RINEX PROJECT REPORT

***SELF INTRODUCTION:***

NAME: S. MUKUNTHAN

COLLEGE: PONDICHERRY TECHNOLOGICAL UNIVERSITY

YEAR: I ST YEAR

**1.MAJOR PROJECT – I**

Choose any dataset of your choice and apply a suitable  ML technique and if

possible deploy it using Heroku and Streamlit(COMPULSORY).

*GITHUB LINK:*

<https://github.com/Mukunthan14/RINEX-4-MAY-2022-DATA-SCIENCE-PROJECT-1.git>

*HEROKU APP LINK:*

<https://rinex4may.herokuapp.com/>

*GOOGLE DRIVE LINK:*

<https://drive.google.com/drive/folders/1_45knFpdOsYBIkQ_P9Zj2pbUFqoGQQcf?usp=sharing>

*SOURCE CODE:*

MODEL:

import pandas as pd

df = pd.read\_csv('https://raw.githubusercontent.com/Mukunthan14/RINEX-4-MAY-2022-DATA-SCIENCE/main/healthcare-dataset-stroke-data.csv')

df

df.info()

df.shape

import seaborn as sns

sns.swarmplot(x='stroke', y='avg\_glucose\_level', data=df) #Swarmplot for glucose level

#DATA CLEANING

df\_clean = df.drop(['id'], axis = 1)

df\_clean

df\_clean.isnull().sum()

df\_clean = df\_clean.dropna()

df\_clean.shape

df\_clean.info()

#ASSIGNING DUMMY VARIABLES

gender\_d = pd.get\_dummies(df\_clean['gender'],drop\_first=True)

smoking\_status\_d = pd.get\_dummies(df\_clean['smoking\_status'],drop\_first=True)

#CONCATENATE NEW CLOUMNS

df\_clean.drop(['gender', 'ever\_married', 'work\_type', 'Residence\_type', 'smoking\_status'],axis=1,inplace=True)

df\_clean = pd.concat([df\_clean,gender\_d,smoking\_status\_d],axis=1)

df\_clean

#ASSIGNING INPUT AND OUPUT

x = df\_clean.loc[:, df\_clean.columns != 'stroke'].values

x

y = df\_clean.loc[:, df\_clean.columns == 'stroke'].values

y

#TRAIN TEST SPLIT

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

x\_train,x\_test,y\_train,y\_test = train\_test\_split(x,y,random\_state = 0)

#Apply CLASSIFIER

from sklearn.linear\_model import LogisticRegression

model = LogisticRegression(solver='lbfgs', max\_iter=6000)

#model fitting

model.fit(x\_train,y\_train.ravel())

#predictor variable

y\_pred = model.predict(x\_test)

y\_pred # predicted outputs

y\_test # actual outputs

y\_pred2 = model.predict(x\_test[100:103]) #INDIVIDUAL PREDICTION

y\_pred2

y\_test[100:103]

#to check the accuracy

from sklearn.metrics import accuracy\_score

accuracy\_score(y\_pred,y\_test)\* 100

import joblib

joblib.dump(model, 'Logistic regression model')

APP.PY:

import streamlit as st

import joblib

import pandas

model = joblib.load('Logistic regression model')

st.title('LOGISTIC REGRESSION MODEL FOR STROKE')

age = st.number\_input("age", min\_value=1, max\_value=100)

hypertension = st.selectbox("hypertension",options=['yes', 'no'])

heart\_disease = st.selectbox("heart disease",options=['yes', 'no'])

avg\_glucose = st.number\_input("average glucose level")

bmi = st.number\_input("bmi")

sex = st.selectbox("Sex",options=['Male' , 'Female'])

smoking = st.selectbox("smoking status",options=['formerly smoked', 'never smoked', 'smokes'])

hypertension = 1 if hypertension == 'yes' else 0

heart\_disease = 1 if heart\_disease == 'yes' else 0

Male = 1 if sex == 'Male' else 0

Female = 1 if Male == 0 else 0

fs, ns, s = 0,0,0

if smoking == 'formerly smoked':

fs = 1

elif smoking == 'never smoked':

ns = 1

elif smoking == 'smokes':

s = 1

input\_data = [[age, hypertension, heart\_disease, avg\_glucose, bmi, Male, Female, fs, ns, s]]

pred = model.predict(input\_data)

if pred[0] == 1:

op = 'Person has suffered stroke'

