Telco Customer Churn

Mouaz Ahmed
Hyder Gritli

November 2025

1 Introduction

This project focuses on predicting customer churn in the telecommunications industry using machine learning classification techniques. We will analyze customer behavior patterns using features such as churn score, number of dependents, referrals, satisfaction score, and tenure in months. Customer churn prediction is a practical business problem with real-world applications in customer retention and revenue optimization. Previous research in this domain has primarily utilized algorithms such as Logistic Regression and Decision Tree Classifiers, and LightGBM Classifier. The previous work is a GitHub repository. [1]

2 Method

We will implement and compare three supervised machine learning algorithms:

- Logistic Regression.
- Decision Tree Classifier.
- Support Vector Machine.

Additionally, we will apply the following techniques to improve model performance:

- Categorical feature embedding.
- Ensemble methods (hard and soft voting).
- SMOTE for handling imbalanced data.
- Statistical analysis for identifying churn timing patterns

[1] https://github.com/Geo-y20/Telco-Customer-Churn-

3 Dataset

This merged Telco Customer Churn dataset provides a comprehensive view of customer attributes, service usage, location data, and churn behavior. This expanded Dataset is a valuable resource for understanding churn patterns, customer segmentation, and developing targeted marketing strategies.
The dataset is sourced from HuggingFace[2] website and structured as:

- Format: CSV.
- Size: 7,043 rows × 49 columns.
- Each row represents a unique customer with their associated attributes.

4 Midterm Milestone

Stage	Milestone	Date
Proposal	Introduce the problem and solution at high level and some information about dataset	15 Nov 2025
Midway Report	Data preprocessing pipeline and implementation of three ML algorithms(Logistic Regression, Decision Tree Classifier, and SVM)	1 Dec 2025
Final Report	Apply The improvements, and Comparison between our work and the work that mentioned in Introduction, and Comparison between The algorithms in our work(Comparison will be based on the results of performance metrics)	18 Dec 2025

[2]HaggingFace website,https://huggingface.co/datasets/aai510-group1/telco-customer-churn