WINE QUALITY PREDICTION USING IBM WATSON MACHINE LEARNING

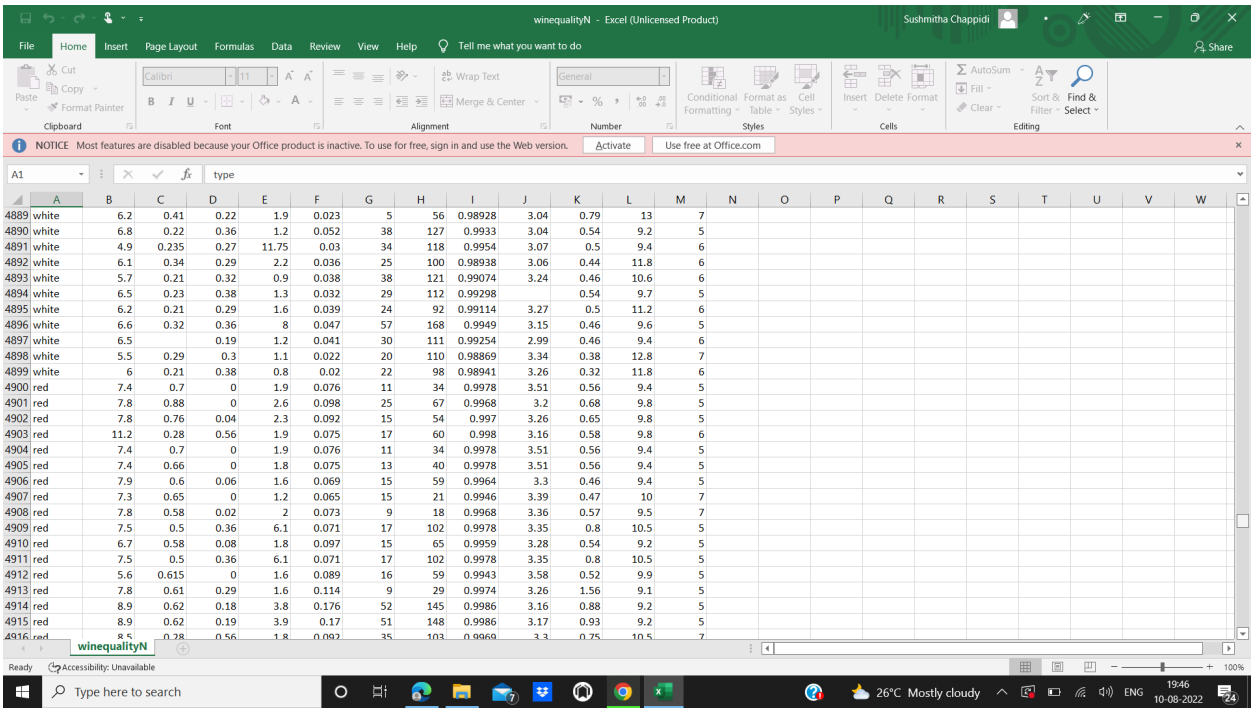
INTRODUCTION:

Project Overview:

Wine is the most commonly used beverage globally, and its values are considered important in society. Wine is an alcoholic drink that is made up of fermented grapes. Quality of wine is important for its consumers, mainly for producers in the present competitive market to raise the revenue. Wine quality refers to the factors that go into producing a wine, as well as the indicators or characteristics that tell you if the wine is of high quality. Historically, wine quality used to be determined by testing at the end of the production.

If you have come across wine then you will notice that wine has also their type, they are red and white wine. According to experts, wine is differentiated according to its smell, flavour, and colour, but we are not wine experts to say that wine is good or bad. Every person has their own opinion about the tastes, so identifying a quality based on a person's taste is challenging. Judging the quality of wine manually is a really tough task, even the professional wine tasters have the accuracy of 71%.

In this project, we present a wine quality prediction technique that utilizes historical data to train simple machine learning models which are more accurate and can help us know the quality of wine. The models can be run on much less resource intensive environments. From this the best model is selected and saved in pkl format. We will be doing flask integration and IBM deployment.



