|  |
| --- |
| ***MINI PROJECT REPORT ON AGRICULTURE*** |
| **RECOMMENDING ADEQUATE CROP** |
| **USING DATA SCIENCE AND ANALYTICS** |

|  |
| --- |
| Nehla Shajahan  21-01-2021 |

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| SL NO: | CONTENT | PAGE NO |
| 1 | Abstract | 3 |
| 2 | Introduction | 4 |
| 3 | Data Description | 5 |
| 4 | Data Visualisation | 8 |
| 5 | Building various models | 9 |
| 5.a | KNN model | 10 |
| 5.b | Decision Tree model | 12 |
| 5.c | SVM model | 13 |
| 5.d | Random Forest model | 14 |
| 6 | Conclusion | 15 |

**ABSTRACT**

In this project, we aim to find an adequate predictive model to maximize the agricultural yield by recommending the appropriate crop with the help of data science and analytics. Here we use the idea of precision agriculture which is the technology of “site-specific” farming. It has provided us with the advantage of efficient input, output and better decisions regarding farming. The dataset used here was built by expanding datasets of rainfall, climate and fertilizer data available for India and was taken from Kaggle. It has various parameters like ratio of Nitrogen content in soil (N), ratio of Phosphorous content in soil (P), ratio of Potassium content in soil (K), temperature in degree Celsius, relative humidity in %, ph value of the soil, rainfall in mm. So, the dataset uses these 7 parameters to predict the appropriate crop that is best suitable for that area. It comprises of 1900 rows and the crop labels given for each of these rows are rice, maize, chickpea, kidney beans, pigeon peas, moth beans, mung bean, black gram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya and coconut.

**INTRODUCTION**

Agriculture is one of the most vital and extensively practiced occupations in India and it has a crucial role in the development of our country. Around 60 percent of the total land in the country is used for agriculture to meet the needs of 1.2 billion people, so taming crop production is therefore seen as a significant aspect of agriculture. Basically, if we have a piece of land, we need to know what kind of crop can be grown in this area. Agriculture depends on the various soil properties. Production of crops is a difficult task since it involves various factors like soil type, temperature, humidity etc. If it is possible to find the crop before sowing it, it would be of great help to the farmers and the other people involved to make appropriate decisions on the storage and business side. There are ML models that solve agricultural problems by monitoring the agricultural area on the basis of soil properties and recommending the most appropriate crop to farmers, thereby helping them to significantly increase productivity and reduce loss.

Machine learning is an important decision support tool for crop prediction, including supporting decisions on what crops to grow and what to do during the growing season of the crops. Several machine learning algorithms have been applied to support crop prediction research.

An ML model can be descriptive or predictive, depending on the research problem and research questions. While descriptive models are used to gain knowledge from the collected data and explain what has happened, predictive models are used to make predictions in the future.

Crop yield prediction is an essential task for the decision-makers at national and regional levels for rapid decision-making. An accurate crop yield prediction model can help farmers to decide on what to grow and when to grow. There are different approaches to crop yield prediction.

