**Fruit Category Prediction**

**By Using**

**Machine Learning**

**Project Report submitted in partial fulfillment of**

**The requirements for the degree of**

**Master of Computer Applications**

**Of**

**Maulana Abul Kalam Azad University of Technology**

**By**

**Satyajit Paine - 29101018022**

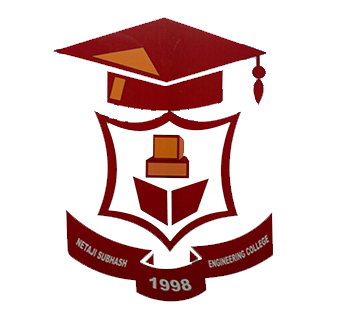
**Sayan Bera - 29101018021**

**Sourav Sahoo - 29101018019**

**Under the guidance of**

**Prof. Mr. Sudip Dey**

**DEPARTMENT OF M.C.A**

****

**NETAJI SUBHASH ENGINEERING COLLEGE**

**TECHNO CITY, GARIA, KOLKATA-700152**

**Academic Year of Pass out: 2020-2021**

**CERTIFICATE**

This is to certify that this project report titled

Fruit Category Prediction By Using Machine Learning

submitted in partial fulfillment of requirements for award of the degree Master of

Computer Application of Maulana Abul Kalam Azad University of Technology

is a faithful record

of the original work carried out by,

Satyajit Paine - 29101018022 - 182910510011

Sayan Bera - 29101018021 -182910510012

Sourav Sahoo - 29101018019 - 182910510014

Under my guidance and supervision.

It is further certified that it contains no material, which to a substantial extent has been submitted for the award of any degree/diploma in any institute or has been published in any form, except the assistances drawn from other sources, for which due acknowledgement has been made.

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

Date: ....................

Guide’s Signature

**Prof. Sudip Dey**

Sd/\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Head of the Department**

**Dr. Atanu Das**

**Netaji Subhash Engineering College**

**Techno City, Garia, Kolkata – 700152**

**DECLARATION**

We hereby declare that this project report titled

**Fruit Category Prediction By Using Machine Learning**

Is our own original work carried out as a under graduate student in Netaji Subhash

Engineering College except to the extent that assistances from other sources are duly

acknowledged.

All sources used for this project report have been fully and properly cited. It contains

no material which to a substantial extent has been submitted for the award of any

degree/diploma in any institute or has been published in any form, except where due

acknowledgement is made.

Students Name Signature Dates

…………………………………. ………………………………….. ……………………

…………………………………. ………………………………….. ……………………

…………………………………. ………………………………….. ……………………

…………………………………. ………………………………….. ……………………

|  |
| --- |
| **CERTIFICATE OF APPROVAL** |
| We here by approve this dissertation titled |
| **Fruit Category** **Prediction By Using Machine Learning** |
| Carried out by |
| Satyajit Paine, Roll no. 29101018022,  Sayan Bera, Roll no. 29101018021,  Sourav Sahoo, Roll no. 29101018019,  Under the guidance of |
| **Prof. Mr. Sudip Dey** |
| Of Netaji Subhash Engineering College, Kolkata in partial fulfillment of requirements | |
| for award of the Master of Computer Application of Maulana Abul  Kalam Azad University of Technology |
|  |
| **Date: ..........................** | |
| **Examiner’s Signatures:** | |
| **…………………………………………** |
| **…………………………………………** |
| **…………………………………………** |

**ACKNOWLEDGEMENTS**

The achievement that is associated with the successful Completion of any task would be incomplete without mentioning the names of those people whose endless cooperation made it possible. The continuous support driven, by our Faculty Coordinator **Mr. Sudip Dey** and Student Representative **Mr. Sayan Bera.** Their constant guidance and encouragement made all our efforts successful. We take this opportunity to express our deep gratitude towards our project mentor **Mr. Sudip Dey** for giving such valuable suggestions guidance and encouragement during the development of this project work. Last but not the least we are grateful to all the faculty members of **Engineering Study Centre** for their support.

Satyajit Paine

Sayan Bera

Sourav Sahoo

Dated: ………………..

TABLE OF CONTENTS

1. **Abstract**

**2. Introduction**

**3. Problem Definition**

**4. Project Goal**

**5. Software Required**

**6. Methodology**

**7. Project Objective**

**8. Project Workflow**

**9. Project Implementation**

**10. Step-by-step Working**

**11. Project Limitations**

**12. Future Scope**

**13. Summary**

**14. Bibliography**

**15. Code**

ABSTRACT

This report describes an approach of creating a system identifying fruit. The system helps the customers to label desired fruits according to its weight and mass or color etc. The purpose of the system is to minimize the number of human computer interactions, speed up the identification process and improve the usability of the graphical user interface compared to existing systems. To accomplish creating a system improving these properties, an idea of implementing machine learning to identify the products aroused. Instead of assigning the responsibility to the user, who usually identifies the products manually, the responsibility is given to a computer to classify an object; we use classification technique such as k-Nearest Neighbor algorithm in this project. Here we use the fruit dataset to predict fruits whether the fruit is ‘orange’ or ‘apple’ or other. We also check the model which has been proved useful in displaying complex system and showing the relationship between variables in a graphical way. In this machine learning model we can get accurate result after various testing.

INTRODUCTION

In the last few years, the most popular analysis techniques that have been used for classifications of fruit category data are based upon color and width or color and mass or color, height and shape analysis. However, classification of fruits is still a complicated task due to the various properties of numerous types of fruits. These analysis methods are still not robust and effective enough to identify and distinguish fruits. Out of 185 type of four different fruits collected for the system, (80%) data are used for training and (20%) for testing the system. The system uses the k-Nearest Neighbor (KNN) algorithm as classifier. The proposed method combines four features i.e. color, mass, width and height. The method classifies and predicts varieties of fruits using nearest neighbor’s classification. Results: Accuracy of the proposed system is 100%. The proposed system has the flexibility for many fruits in various colors, mass and height. The proposed method is robust, accurate and flexible that can process, analyze, classify and identify the varieties of fruits using their features such as color, height and mass. The KNN algorithm is used for fruits category prediction which accurately classifies fruits.

Python- Python is an interpreted, high-level and general-purpose programming language. The quantity of highlights in the actual language is modest, easy and requiring moderately little venture of time or exertion to create small programs and projects. The Python language is intended to be straightforward and clear. This straightforwardness makes Python an ideal language, and it allows novices to get and adopt it rapidly. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

[//Libraries]

Machine Learning (ML) - ML is a utilization of man-made consciousness (AI) that gives frameworks the capacity to consequently take in and improve as a matter of fact without being expressly customized. Machine learning focusses around the improvement of PC programs that can get to information and use it to find out on their own.