	A	B	C	D	E	F	G	H	I	J	K	L	M
4889	white	6.2	0.41	0.22	1.9	0.023	5	56	0.98928	3.04	0.79	13	7
4890	white	6.8	0.22	0.36	1.2	0.052	38	127	0.9933	3.04	0.54	9.2	5
4891	white	4.9	0.235	0.27	11.75	0.03	34	118	0.9954	3.07	0.5	9.4	6
4892	white	6.1	0.34	0.29	2.2	0.036	25	100	0.98938	3.06	0.44	11.8	6
4893	white	5.7	0.21	0.32	0.9	0.038	38	121	0.99074	3.24	0.46	10.6	6
4894	white	6.5	0.23	0.38	1.3	0.032	29	112	0.99298		0.54	9.7	5
4895	white	6.2	0.21	0.29	1.6	0.039	24	92	0.99114	3.27	0.5	11.2	6
4896	white	6.6	0.32	0.36	8	0.047	57	168	0.9949	3.15	0.46	9.6	5
4897	white	6.5		0.19	1.2	0.041	30	111	0.99254	2.99	0.46	9.4	6
4898	white	5.5	0.29	0.3	1.1	0.022	20	110	0.98869	3.34	0.38	12.8	7
4899	white	6	0.21	0.38	0.8	0.02	22	98	0.98941	3.26	0.32	11.8	6
4900	red	7.4	0.7	0	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	5
4901	red	7.8	0.88	0	2.6	0.098	25	67	0.9968	3.2	0.68	9.8	5
4902	red	7.8	0.76	0.04	2.3	0.092	15	54	0.997	3.26	0.65	9.8	5
4903	red	11.2	0.28	0.56	1.9	0.075	17	60	0.998	3.16	0.58	9.8	6
4904	red	7.4	0.7	0	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	5
4905	red	7.4	0.66	0	1.8	0.075	13	40	0.9978	3.51	0.56	9.4	5
4906	red	7.9	0.6	0.06	1.6	0.069	15	59	0.9964	3.3	0.46	9.4	5
4907	red	7.3	0.65	0	1.2	0.065	15	21	0.9946	3.39	0.47	10	7
4908	red	7.8	0.58	0.02	2	0.073	9	18	0.9968	3.36	0.57	9.5	7
4909	red	7.5	0.5	0.36	6.1	0.071	17	102	0.9978	3.35	0.8	10.5	5
4910	red	6.7	0.58	0.08	1.8	0.097	15	65	0.9959	3.28	0.54	9.2	5
4911	red	7.5	0.5	0.36	6.1	0.071	17	102	0.9978	3.35	0.8	10.5	5
4912	red	5.6	0.615	0	1.6	0.089	16	59	0.9943	3.58	0.52	9.9	5
4913	red	7.8	0.61	0.29	1.6	0.114	9	29	0.9974	3.26	1.56	9.1	5
4914	red	8.9	0.62	0.18	3.8	0.176	52	145	0.9986	3.16	0.88	9.2	5
4915	red	8.9	0.62	0.19	3.9	0.17	51	148	0.9986	3.17	0.93	9.2	5
4916	red	8.5	0.28	0.56	1.8	0.092	35	103	0.9969	3.3	0.75	10.5	7

as you can see in the data that there is red and white wine, and some other features. you can see that several features will be used to classify the quality of wine, many of them are chemical, so we need to have a basic understanding of such chemicals.

volatile acidity : Volatile acidity is the gaseous acids present in wine.

fixed acidity : Primary fixed acids found in wine are tartaric, succinic, citric, and malic

residual sugar : Amount of sugar left after fermentation.

citric acid : It is weak organic acid, found in citrus fruits naturally.

chlorides : Amount of salt present in wine.

free sulfur dioxide : So2 is used for prevention of wine by oxidation and microbial spoilage.

total sulfur dioxide pH : In wine pH is used for checking acidity density

sulphates : Added sulfites preserve freshness and protect wine from oxidation, and bacteria.

alcohol : Percent of alcohol present in wine. Rather than chemical features, you can see that there is one feature named Type it contains the types of wine we here discuss on red and white wine, the percent of red wine is greater than white.

purpose:

Here we will predict the quality of wine on the basis of giving features. We use the wine quality dataset from Kaggle. This dataset has the fundamental features which are responsible for affecting the quality of the wine. By the use of several Machine learning models, we will predict the quality of the wine. Here we will only

deal with the white type wine quality, we use classification techniques to check further the quality of the wine i.e. is it good or bed.

Data Pre-Processing

As we have understood how the data is. Lets pre-process the collected data.

The download data set is not suitable for training the machine learning model as it might have so much of randomness so we need to clean the dataset properly in order to fetch good results. This activity includes the following steps.

