



CUSTOMER CHURN PREDICTION

MACHINE LEARNING

Introduction

Customer churn means when a customer stops using a company's product or service.

Predicting churn helps businesses take action to retain valuable customers.

The goal of this project is to predict which customers are likely to leave using Machine Learning techniques.

01

NumPy , Pandas

For numerical computations and data manipulation

02

Matplot and seaborn

For Visualization

03

Scikit-learn

For building and evaluating Machine Learning models.

Data Preprocessing

- Handled missing values and removed duplicate records.
- Converted categorical features using encoding (label encoding or one-hot encoding).
- Scaled numerical features to improve model performance.
- Split the dataset into training and testing sets for model building.

EDA

- Exploratory Data Analysis (EDA) Checked the distribution of the target variable to see the proportion of churners and non-churners.
- Analyzed correlations between features to identify important predictors.
- Visualized numerical and categorical variables using graphs to understand trends and patterns.
- Observed that contract type, tenure, and payment method are strong indicators of churn.

Model Building

Logistic Regression

Random Forest

Xgboost

ACCURACY

74%

logistic
regression

84%

Random
Forest

80%

XG
boost

Evaluation of all 3 models

	Precision	Recall	F! Score
Logistic	0.50	0.73	0.60
Random Forest	0.66	0.81	0.73
Xgboost	0.67	0.53	0.58

CONCLUSION

Logistic Regression had good recall but lower precision, catching many churners with more false alarms.

Random Forest was the most reliable overall with high accuracy and balanced precision & recall; XGBoost had good precision but missed some churners.

