Tech Saksham

Final case study

Name: Raksha U

Department: MCA (BCU)

Semester: Third

USN: P18BR23S126017

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**1. Title**

Predicting Customer Churn Using Logistic Regression

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**2. Objective:**

The objective of this project is to build a machine learning model using logistic regression to predict customer churn for a given dataset. By identifying customers who are likely to churn, businesses can take proactive measures to retain them and improve customer satisfaction.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**3. Problem Statement:**

Customer churn is a critical issue for businesses, as losing customers directly impacts revenue and growth. It is essential to identify patterns in customer behaviour that may indicate a higher likelihood of churn. The dataset contains customer information, and the task is to use this data to predict whether a customer will churn (Yes) or not (No).

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**4. Solution:**

The solution involves the following steps:

* Loading the Dataset: Importing and exploring the customer churn dataset.
* Data Pre-processing: Cleaning and pre-processing the data by handling missing values, encoding categorical variables, and addressing any class imbalance.
* Feature Splitting: Splitting the data into training and testing subsets to evaluate the model's performance.
* Model Training: Training a logistic regression model to predict customer churn.
* Evaluation: Evaluating the model using metrics such as accuracy, classification report, and confusion matrix to assess its performance.

**5. Code Implementation:**

# Import required libraries

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import classification\_report, confusion\_matrix, accuracy\_score

# Step 1: Load the dataset

file\_path = "customer\_churn.csv" # Update this to your file path

data = pd.read\_csv(file\_path)

# Step 2: Data Preprocessing

# Drop unnecessary columns

data = data.drop(['customerID'], axis=1)

# Convert 'TotalCharges' to numeric

data['TotalCharges'] = pd.to\_numeric(data['TotalCharges'], errors='coerce')

# Handle missing values in 'TotalCharges'

data['TotalCharges'] = data['TotalCharges'].fillna(data['TotalCharges'].median())

# Encode categorical variables

data = pd.get\_dummies(data, drop\_first=True)

# Check for imbalance in target variable

print("\nClass Distribution in Target Variable:")

print(data['Churn\_Yes'].value\_counts())

# Step 3: Splitting the dataset

X = data.drop('Churn\_Yes', axis=1)

y = data['Churn\_Yes']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42, stratify=y)

# Step 4: Model Training

model = LogisticRegression(max\_iter=1000)

model.fit(X\_train, y\_train)

# Step 5: Predictions and Evaluation

y\_pred = model.predict(X\_test)

# Evaluation Metrics

print("\nAccuracy:", accuracy\_score(y\_test, y\_pred))

print("\nClassification Report:\n", classification\_report(y\_test, y\_pred, zero\_division=1))

# Confusion Matrix

conf\_matrix = confusion\_matrix(y\_test, y\_pred)

sns.heatmap(conf\_matrix, annot=True, fmt='d', cmap='Blues')

plt.title("Confusion Matrix")

plt.xlabel("Predicted")

plt.ylabel("Actual")

plt.show()

**6. Output:**