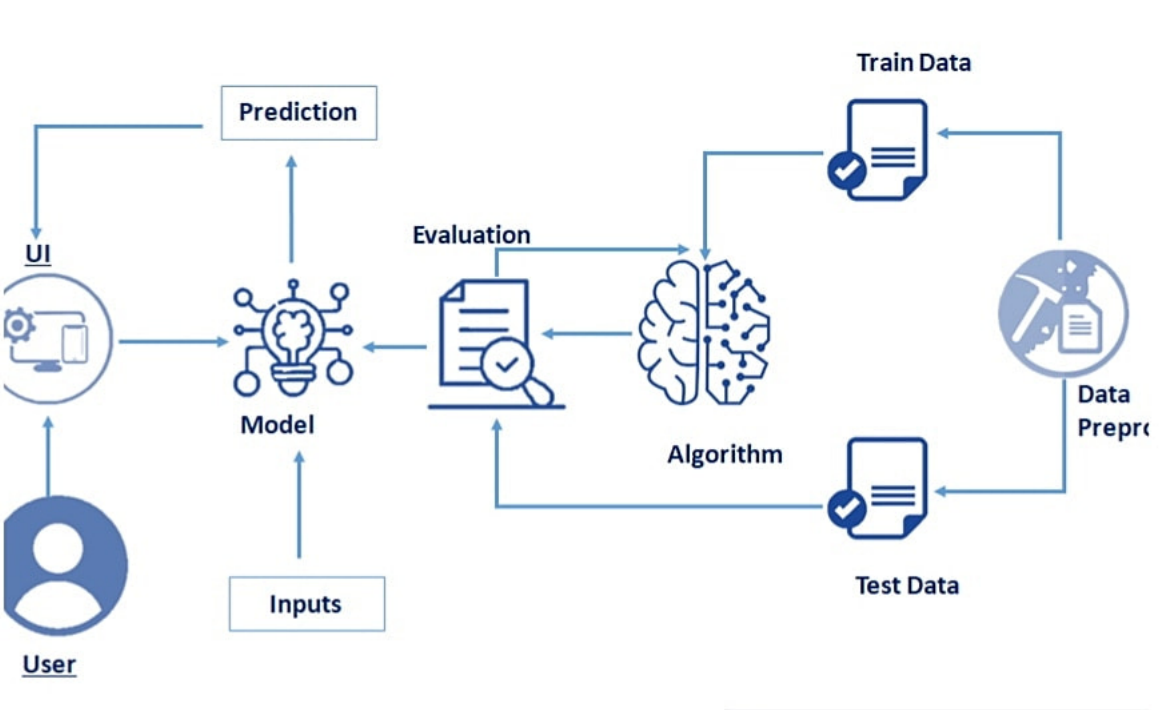
- **Removing unwanted columns**
- **Handling missing values**
- **Converting the target variable into binary class variable**
- **Handling categorical data**
- **Splitting dataset into training and test set**
- **Scaling Techniques**

Note: These are the general steps of pre-processing the data before using it for machine learning. Depending on the condition of your dataset, you may or may not have to go through all these steps.

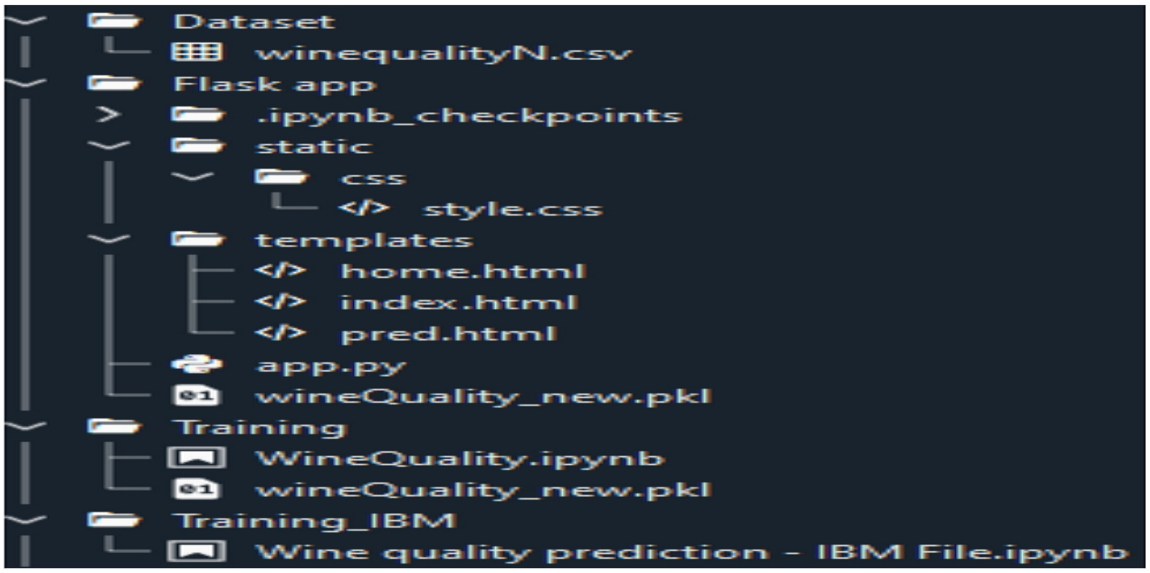
In the data frame, head() function is used to display the first 5 data. Our dataset has employee id (unique values), department (totally 9 dept.), region (location), education, gender, recruitment channel, age, no. of trainings, previous year ratings, length of service, KPIs, award won, average training score and is_promoted (target variable) columns.