The way toward learning starts with perceptions or information, like models, direct insight, or guidance, to search for designs in information and settle on better choices later on dependent on the models that we give. The essential point is to permit the PCs adapt naturally without human intercession or help and change activities likewise.

Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist. In regression analysis, logistic regression is estimating the parameters of a logistic model.

PROBLEM DEFINITION

The main objective of this project is to predict the fruits whether it is “Orange” or “lemon” or other. Based on the dataset. Project uses KNN classification algorithm to predict whether fruit is Orange or lemon or apple. This integrated technique of classification gives a promising classification results with utmost accuracy rate. We use this types of classification when two or more than two class are present in the classifier.

PROJECT GOAL

The goal is to predict whether the fruit is ‘Orange’ or 'lemon' or 'apple'. The goal is achieved by using KNN algorithm in machine learning.

SOFTWARE REQUIRED

TOOLS USED

1.Anaconda Navigator-1.9.6

2.Jupyter Notebook-5.7.4

LIBRARY USED

1.Analyzing:Numpy,Pandas,Sci-kit Learn

2.Visualization:Matplotlib,Seaborn

METHODOLOGY

• Data Selection: Data is the establishment for any AI project. The work is to discover ways and sources of gathering significant furthermore, thorough information, deciphering it, and examining results with the assistance of measurable strategies

• Data Visualization:A large chunk of data addressed in graphical structure is more obvious and to analysis the data. A few organizations determine that an information should realize how to make slides, graphs, diagrams, and formats.

• Data cleaning: This arrangement of methodology takes into account of eliminating noise and fixing irregularities in data. A data researcher can fill in missing information utilizing attribution methods. An expert likewise distinguishes exceptions — perceptions that deviate significantly from the rest of distribution.

• Data Splitting: A dataset utilized for AI should be partitioned into three subsets — test, training and approval sets.

• Model Selection: After a data researcher has preprocessed the gathered information and part it into three subsets, now the individual can continue with a model preparing. This interaction involves "feeding" the calculation with the preparing information. Algorithm will handle information and generate a model that can discover an objective incentive in new information. The motivation behind model preparing is to build up a model.

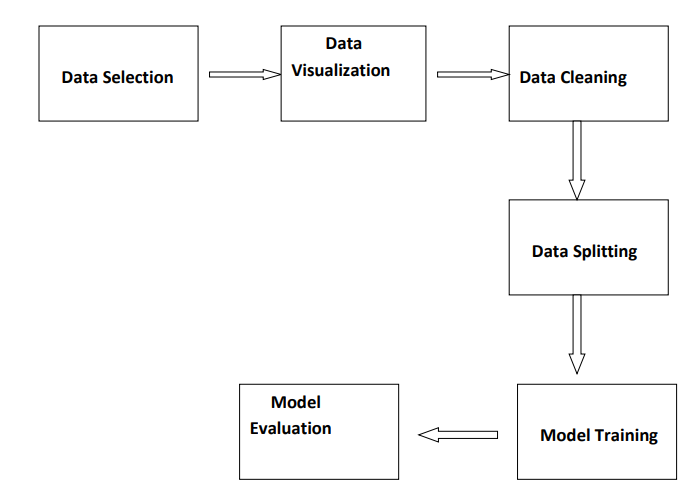
• Model Evaluation: The objective or the goal of this progression is to build up the simple model ready to detail an objective worth quick and all able to formulate a target value fast and well enough and check the accuracy.

PROJECT OBJECTIVE

The main objective of this project is to find The Fruit Category Prediction by using machine learning.

PROJECT WORKFLOW

This is the detailed work architecture where we are showing the process of Fruit Category Prediction with the help of KNN algorithm for finding accuracy.



PROJECT IMPLEMENTATION

• SELECTION OF DATA: The process of selecting data depends on the type of project we desire to. The data set can be collected from various sources such as a file, database, sensor and many other such sources.

• VISUALIZATION OF DATA: Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in.

• DATA PRE-PROCESSING: As we know that data pre-processing is a process of cleaning the raw data into clean data, so that can be used to train the model. So, we definitely need data pre-processing to achieve good results from the applied model in machine learning and deep learning projects.

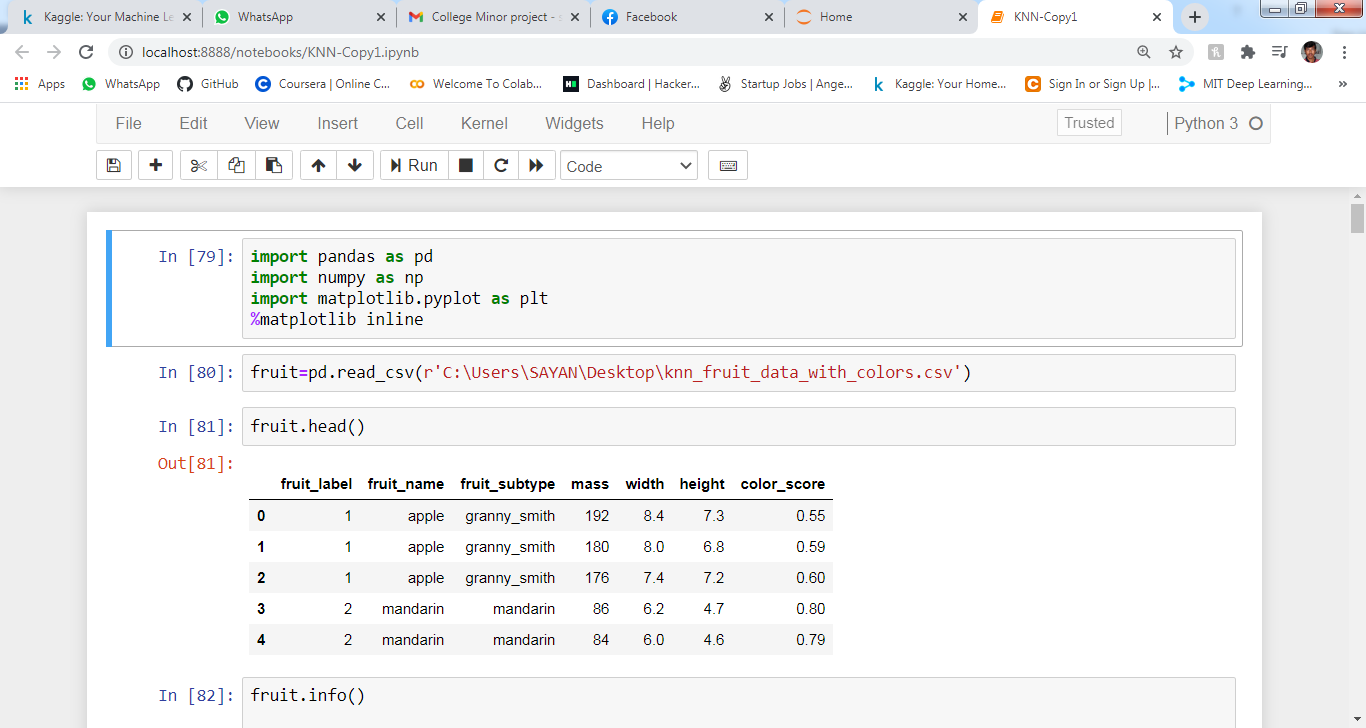
• SELECTION OF DEPENDENT AND INDEPENDENT DATA: We need to select the dependent and independent data and store them in y and x.

