Assesment Report

on

"Classify Customer Churn:"

submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY DEGREE

SESSION 2024-25

in

Computer Science & Engineering (AI & ML)

By

Vishesh Agarwal (202401100400215)

Under the supervision of

"Abhishek Shukla"

KIET Group of Institutions, Ghaziabad

Affiliated to

Dr. A.P.J. Abdul Kalam Technical University, Lucknow

(Formerly UPTU)

April, 2025

INTRODUCTION

Churn prediction is a crucial task for telecom companies to retain customers. Understanding the factors that lead to churn helps businesses take preventive actions and reduce customer loss.

This project uses a public dataset containing customer service records. We'll train a classification model using logistic regression to predict whether a customer will churn based on their service usage, contract type, monthly charges, etc.

METHODOLOGY

1. Data Upload & Cleaning:
   - The dataset is uploaded via Google Colab.
   - Missing values are detected and handled.
   - Non-numeric values in the TotalCharges column are converted and filled with median values.

2. Feature Engineering:
   - Non-numeric (categorical) variables are encoded using LabelEncoder.
   - Irrelevant columns (like customerID) are removed.

3. Model Building:
   - Features are scaled using StandardScaler.
   - Data is split into training and testing sets (80-20 split).
   - Logistic Regression is used as the classification algorithm.

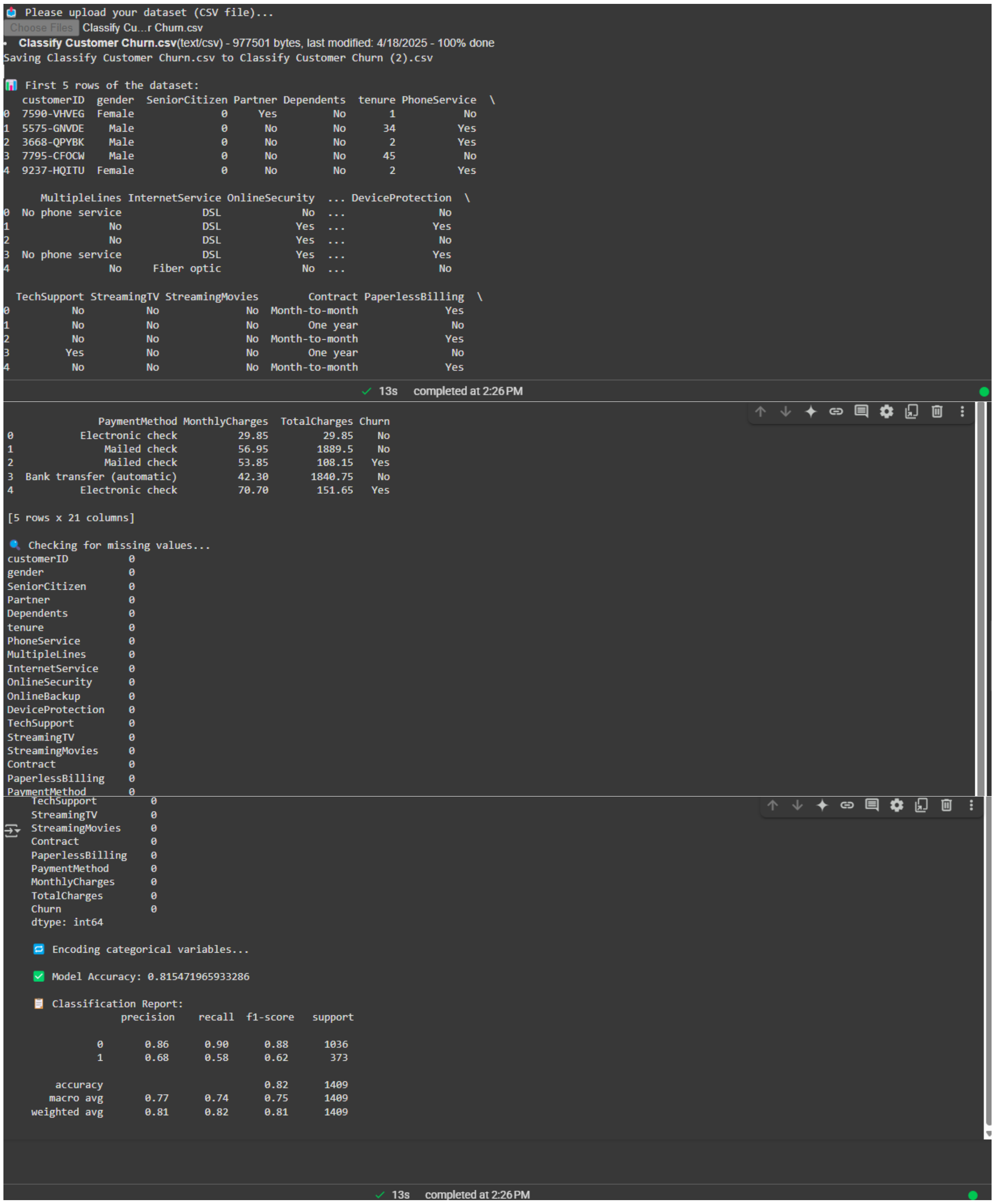
4. Evaluation:
   - Accuracy and a classification report are used to evaluate the model's performance.