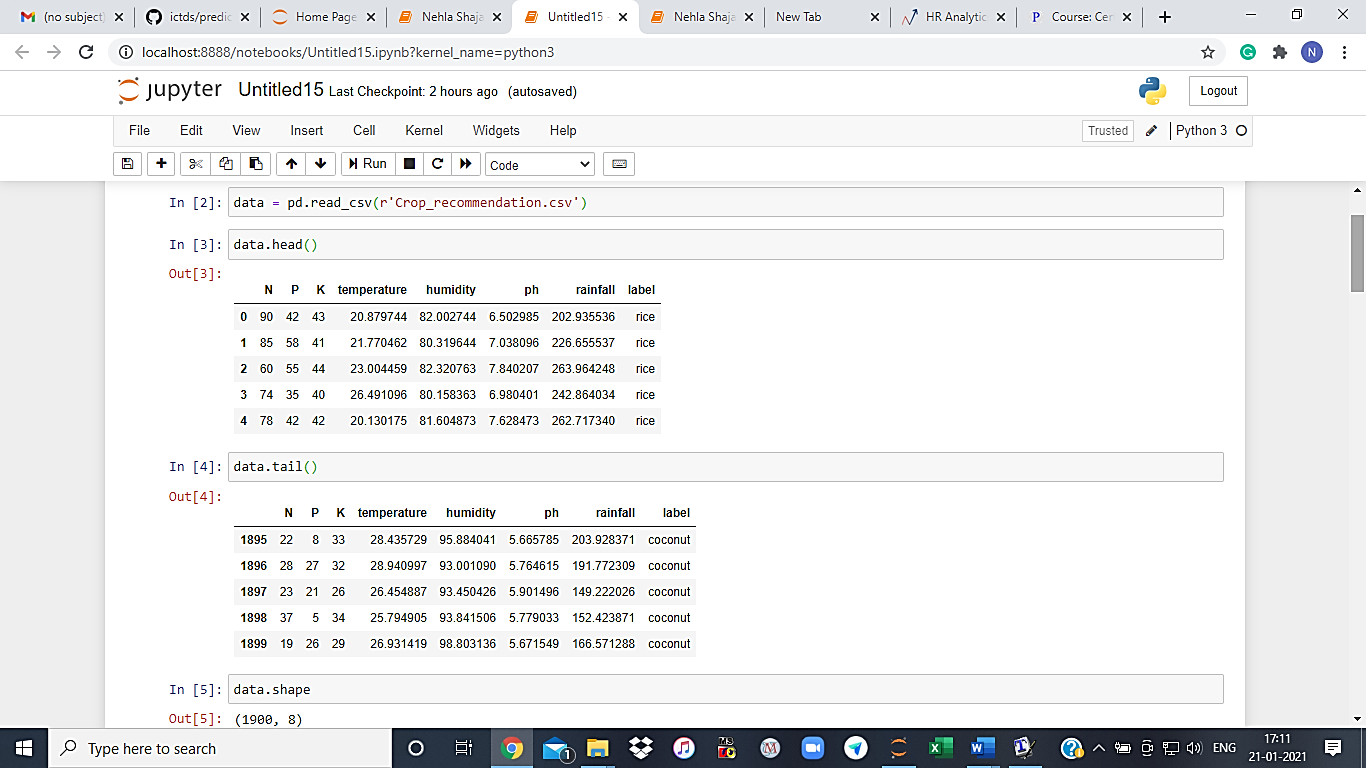
In this project, we are trying to devise an ML model using data science and analytics to predict the best suitable crop for a given area. Our project is a recommendation system which makes use of different machine learning techniques such that it recommends the suitable crops based on the input soil parameters.

**DATA DESCRIPTION**

The dataset used here is basically taken from Kaggle and was titled “Crop recommendation”. The data set is quite large with 1900 rows and 8 columns, 7 of them being the parameters responsible to recommend the appropriate crop which is indicated as ‘label’ in the 8th column.