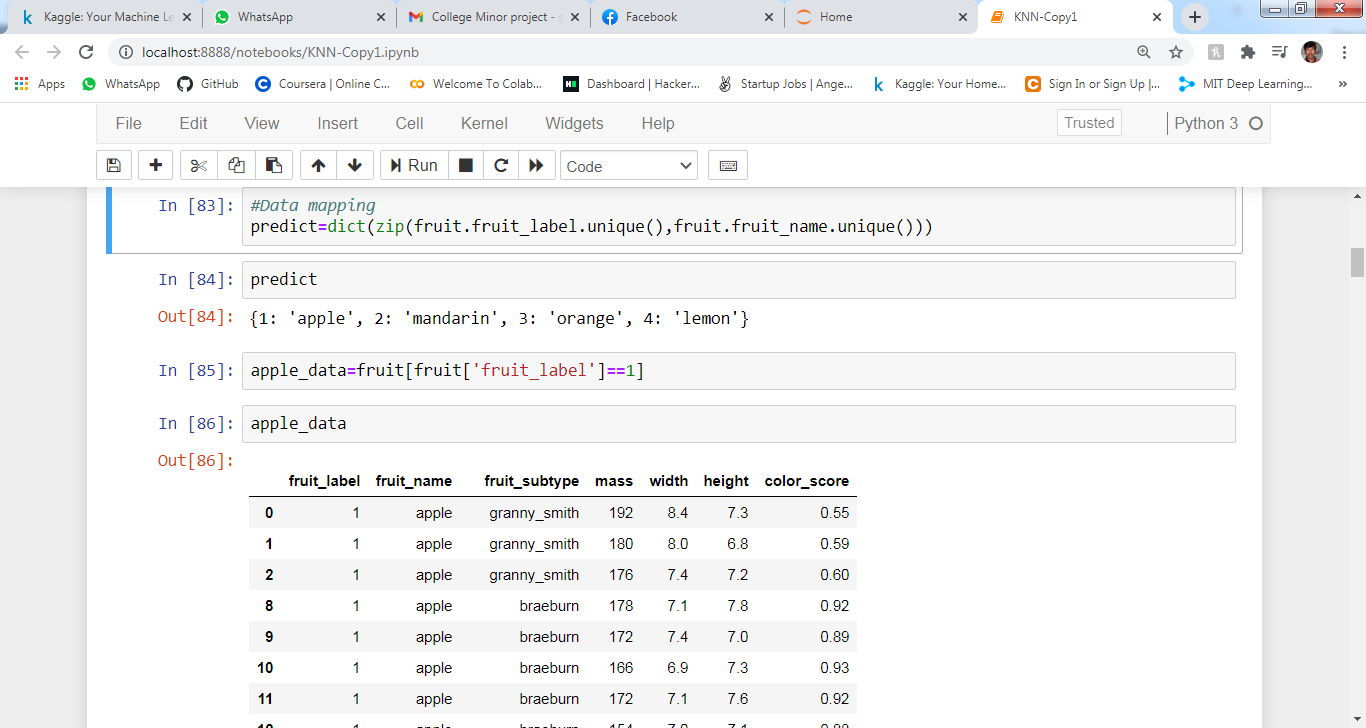
• SPLITTING OF THE DATA: We train the classifier using ‘titanic\_ data’, then we test the performance and accuracy.

• FITTING THE MODEL: In a data set, a training set is implemented to build up a model. Once the model is trained, we can use the same trained model to predict using the testing data i.e. the unseen data. Once this is done, we can develop a confusion matrix, this tells us how well our model is trained.

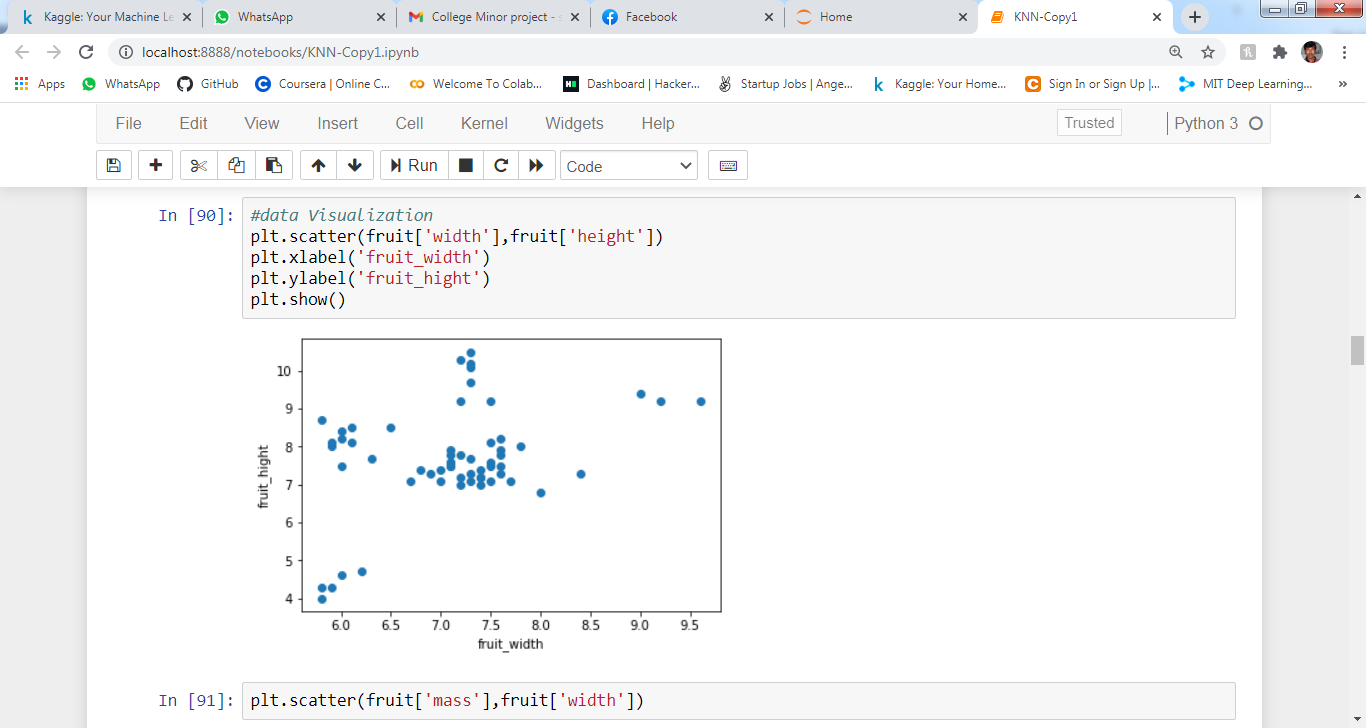
• MODEL EVALUATION: It is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future.