Theoritical Analysis:

block diagram



Hardware and Software Designing



Experimental Investigations:

Multivariate Analysis

In simple words, multivariate analysis is to find the relation between multiple features. Here we have used a heatmap from the seaborn package.

- **Correlation is a statistical measure that expresses the extent to which two variables are linearly related. It's a common tool for describing simple relationships without making a statement about cause and effect.**
- **To visualize the correlation heat map() function is used. From the below image we can easily find the highly correlated feature. Abs() method is used to convert the negative correlation to positive correlation.**

- **From the above correlation plot for the given dataset for wine quality prediction, we can easily see which items are related strongly with each other and which items are related weekly with each other. For Example,**
- **The strongly correlated items are :**

  1.fixed acidity and citric acid. 2.free sulphur dioxide and total sulphor dioxide. 3.fixed acidity and density. 4. alcohol and quality.

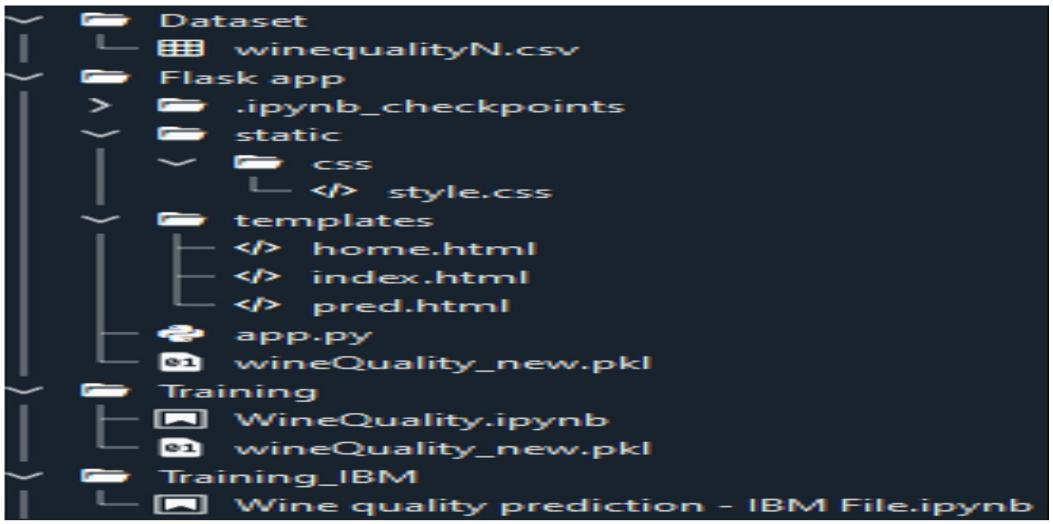
  so, from above points there is a clear inference that alcohol is the most important characteristic to determine the quality of wine.