CODE

```
# Customer Churn Classification in Google Colab (Improved Version)

# Step 1: Import necessary 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.preprocessing import LabelEncoder, StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, accuracy_score

from google.colab import files
from io import StringIO

# Step 2: Upload the dataset
print("📤 Please upload your dataset (CSV file)...")
uploaded = files.upload()

# Step 3: Load the dataset
file_name = list(uploaded.keys())[0]
df = pd.read_csv(StringIO(uploaded[file_name].decode('utf-8')))

# Step 4: Display the first few rows
print("\n📊 First 5 rows of the dataset:")
print(df.head())

# Step 5: Check for missing values
print("\n🔍 Checking for missing values...")
print(df.isnull().sum())

# Step 6: Drop unneeded columns (customerID is just an identifier)
if 'customerID' in df.columns:
   df.drop('customerID', axis=1, inplace=True)

# Step 7: Convert TotalCharges to numeric (some may be blanks or spaces)
df['TotalCharges'] = pd.to_numeric(df['TotalCharges'], errors='coerce')
df['TotalCharges'] = df['TotalCharges'].fillna(df['TotalCharges'].median())

# Step 8: Encode categorical variables
print("\n🔄 Encoding categorical variables...")
le = LabelEncoder()
```

```
for col in df.select_dtypes(include='object').columns:
    df[col] = le.fit_transform(df[col])

# Step 9: Define features and target
X = df.drop('Churn', axis=1)
y = df['Churn']

# Step 10: Scale features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Step 11: Split the data
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2,
random_state=42)

# Step 12: Train Logistic Regression model
model = LogisticRegression(max_iter=500, solver='lbfgs')
model.fit(X_train, y_train)

# Step 13: Predict and evaluate
y_pred = model.predict(X_test)

print("\n✅ Model Accuracy:", accuracy_score(y_test, y_pred))
print("\n📋 Classification Report:\n", classification_report(y_test, y_pred))
```

OUTPUT / RESULT



```
📤 Please upload your dataset (CSV file)...
Choose Files  Classify Cu...r Churn.csv
• Classify Customer Churn.csv(text/csv) - 977501 bytes, last modified: 4/18/2025 - 100% done
Saving Classify Customer Churn.csv to Classify Customer Churn (2).csv

📊 First 5 rows of the dataset:
   customerID  gender  SeniorCitizen Partner Dependents  tenure PhoneService  \
0  7590-VHVEG  Female              0     Yes         No       1           No
1  5575-GNVDE    Male              0      No         No      34          Yes
2  3668-QPYBK    Male              0      No         No       2          Yes
3  7795-CFOCW    Male              0      No         No      45           No
4  9237-HQITU  Female              0      No         No       2          Yes

      MultipleLines InternetService OnlineSecurity  ... DeviceProtection  \
0  No phone service             DSL             No  ...               No
1                No             DSL            Yes  ...              Yes
2                No             DSL            Yes  ...               No
3  No phone service             DSL            Yes  ...              Yes
4                No     Fiber optic             No  ...               No

  TechSupport StreamingTV StreamingMovies        Contract PaperlessBilling  \
0          No          No              No  Month-to-month              Yes
1          No          No              No        One year               No
2          No          No              No  Month-to-month              Yes
3         Yes          No              No        One year               No
4          No          No              No  Month-to-month              Yes
```

✓ 13s completed at 2:26 PM

```
              PaymentMethod MonthlyCharges  TotalCharges Churn
0          Electronic check          29.85         29.85    No
1              Mailed check          56.95        1889.5    No
2              Mailed check          53.85        108.15   Yes
3 Bank transfer (automatic)          42.30       1840.75    No
4          Electronic check          70.70        151.65   Yes

[5 rows x 21 columns]

🔍 Checking for missing values...
customerID          0
gender              0
SeniorCitizen       0
Partner             0
Dependents          0
tenure              0
PhoneService        0
MultipleLines       0
InternetService     0
OnlineSecurity      0
OnlineBackup        0
DeviceProtection    0
TechSupport         0
StreamingTV         0
StreamingMovies     0
Contract            0
PaperlessBilling    0
PaymentMethod       0
```

```
TechSupport         0
StreamingTV         0
StreamingMovies     0
Contract            0
PaperlessBilling    0
PaymentMethod       0
MonthlyCharges      0
TotalCharges        0
Churn               0
dtype: int64

🔁 Encoding categorical variables...

✅ Model Accuracy: 0.815471965933286

📋 Classification Report:
              precision    recall  f1-score   support

           0       0.86      0.90      0.88      1036
           1       0.68      0.58      0.62       373

    accuracy                           0.82      1409
   macro avg       0.77      0.74      0.75      1409
weighted avg       0.81      0.82      0.81      1409
```

✓ 13s completed at 2:26 PM

✅ Model Accuracy: 0.8133427963094393

References / Credits

- Dataset: https://www.kaggle.com/blastchar/telco-customer-churn
- Tools Used: Python, Google Colab, pandas, scikit-learn
- Libraries: pandas, numpy, seaborn, matplotlib, sklearn