Dataset Documentation

X_train_scaled

Shape: (5392, 11)

Features:

- SeniorCitizen (int)

- tenure (float, scaled)

- MonthlyCharges (float, scaled)

- TotalCharges (float, scaled)

- gender_Male (bool)

- Dependents_Yes (bool)

- PhoneService_Yes (bool)

- MultipleLines_Yes (bool)

- InternetService_Fiber optic (bool)

- Contract_One year (bool)

- Contract_Two year (bool)

Features are standardized (mean ~0, std ~1), except for 'SeniorCitizen' which is binary.

X_test_scaled

Shape: (1349, 11)

Same features and scaling as X_train_scaled.

Means and standard deviations align closely, indicating consistent scaling.

y_train

Shape: (5392, 1)

Target: Churn_Yes (binary classification)

Class Distribution:

- False (No Churn): 3963 (~73.5%)

- True (Churn): 1429 (~26.5%)

y_test

Shape: (1349, 1)

Target: Churn_Yes (binary classification)

Class Distribution:

- False (No Churn): 987 (~73.2%)

- True (Churn): 362 (~26.8%)