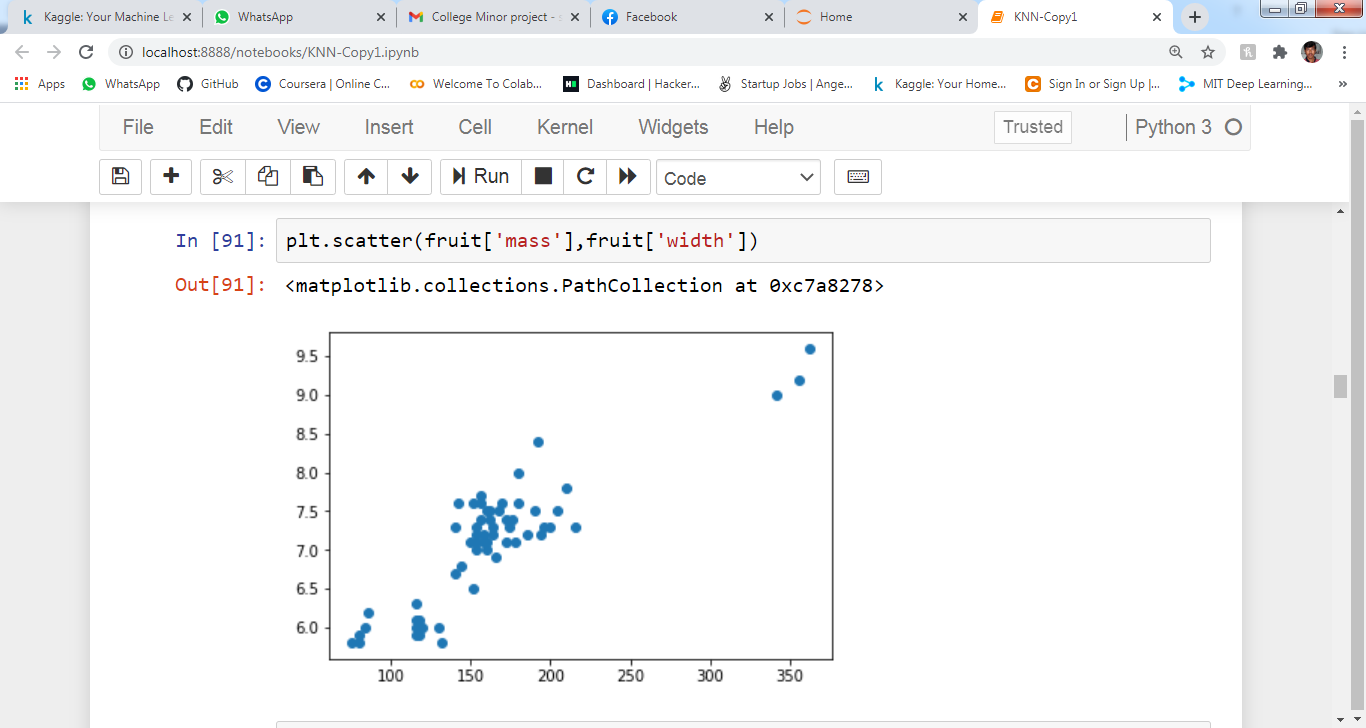
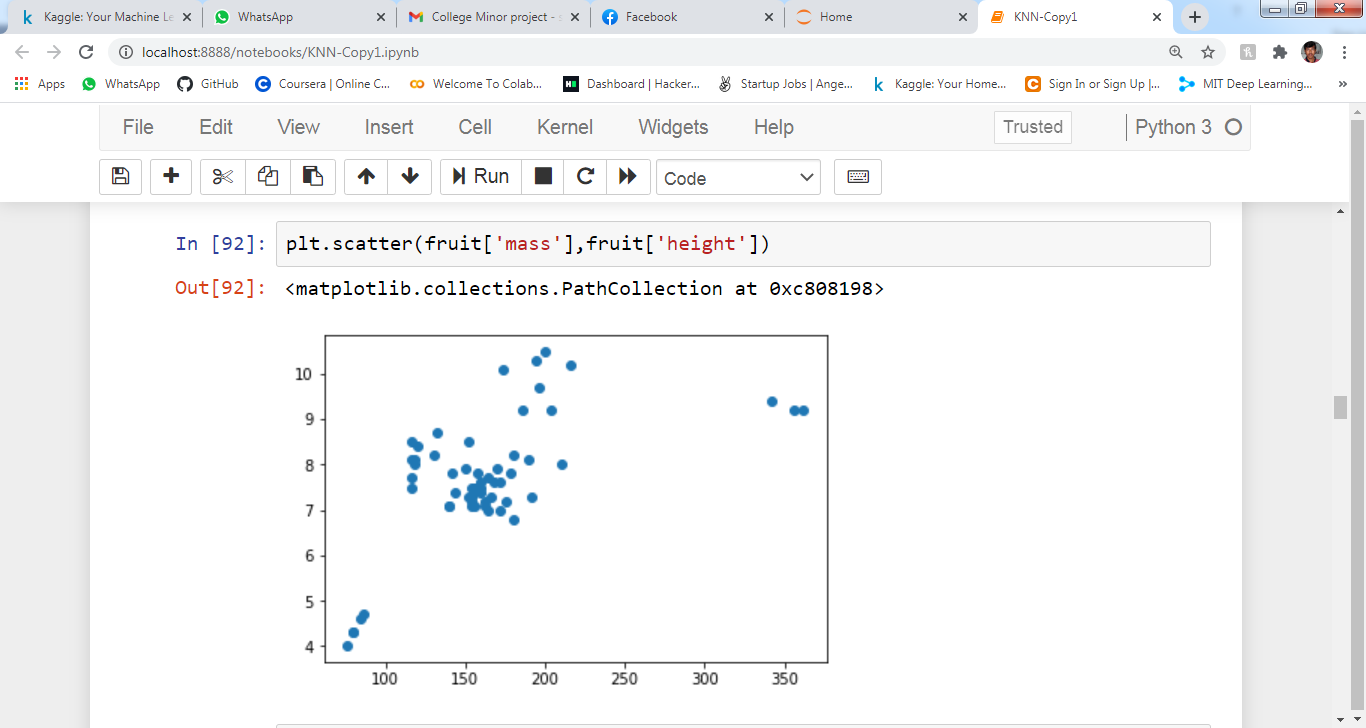
STEP BY STEP WORKING:

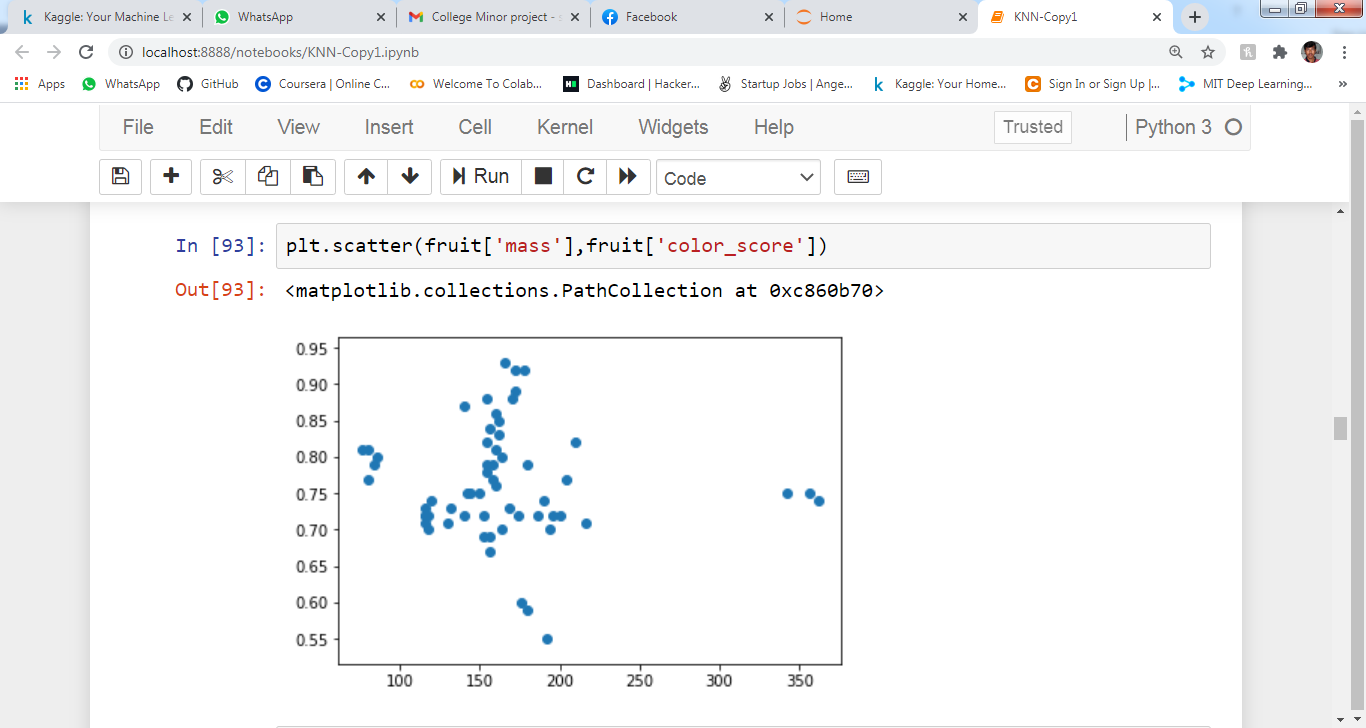
* SELECTION OF DATA: 

