

Customer Churn Prediction Documentation

Table of Contents

- Introduction
- Data Preprocessing and Cleaning
- Feature Engineering
- Model Selection and Optimization
- Model Deployment and Integration
- Code Clarity and Organization
- Documentation and Reporting

1. Introduction

This documentation outlines the process and methodology used for the Customer Churn Prediction project. The objective of the project is to develop a machine learning model that predicts customer churn based on historical customer data. The project follows a typical machine learning project pipeline, encompassing data preprocessing, feature engineering, model selection, optimization.

2. Data Preprocessing and Cleaning

In this phase, the collected historical customer data was prepared for analysis and model training. Key steps included:

- Reading data from an Excel file using the `pd.read_excel()` function.
- Handling missing values: No missing values were found in the dataset.
- Handling duplicates: No duplicate rows were found in the dataset.
- Outlier analysis using boxplots to identify potential outliers.

3. Feature Engineering

Feature engineering aimed to enhance the predictive power of the model by creating new features:

- `Sub_More_ThanYear`: A boolean feature indicating whether the subscription length is more than a year.
- `Total_Money_Paid`: A feature calculated by multiplying subscription length and monthly bill.
- Categorical features were encoded using `LabelEncoder` to convert them into numerical values.

4. Model Selection and Optimization

- Multiple models are trained and evaluated on testing set but `SGDClassifier` has the highest `F1_score` and Precision. That's why i choose SGD Classifier for this Task
- I have done hyper parameter tuning by giving multiple parameters to the model and increases the the performance of the model
- `F1_score` obtained 0.67 and precision 1

5. Code Clarity and Organization

Code readability and organisation were maintained throughout the project:

- Descriptive variable names and meaningful comments were used.

- Code was organised into distinct sections for each step of the pipeline.
- Functions were defined for repeated tasks such as data reading, encoding

In conclusion, the Customer Churn Prediction project successfully developed a machine learning model to predict customer churn. The project demonstrated proficiency in data preprocessing, feature engineering, model selection, optimization, and documentation. The clear and organised codebase ensures maintainability and further development.