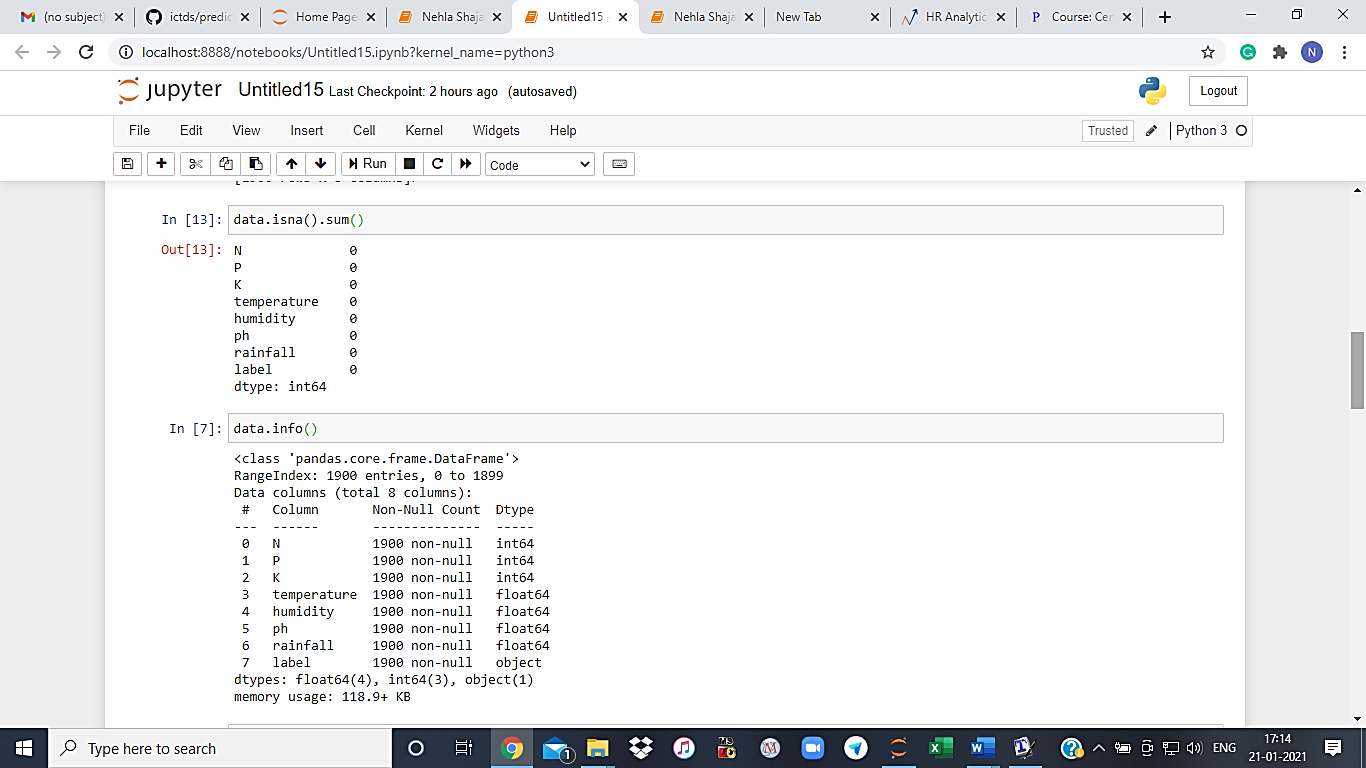
STEP 1

First of all we load the dataset into our Jupyter notebook and assign it to a variable. We further explore the dataset by using functions such as data.head(), data.tail(), data.shape etc to derive meaningful insights.