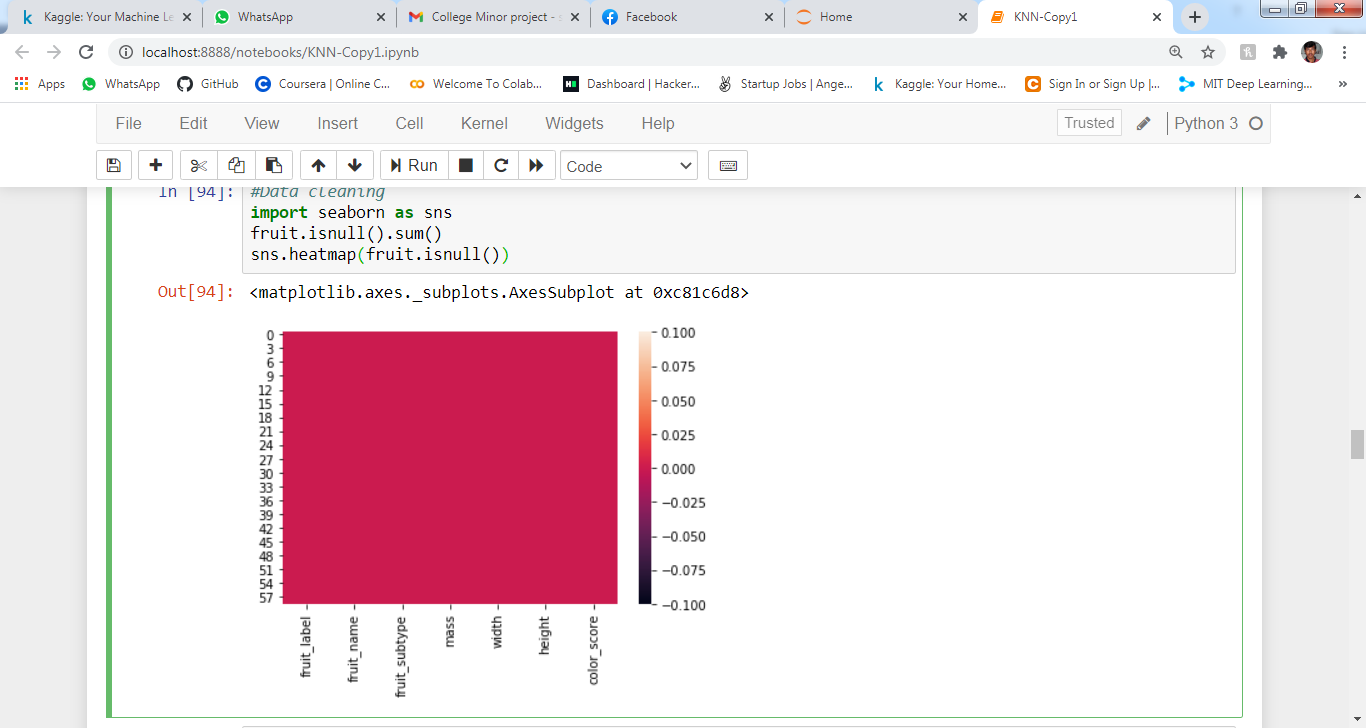
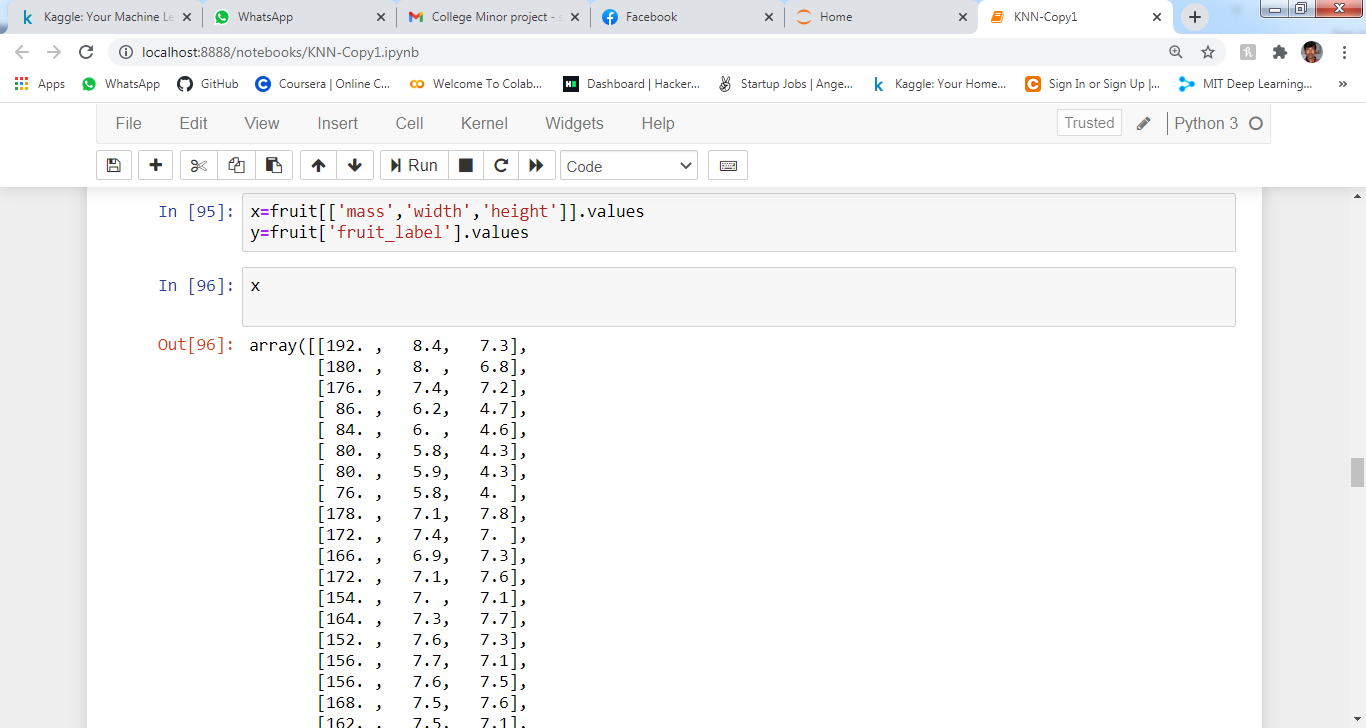


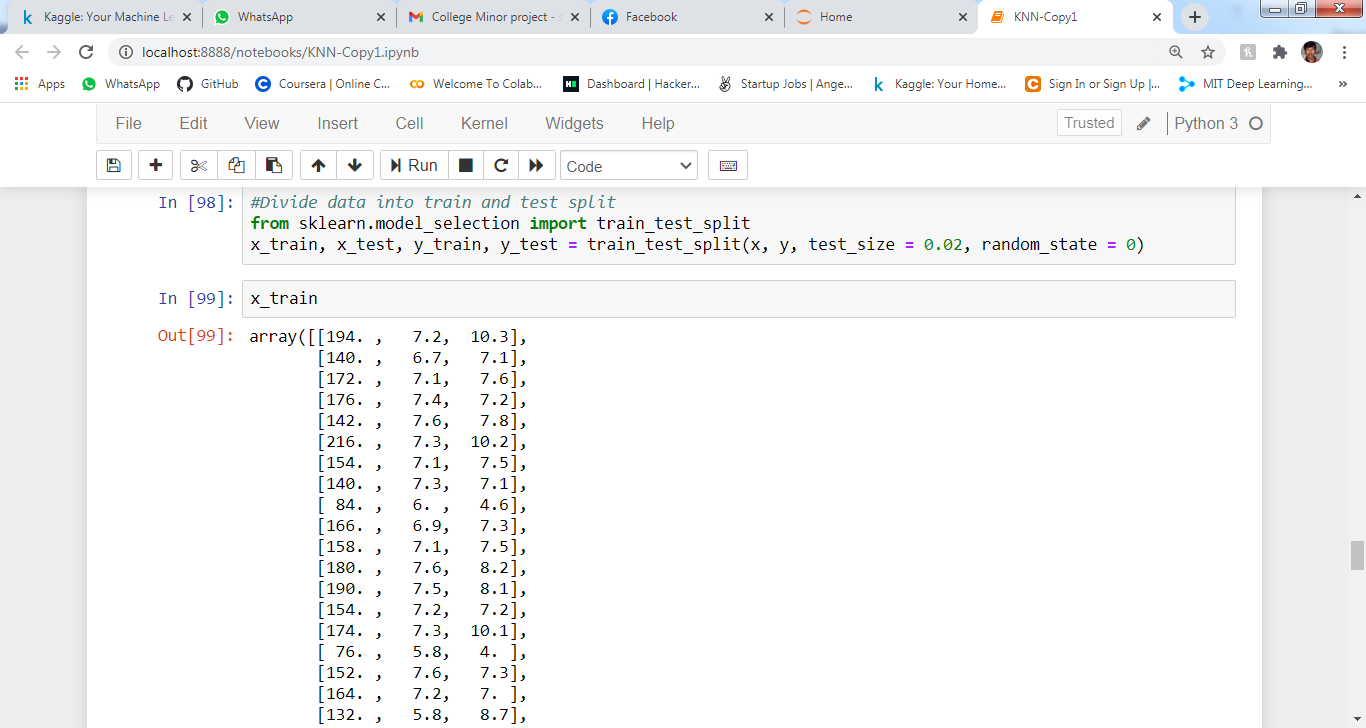
* VISUALIZATION OF DATA:

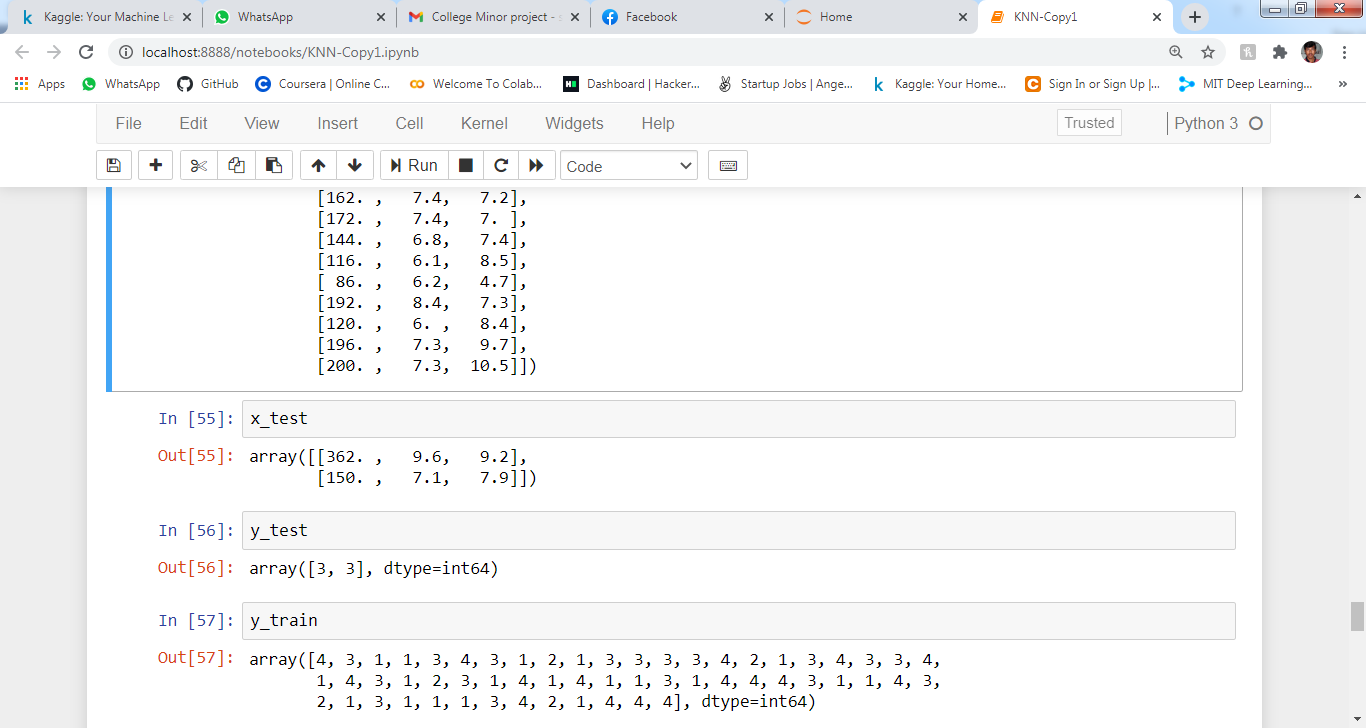




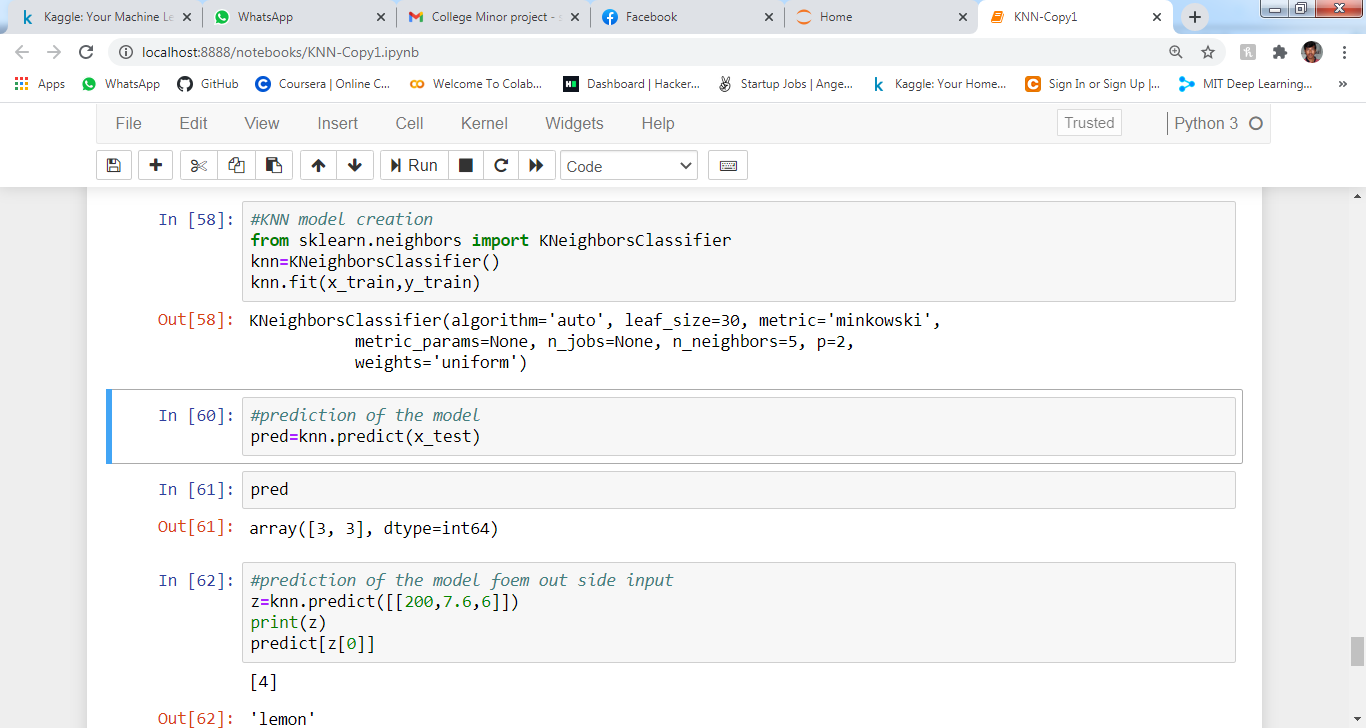


* CLEANING DATA :
* DEPENDENT AND INDEPENDENT DATA :
* SPLITTING THE DATA :