- **The weakly correlated items are :**

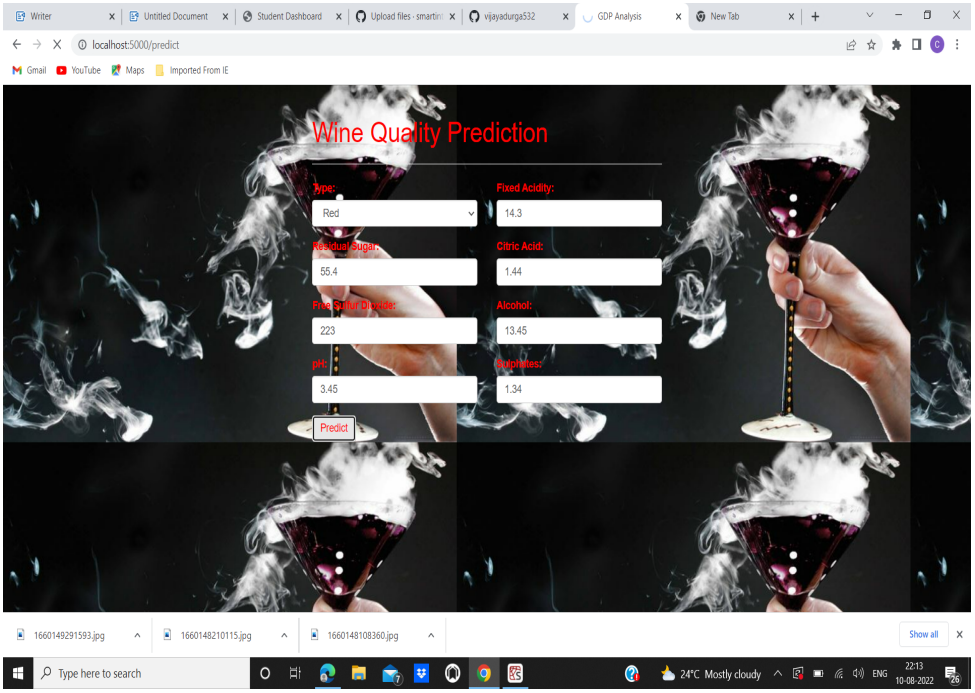
  1.citric acid and volatile acidity. 2.fixed acidity and ph. 3.density and alcohol.

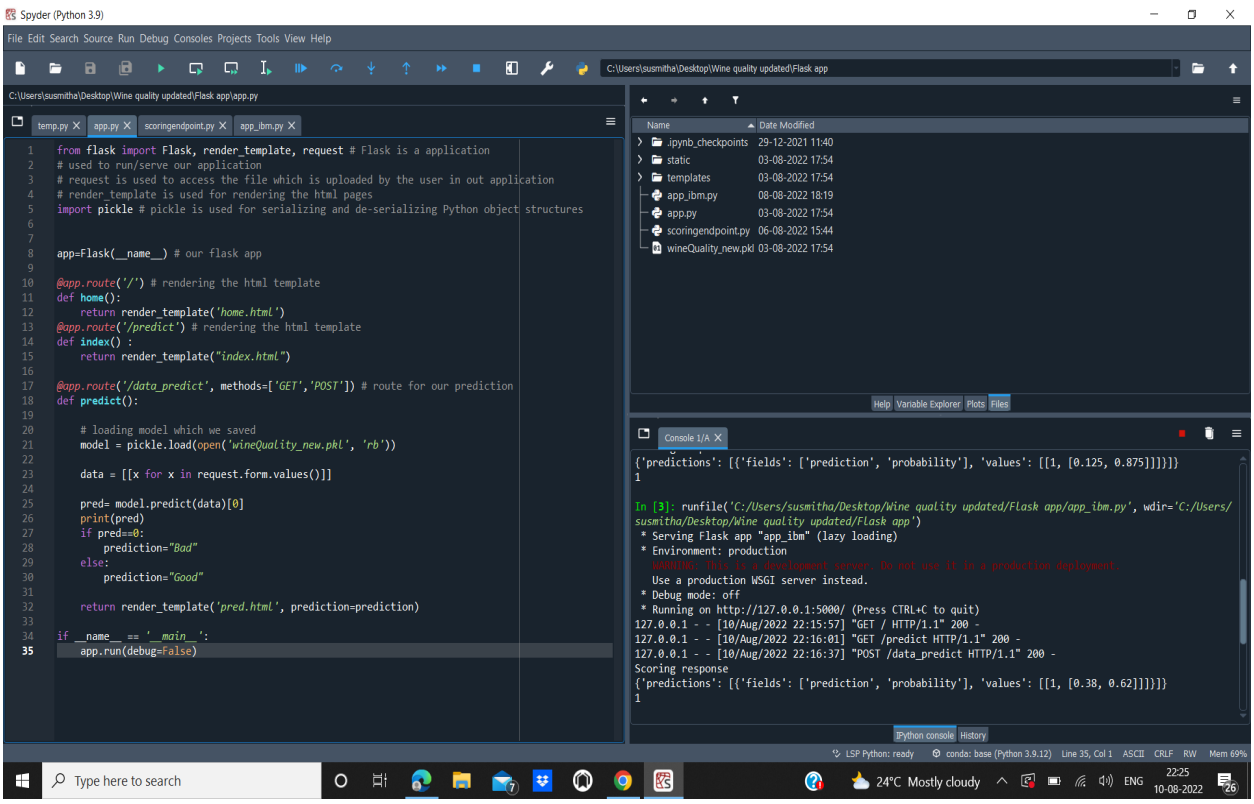
  These are some relations which do not depend on each other at all.

