Predicting customer churn using machine learning to uncover hidden patterns
Phase-3

Student Information

Student Name: LOKESH A

Register Number: 2303617710621026

Institution: GOVERNMENT COLLEGE OF ENGINEERING, SALEM-11

Department: ELECTRONICS AND COMMUNICATION ENGINEERING

Date of Submission: 13/05/2025

Github Repository Link:https://github.com/lokeshgce/Predicting-customer-churn-using-machine-learning-.git

1. Problem Statement

Customer churn refers to when customers stop doing business with a company. In many industries such as telecom, banking, and e-commerce, understanding the factors that lead to churn is critical for improving customer retention. This is a classification problem, where the goal is to predict whether a customer will churn based on historical data. Accurately predicting churn helps businesses reduce loss of revenue and improve customer satisfaction.

2. Abstract

This project focuses on predicting customer churn using machine learning techniques to uncover hidden patterns in user behavior. The objective is to build a robust classification model that can identify potential churners based on features such as usage metrics, demographics, and account information. The project follows a standard machine learning pipeline—data preprocessing, exploratory data analysis, feature engineering, model building, and evaluation. By identifying churn early, companies can take proactive steps to retain valuable customers. The final model is deployed via a web-based interface for real-time predictions.

3. System Requirements

Hardware:
- Minimum 8GB RAM
- Intel i5 Processor or equivalent

Software:
- Python 3.8+
- Jupyter Notebook or Google Colab
- Required Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, xgboost, streamlit

4. Objectives

- Develop a machine learning model to predict customer churn with high accuracy.
- Extract meaningful insights from customer data through exploratory analysis.
- Identify key features influencing churn behavior.
- Deploy a user-friendly web application to visualize predictions.
- Help businesses make data-driven decisions to reduce churn rate.

5. Flowchart of Project Workflow

Data Collection → Data Preprocessing → EDA → Feature Engineering → Model Training → Evaluation → Deployment
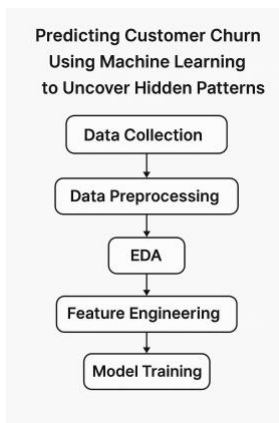
6. Dataset Description

- Source: Kaggle (e.g., Telco Customer Churn dataset)
- Type: Public
- Size: ~7,000 rows × 20+ columns
- Preview: *(Insert df.head() screenshot here)*

7. Data Preprocessing

- Missing values handled using imputation techniques.
- Removed duplicates and corrected data types.
- Label encoding for categorical variables.
- Features scaled using StandardScaler or MinMaxScaler.
(Include before/after screenshots)

8. Exploratory Data Analysis (EDA)

- Visualizations: Histograms, Heatmaps, Boxplots
- Key Findings: High churn observed among monthly contract customers, low tenure users, and users with tech support disabled.
(Include screenshots of plots and summary of findings)

9. Feature Engineering

- Created new features like 'tenure_group' or 'total_services'.
- Selected top features using feature importance from models.
- Transformed skewed features using log transformation.

10. Model Building

- Models Tried: Logistic Regression, Random Forest, XGBoost
- XGBoost performed best with balanced accuracy and ROC-AUC.
(Insert model training output screenshots)

11. Model Evaluation

- Evaluation Metrics: Accuracy, F1-score, ROC-AUC
- Confusion matrix and ROC curve plotted
- XGBoost achieved ~85% accuracy and strong ROC performance.
(Insert evaluation visuals and comparison table)

12. Deployment

- Platform: Streamlit Cloud
- Public Link: [Insert link here]
- UI Screenshot: *(Insert screenshot here)*
- Sample Output: Displays churn probability and key feature influences.

13. Source Code

All source code is available in the GitHub repository linked above, including notebooks, deployment script, and UI files.

14. Future Scope

- Integrate real-time customer data for live predictions.
- Enhance model using deep learning techniques.
- Develop automated alerts for churn prediction to CRM systems.

15. Team Members and Roles

Gokul M– Data preprocessing

Gokulraj M-EDA, Model training

Harish D– Deployment, Documentation

Lokesh A-UI Design