Preprocessing Report – Customer Churn Prediction

Introduction

This report documents the complete data preprocessing workflow used in the Customer Churn Prediction project. The aim is to convert raw customer data into a clean and machine-learning-ready format.

Dataset Overview

The dataset consists of 500 customer records containing demographic details, service contracts, and billing information. The target variable is customer churn.

Data Cleaning

Missing values were identified and removed. Data types were validated to ensure compatibility with preprocessing and modeling steps.

Categorical Data Handling

Label Encoding, One-Hot Encoding, and Binary Encoding were applied to convert categorical data into numerical form suitable for machine learning models.

Feature Scaling

Min-Max Scaling and Standard Scaling were applied to numerical variables to normalize feature ranges and improve model performance.

Outlier Detection

Outliers were detected using both the Interquartile Range (IQR) method and Z-score analysis, ensuring robust identification of extreme values.

Preprocessing Pipeline

A complete preprocessing pipeline was built using ColumnTransformer and Pipeline to ensure reproducibility and prevent data leakage.

Conclusion

The preprocessing steps resulted in a high-quality dataset suitable for predictive modeling and real-world deployment.