  Flowchart:inpiu


Dataset
winequalityN.csv
Flask app
.ipynb_checkpoints
static
css
style.css
templates
home.html
index.html
pred.html
app.py
wineQuality_new.pkl
Training
WineQuality.ipynb
wineQuality_new.pkl
Training_IBM
Wine quality prediction - IBM File.ipynb

INPUT:


localhost:5000/predict
Wine Quality Prediction
Type:
Red
Fixed Acidity:
14.3
Residual Sugar:
55.4
Citric Acid:
1.44
Free Sulfur Dioxide:
223
Alcohol:
13.45
pH:
3.45
Sulphates:
1.34
Predict
Type here to search
24°C Mostly cloudy
ENG
22:13
10-08-2022



OUTPUT:

Advantages and Disadvantages:

we can predict the quality of wine whether it is a good wine or bad wine
Application:
we can make use of this application in wine quality checking at wine industries.
Conclusion:
by making use of this application we can find out the wine quality by providing attributevalues.
Future Scpoe:
we can make the prediction easy and we may develop the application
Bibiliography:
I had followed the hackers realm youtube channel,google,and smartinternz mentors methods and their suggestions.
Appendix:
SOURCE CODES:
app.py

```
from flask import Flask, render_template, request # Flask is a application
# used to run/serve our application
# request is used to access the file which is uploaded by the user in out application
# render_template is used for rendering the html pages
import pickle # pickle is used for serializing and de-serializing Python object structures


app=Flask(__name__) # our flask app

@app.route('/') # rendering the html template
def home():
    return render_template('home.html')
@app.route('/predict') # rendering the html template
def index() :
    return render_template("index.html")

@app.route('/data_predict', methods=['GET','POST']) # route for our prediction
```

```
def predict():

    # loading model which we saved
    model = pickle.load(open('wineQuality_new.pkl',
  'rb'))

    data = [[x for x in request.form.values()]]

    pred= model.predict(data)[0]
    print(pred)
    if pred==0:
        prediction="Bad"
    else:
        prediction="Good"

    return render_template('pred.html',
  prediction=prediction)

if __name__ == '__main__':
    app.run(debug=False)
```

app_ibm.py

```
# -*- coding: utf-8 -*-
"""
Created on Mon Aug  8 18:01:19 2022

@author: susmitha
"""

from flask import Flask, render_template, request #
  Flask is a application
# used to run/serve our application
# request is used to access the file which is uploaded
  by the user in out application
# render_template is used for rendering the html pages
```

```
import pickle # pickle is used for serializing and de-serializing Python object structures
import requests

# NOTE: you must manually set API_KEY below using information retrieved from your IBM Cloud account.
API_KEY = "xhhhZHbB1Of7q9TEn62_PcljX0g4dXVCPdbPl3ox-q-G"
token_response = requests.post('https://iam.cloud.ibm.com/identity/token

 API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-

mltoken = token_response.json()["access_token"]

header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}

app=Flask(__name__) # our flask app

@app.route('/') # rendering the html template
def home():
    return render_template('home.html')
@app.route('/predict') # rendering the html template
def index() :
    return render_template("index.html")

@app.route('/data_predict', methods=['GET','POST']) # route for our prediction
def predict():

    # loading model which we saved
    #model = pickle.load(open('wineQuality_new.pkl', 'rb'))

```

```
    data = [[x for x in request.form.values()]]
    payload_scoring = {"input_data": [{"fields":
["f0","f1","f2","f3","f4","f5","f6","f7"],"values":
data}]}

    response_scoring = requests.post('https://us-
-43f3-

json=payload_scoring,
     headers={'Authorization': 'Bearer ' + mltoken})
    print("Scoring response")
    print(response_scoring.json())
    pred= response_scoring.json()
    output= pred['predictions'][0]['values'][0][0]
    print(output)
    #pred= model.predict(data)[0]
    #print(pred)
    if output==0:
        prediction="Bad"
    else:
        prediction="Good"

    return render_template('pred.html',
prediction=prediction)

if __name__ == '__main__':
    app.run(debug=False)
```

THANK YOU