STEP 2

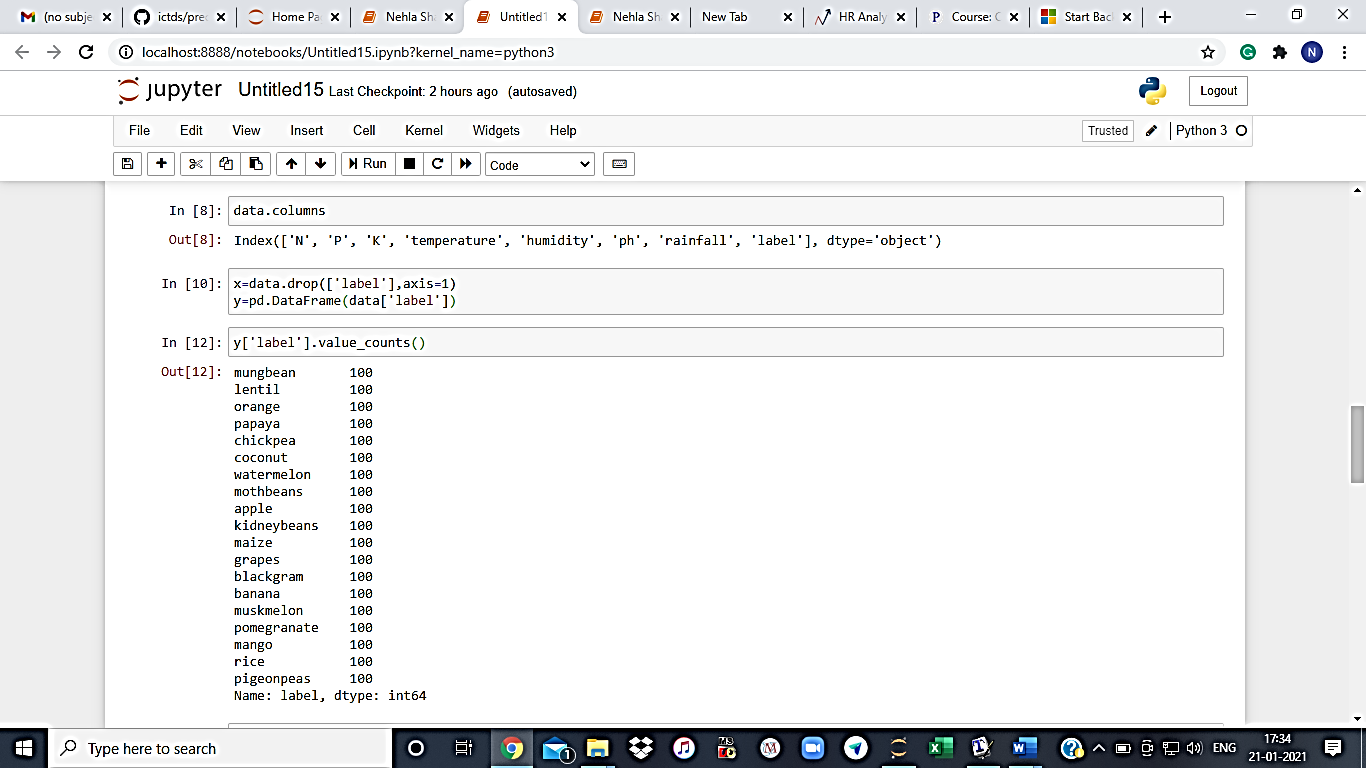
We now look for null values in the dataset and for the datatypes present. We can see that there are no null values present and only the datatype of column 8, label is object and every other column is either of float or integer type.