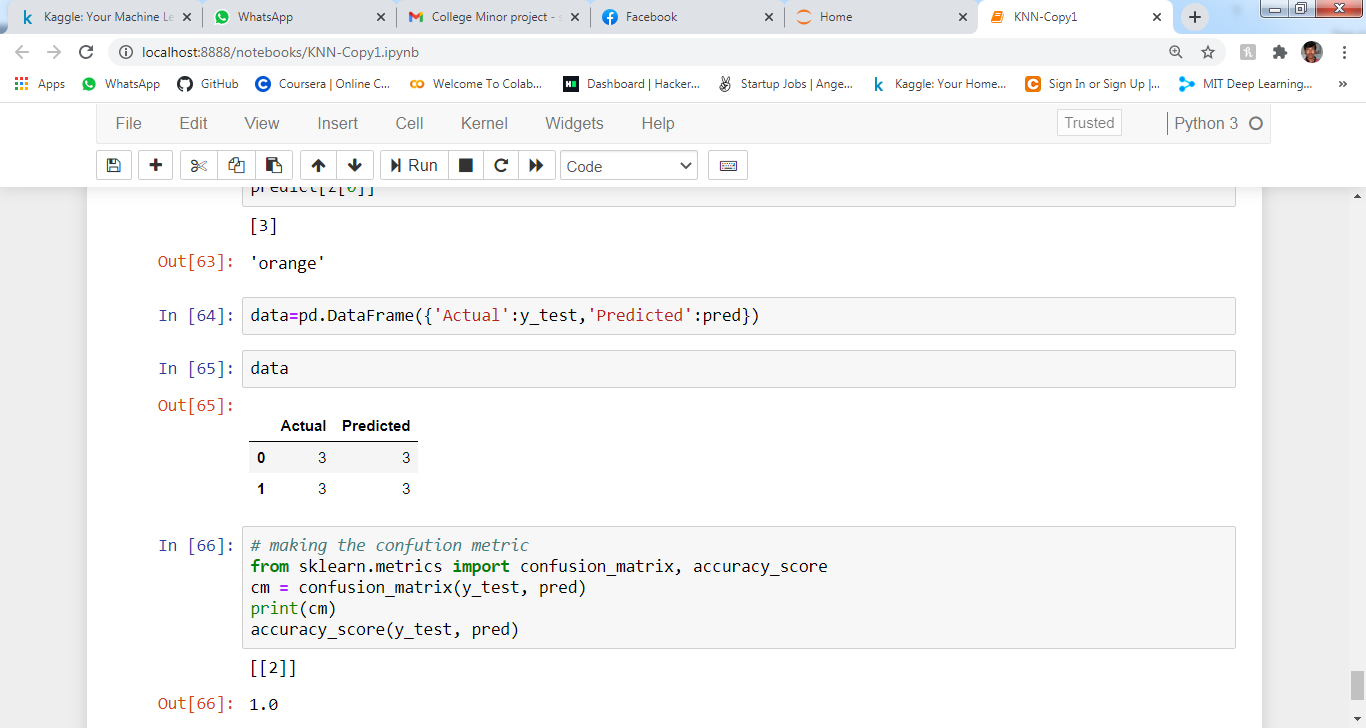




* MODEL TRAINING:



* MODEL EVALUATION :



PROJECT LIMITATIONS:

❖We worked on the **backend part of the system** thus

there is no frontend work associated which can result in a more realistic look and focus on user experience.

❖While using dataset greater than 1gb, the project won’t work properly.

FUTURE SCOPE

❖Since here only the backend part of the system is built, we can create a custom frontend which can result in a more realistic look and focus on user experience.

❖We can work on the project and use it for greater size distributed systems like hadoop.

SUMMARY

We load previous datasets to the system. Visualization of data is done to properly know about the data. Data pre-processing is not required as there was no Nan. Following operations are performed on the dataset after that. User input data to the system in order to check whether the passenger is survived or not. Building model using Logistic Regression Algorithm and train the data set. Test the dataset using model. Get the evaluation result. Get the predicted voting from all classifiers and gives the survival result.

BIBLIOGRAPHY

* <https://www.wikipedia.org/>
* <https://stackify.com/aws-lambda-with-python-a-complete-getting-started-guide/>
* <https://www.kaggle.com>

Code:

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

%matplotlib inline

fruit=pd.read\_csv(r'C:\Users\SAYAN\Desktop\knn\_fruit\_data\_with\_colors.csv')

fruit.head()

fruit.info()

predict=dict(zip(fruit.fruit\_label.unique(),fruit.fruit\_name.unique()))

predict

apple\_data=fruit[fruit['fruit\_label']==1]

apple\_data

lemon\_data=fruit[fruit['fruit\_label']==4]

lemon\_data

len(lemon\_data.index)

plt.scatter(fruit['width'],fruit['height'])

plt.xlabel('fruit\_width')

plt.ylabel('fruit\_hight')

plt.show()

plt.scatter(fruit['mass'],fruit['width'])

plt.scatter(fruit['mass'],fruit['height'])

plt.scatter(fruit['mass'],fruit['color\_score'])

import seaborn as sns

fruit.isnull().sum()

sns.heatmap(fruit.isnull())

x=fruit[['mass','width','height']].values

y=fruit['fruit\_label'].values

x

y

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

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size = 0.02, random\_state = 0)

x\_train

x\_test

y\_test

y\_train

from sklearn.neighbors import KNeighborsClassifier

knn=KNeighborsClassifier()

knn.fit(x\_train,y\_train)

pred=knn.predict(x\_test)

pred

z=knn.predict([[200,7.6,6]])

print(z)

predict[z[0]]

data=pd.DataFrame({'Actual':y\_test,'Predicted':pred})

data

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

cm = confusion\_matrix(y\_test, pred)

print(cm)

accuracy\_score(y\_test, pred)

THANK YOU