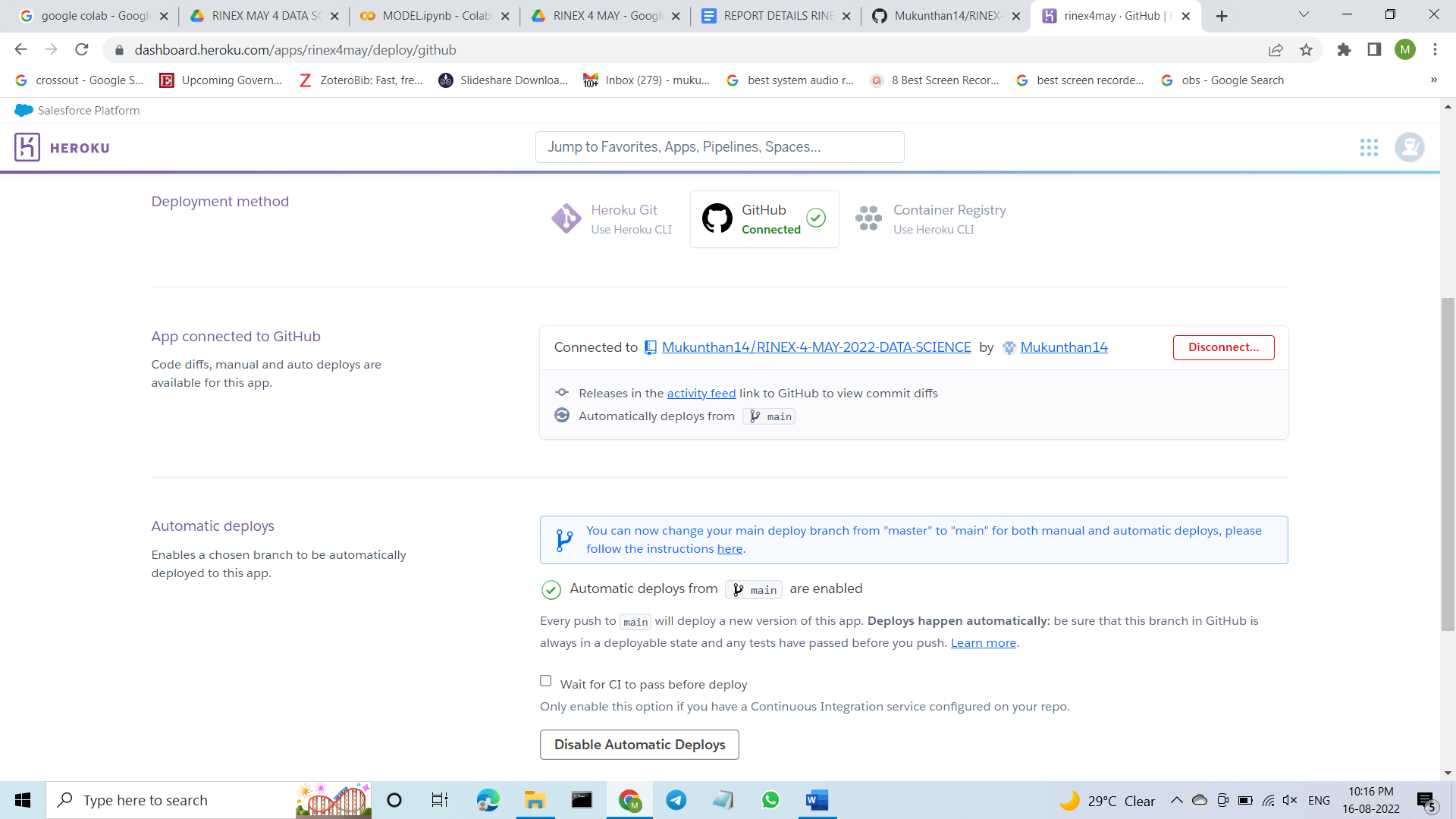
else:

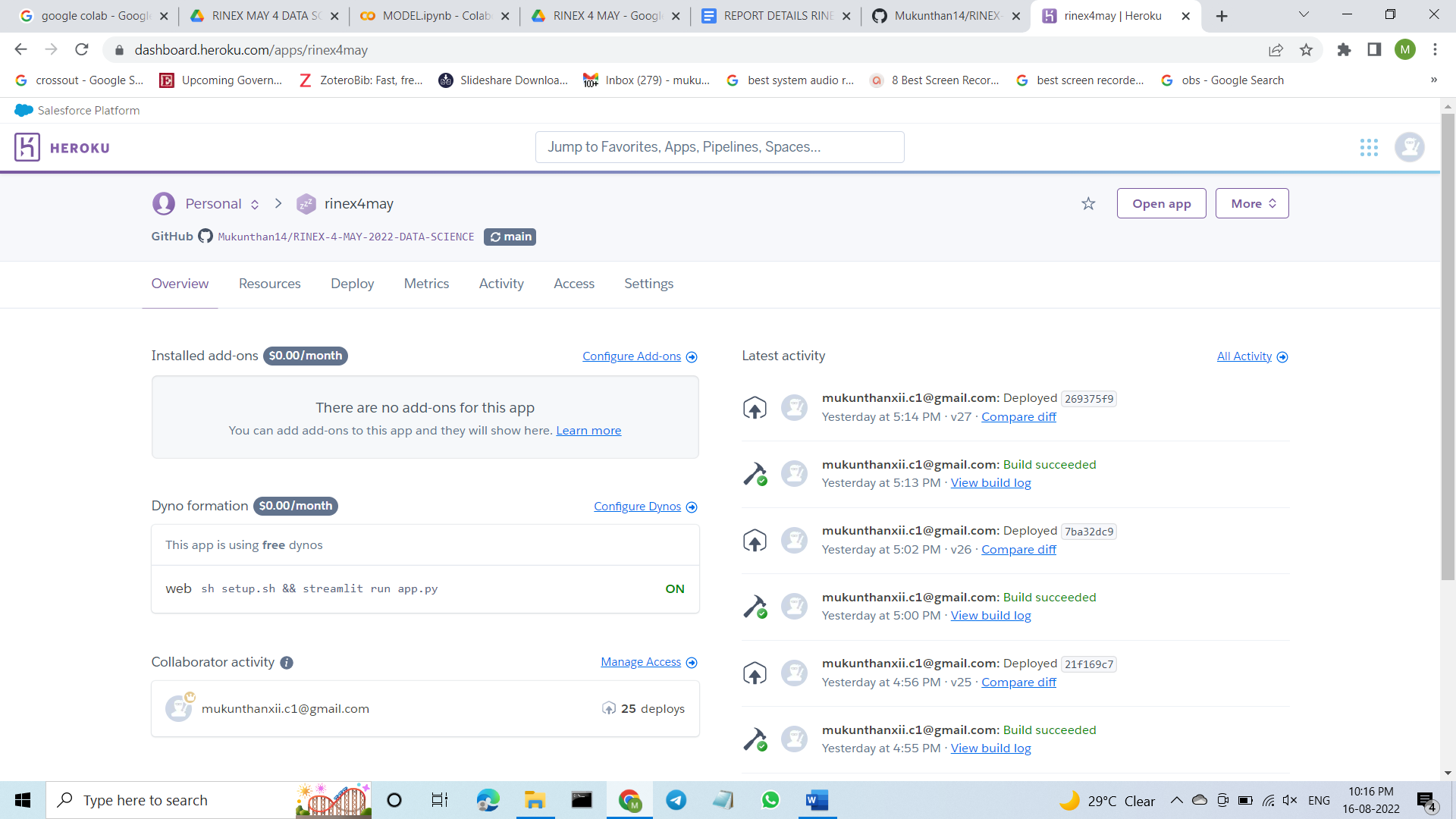
op = 'Person has not suffered stroke'

if st.button('Predict'):

st.subheader(op)

HEROKU SCREENSHOTS:





**2.MAJOR PROJECT – II:**

(a).Take any dataset  and perform Exploratory Data Analysis(EDA)

EDA should be done for atleast 15 different parameters/Analysis

*GITHUB LINK:*

<https://github.com/Mukunthan14/RINEX-4-MAY-PROJECT-2.git>

*GOOGLE DRIVE LINK:*

<https://drive.google.com/drive/folders/1_45knFpdOsYBIkQ_P9Zj2pbUFqoGQQcf?usp=sharing>

THANK YOU