STEP 3

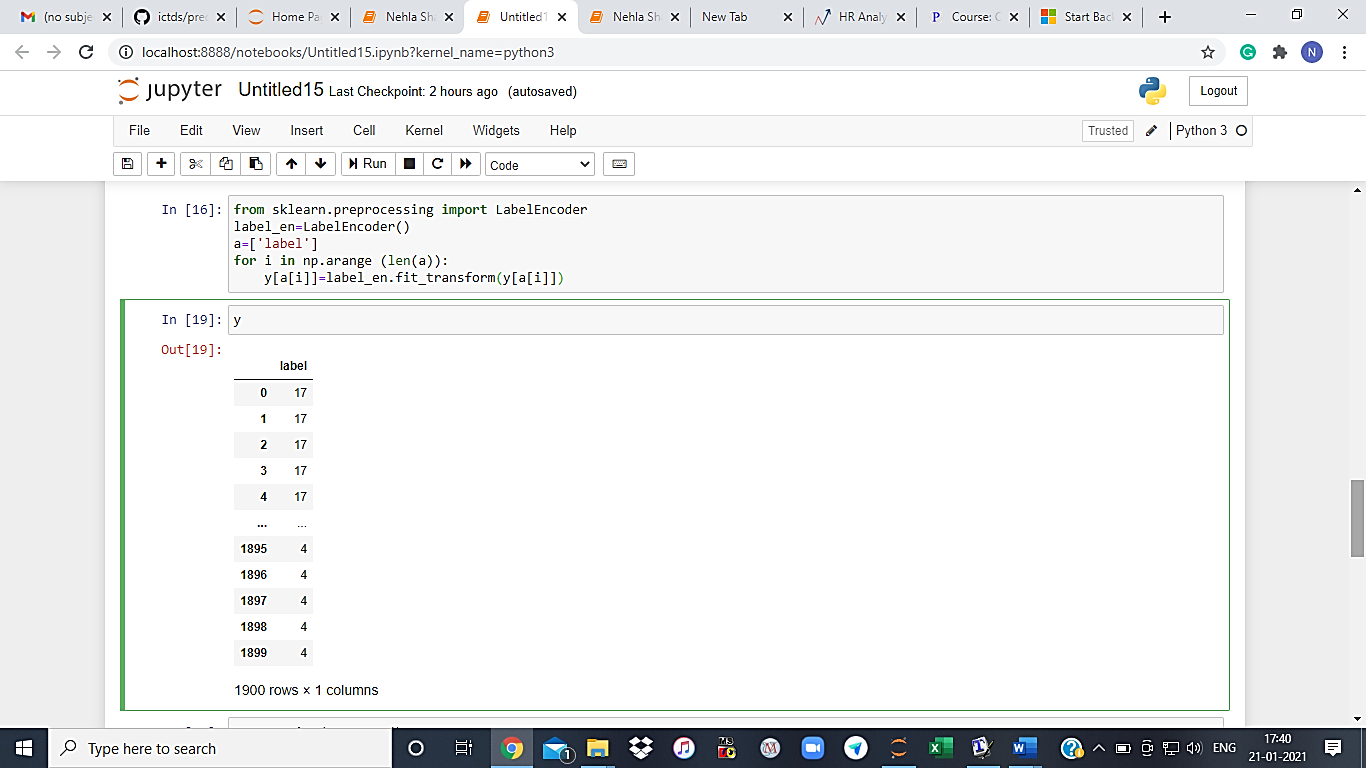
Now we are pre-processing the dataset to determine the target variable and the other parameters. The target is ‘label’ stored in ‘y’ and every other feature is stored in ‘x’.

