

Machine Learning

Predicting

Customer Churn





Team Members

Elbethel Shineda

Leul Mintesinot

Mihret Agegnehu

Mikiyas Bedasa

Natanim Ashenafi

ETS0406/

ETS0766/13

ETS0874/13

ETS0888/13

ETS0979/13

Content

- 1. Introduction**
- 2. Data Description**
- 3. Methodology**
- 4. Implementation**
- 5. Usage**
- 6. Conclusion**

Introduction

1.1 Background

Customer churn is a significant challenge for businesses, especially in subscription-based industries. It refers to the rate at which customers stop doing business with an entity. Understanding and predicting customer churn can help companies implement strategies to retain customers and reduce revenue loss.

Problem Statement

The goal of this project is to develop a machine learning model that accurately predicts customer churn based on various features such as customer demographics, service usage patterns, and account information. Accurate predictions will enable the company to proactively engage at-risk customers and reduce churn rates.



Objectives

04

To evaluate the models using appropriate metrics and select the best performing model.

01

To collect and preprocess a comprehensive dataset on customer churn.

02

To perform exploratory data analysis (EDA) to identify key features influencing churn.

03

To build and compare several machine learning models for churn prediction.

Data Description



Data Source

The dataset used in this project is sourced from the Telco Customer Churn dataset available on Kaggle. It contains information about a telecom company's customers and their account details.



Data Structure

The dataset comprises 7043 rows and 21 columns, including features such as gender, SeniorCitizen, Partner, Dependents, tenure, PhoneService, TotalCharges ..., and the target variable Churn.



EDA

Churn rate: Approximately 26.5% of customers in the dataset churned.
Visualizations: Histograms, bar plots, and heatmaps were used to visualize the distribution of features and their relationships with churn.



Data analytics

Placeholder text: Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque facilisis tincidunt risus sit amet sagittis. Praesent tempus vitae felis maximus pretium. Mauris quam dui, scelerisque.

Methodology

Model Selection

Random Forest, Gradient Boosting, Support Vector Machine', SVC, Logistic Regression, K-Nearest Neighbors, Decision Tree, Ada Boost, XG Boost, Naive Bayes'

Hyperparameter Tuning

- Logistic Regression: C parameter tuning.
- Decision Trees: Max depth and min samples split.
- Random Forest: Number of estimators and max features.
- Gradient Boosting: Learning rate and number of estimators.
- SVM: C and gamma parameters.

Implementation

Tools and Libraries

- pandas for data manipulation.
- numpy for numerical computations.
- matplotlib and seaborn for data visualization.
- scikit-learn for machine learning models and evaluation.

Model Evaluation

- Accuracy: Proportion of correctly classified instances.
- Precision: Proportion of positive predictions that are correct.
- Recall: Proportion of actual positives that are correctly identified.
- F1-score: Harmonic mean of precision and recall.

Usage

Installation

- clone the repository: `git clone https://github.com/mihretgold/chrun_prediction_group2.git`
- Navigate to the project directory: `cd churn-prediction`

Running the Code

- Data Preprocessing: `python Machine_learning_group2.py`

Conclusion

Summary

This project successfully developed a machine learning model to predict customer churn with an accuracy of 98.2%. The K-Neighbors Classifier model was identified as the best performing model. Key features influencing churn were identified, providing actionable insights for customer retention strategies.