**PROJECT REPORT**

**CUSTOMER CHURN PREDICTION USING MACHINE LEARNING**

SUBMITTED BY: KARAN KULRAJ SINGH

SUBMITTED ON: 06/09/2025

***ABSTRACT***

Customer churn, or customer attrition, is when customers stop using a company’s services. Predicting churn is crucial for businesses (especially telecom, SaaS, and banking) to identify customers at risk and take proactive measures to retain them.

In this project, I have built a machine learning model to predict customer churn using historical data. The model not only predicts whether a customer will churn but also gives the probability of churn. Additionally, the project includes a Streamlit web app where users can input customer details, get predictions, and view actionable business recommendations for retention.

***INTRODUCTION***

Problem Statement: High customer churn increases revenue loss and acquisition costs. Businesses need predictive tools to identify at-risk customers.

Objective:

* Train a machine learning model to predict customer churn.
* Provide business recommendations to reduce churn.
* Build a user-friendly Streamlit app for live predictions.

***TOOLS USED***

Programming Language: Python

Libraries:

* pandas, numpy → Data cleaning & analysis
* matplotlib, seaborn → Visualization
* scikit-learn → Machine Learning (training, evaluation, preprocessing)
* pickle → Saving and loading models
* streamlit → Deploying the app as an interactive web tool

Dataset: Telco Customer Churn dataset (with features like tenure, monthly charges, contract type, etc.)

***STEPS INVOLVED***

(a) Data Preprocessing:

* Loaded dataset into pandas DataFrame.
* Handled missing values (e.g., empty “TotalCharges”).
* Converted categorical variables to numeric using Label Encoding.
* Scaled numeric features (tenure, MonthlyCharges, TotalCharges) using StandardScaler.

(b) Exploratory Data Analysis (EDA)

* Visualized churn vs. non-churn distribution.
* Checked correlation between features (e.g., contract type strongly affects churn).
* Plotted revenue distribution across churned customers.

(c) Model Building

* Split data into training (80%) and testing (20%).
* Trained multiple ML models (Logistic Regression, Random Forest, etc.).
* Selected the best-performing model (Random Forest in our case).
* Saved the trained model using pickle.

(d) Model Evaluation

* Evaluated performance using accuracy, precision, recall, and F1-score.
* Generated confusion matrix to compare predicted vs. actual churn.

(e) Deployment (Streamlit App)

* Built a form where users can enter customer details (contract type, charges, etc.).
* App predicts churn probability and label (Yes/No).
* Added business recommendations (e.g., discounts, free tech support, security bundles).
* Displayed feature importance chart to understand which factors drive churn.

***BUSINESS RECOMMENDATIONS***

Based on model insights, businesses can reduce churn by:

* Offering discounted yearly contracts for month-to-month customers.
* Providing loyalty discounts for high monthly charges customers.
* Giving free trials of tech support and online security bundles.
* Engaging stable customers through rewards and engagement activities.

***CONCLUSION***

This project successfully demonstrates the end-to-end lifecycle of a data science solution:

* Understanding the business problem (churn).
* Cleaning and preprocessing data.
* Training a predictive ML model.
* Deploying results via a user-friendly Streamlit app.
* Translating predictions into actionable business recommendations.

Such a solution can help businesses improve customer retention, reduce revenue loss, and increase long-term customer satisfaction.