We further use the function ‘value\_counts’ to determine the number of times each of these labels are occurring.



STEP 4

Since the target column is of object datatype, we have to perform Label encoding to convert it.

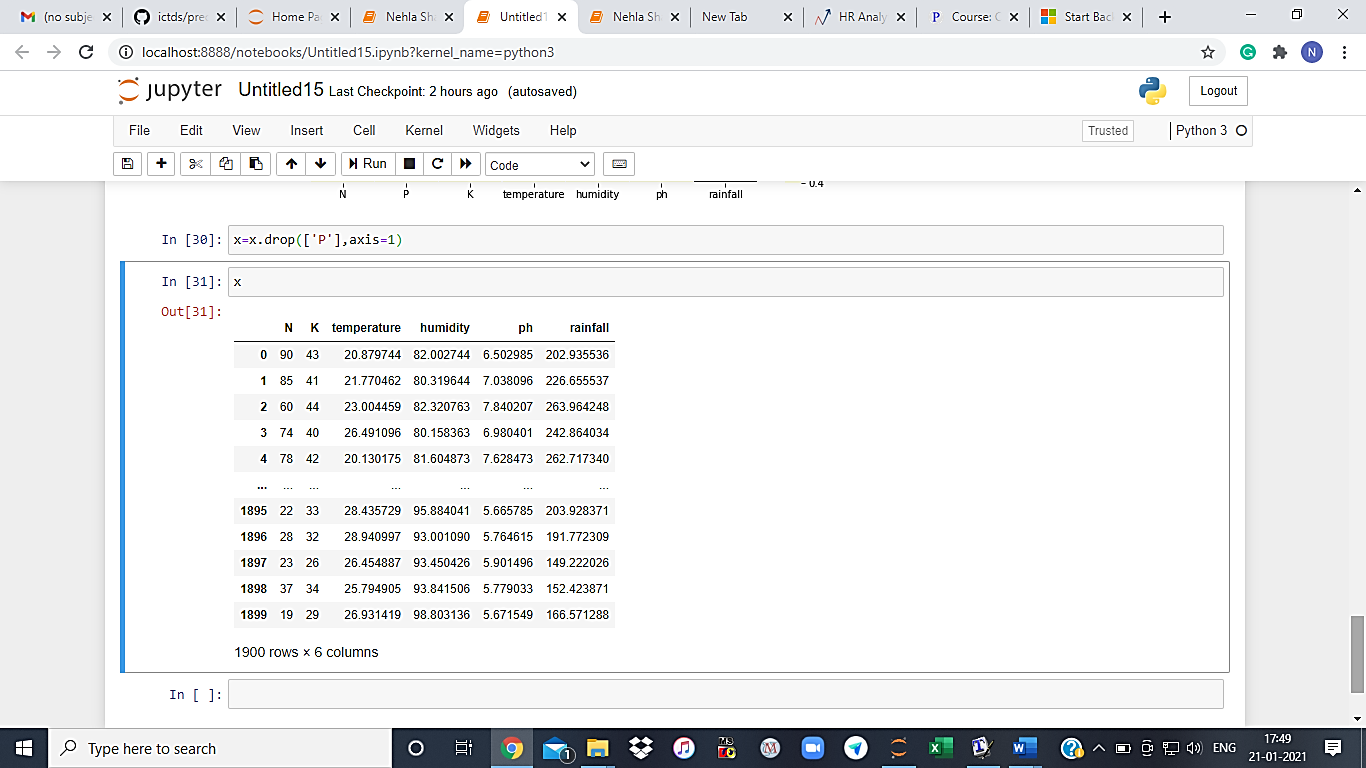


**DATA VISUALISATION**

Now we are plotting a heatmap to find out variables that are correlated