Testing set: 20% of the data

The split was performed using train_test_split() from sklearn.model_selection and the random_state was 42.

Size of the Data

Total data points: 7043
Training set size: 5634
Testing set size: 1409

Features Included

- SeniorCitizen (encoded: 0/1)
- Dependents (encoded: 0/1)
- Tenure (numeric)
- PhoneService (encoded: 0/1)
- MultipleLines (encoded: 0/1)
- MonthlyCharges (numeric)
- Churn (encoded: 0/1)
- Gender (encoded: 0 = female, 1 = male)
- Contract_One year (encoded: 0/1)
- Contract_Two year (encoded: 0/1)
- InternetService (encoded: 0/1)

File Format

- The training set is saved as train_set.csv
- The testing set is saved as test_set.csv