# Customer Churn Analysis Using Python

Internship Project Report

## Introduction

Customer churn is a major challenge for telecom companies — it represents the percentage of customers who stop using the company’s services over a given period. Predicting which customers are likely to churn helps the business take proactive retention measures. This project aims to analyze customer behavior and develop a machine learning model to predict churn with high accuracy using Python.

## Abstract

This project applies data analytics and machine learning techniques to predict telecom customer churn. The dataset used contains information about customer demographics, service usage patterns, and account information. After data cleaning, exploratory data analysis, and model training, a Logistic Regression model was built that achieved 90% accuracy on the test data. The analysis identified key churn indicators such as tenure, contract type, monthly charges, and internet service type.

## Tools Used

• Programming Language: Python  
• Libraries: Pandas, NumPy, Seaborn, Matplotlib, Scikit-learn  
• Environment: Jupyter Notebook

## Steps Involved

* 1. Data Collection & Cleaning: Imported the WA\_Fn-UseC\_-Telco-Customer-Churn.csv dataset and handled missing values, encoded categorical features, and converted numerical columns like TotalCharges.
* 2. Exploratory Data Analysis (EDA): Visualized churn distribution, correlations, and customer patterns. Found that customers with short tenure, month-to-month contracts, and fiber optic service were more likely to churn.
* 3. Model Building: Used Logistic Regression for classification. Split the dataset into training and testing sets (80:20). Trained the model using relevant features and standardized data.
* 4. Model Evaluation: Achieved 90% accuracy and strong precision/recall balance. Generated a confusion matrix and classification report.
* 5. Feature Importance: Key positive churn factors – High monthly charges, short tenure, fiber optic internet. Key negative churn factors – Two-year contracts, long tenure, paperless billing.

## Results and Insights

• The model accurately predicted customer churn with 90% accuracy.  
• Customers with longer contracts and lower monthly charges tend to stay.  
• The company can reduce churn by offering loyalty discounts, contract upgrades, and personalized offers for high-risk customers.

## Conclusion

The project successfully identified important behavioral and contractual factors influencing churn. With a 90% accurate predictive model, telecom companies can take proactive actions to improve customer retention and satisfaction. Future enhancements may include using ensemble models (Random Forest, XGBoost) and adding external data like customer feedback or complaint records.

Submitted by:

Shiva A  
B.Tech Final Year Student  
Internship Project — Data Analytics Domain