with each other and hence exhibit similar behaviour.



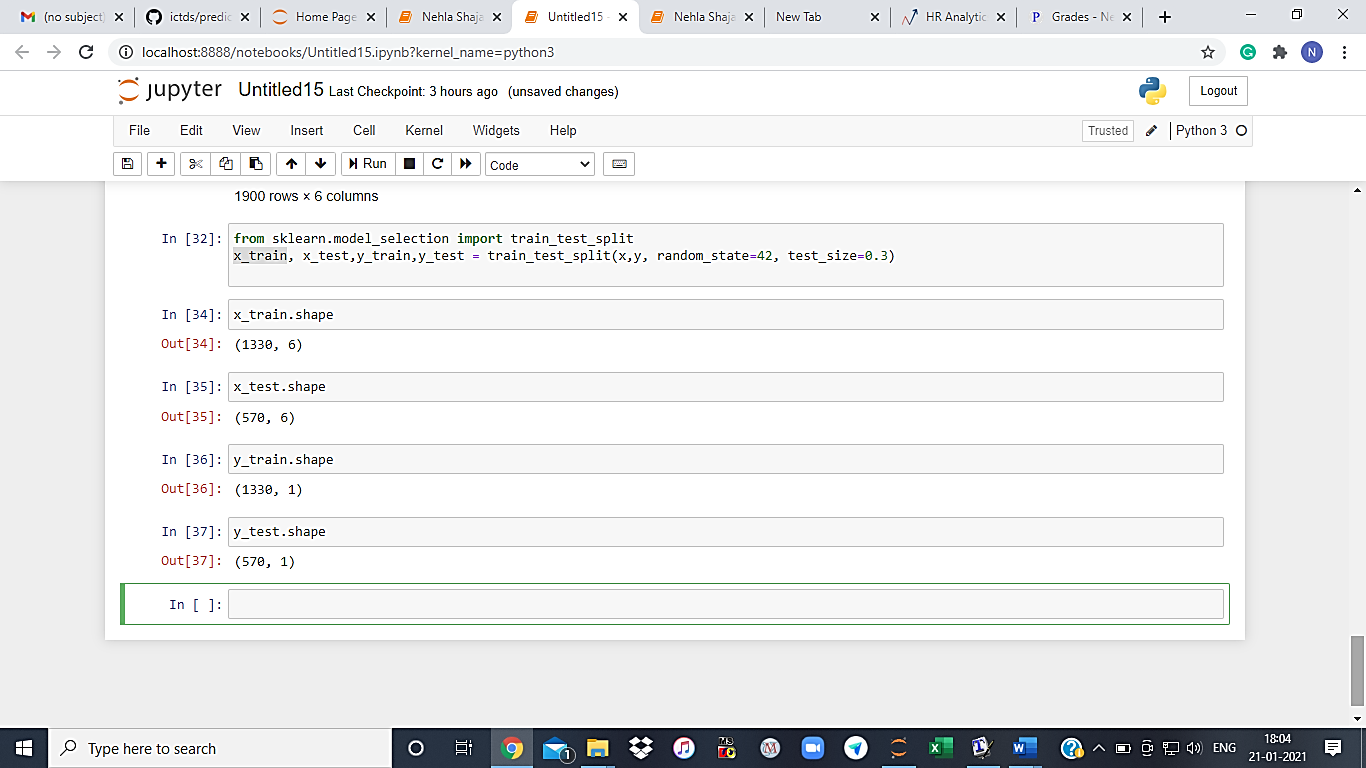
Here we can clearly see that there exists a pretty strong positive correlation of 0.74 between K & P which indicates that these two features will have almost the same effect on the crop. Hence, we can drop either of these two features due to multicollinearity.



**BUILDING VARIOUS MODELS**

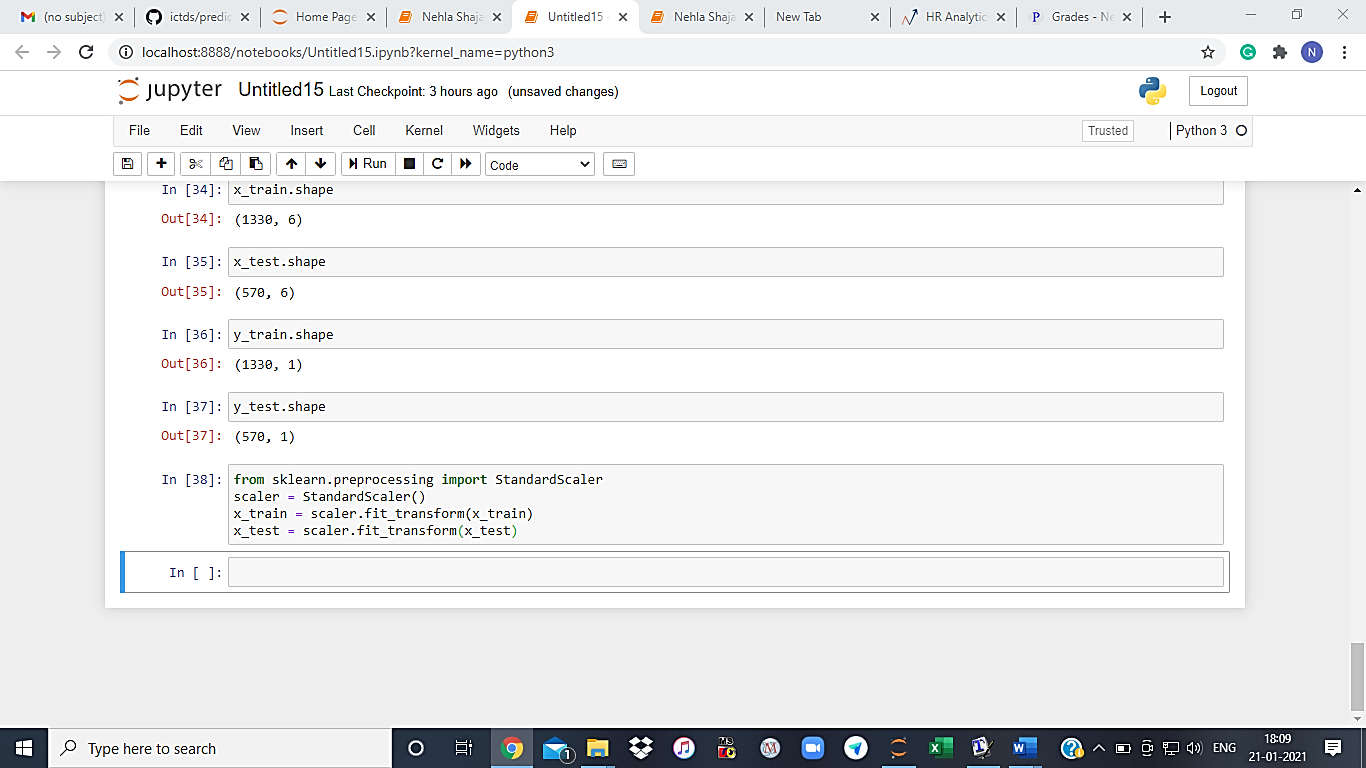
**Step 1**

**Splitting the data into train and test** and checking the shape of each of these 4 datasets.



**Step 2**

**Scaling the dependent variables so that all of the variables come under one standard scale.**

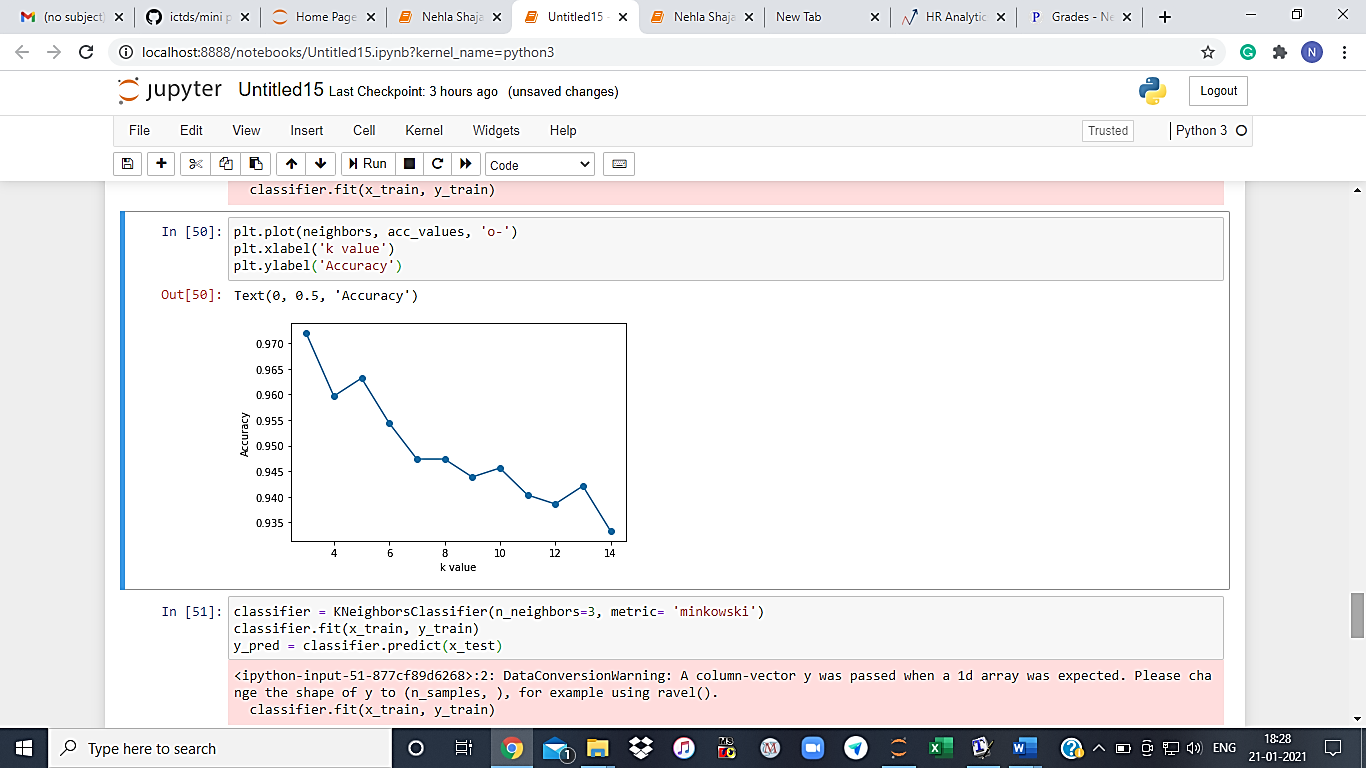


**Step 3: Building various ML models and evaluating their performance**

*1.Building a KNN model*

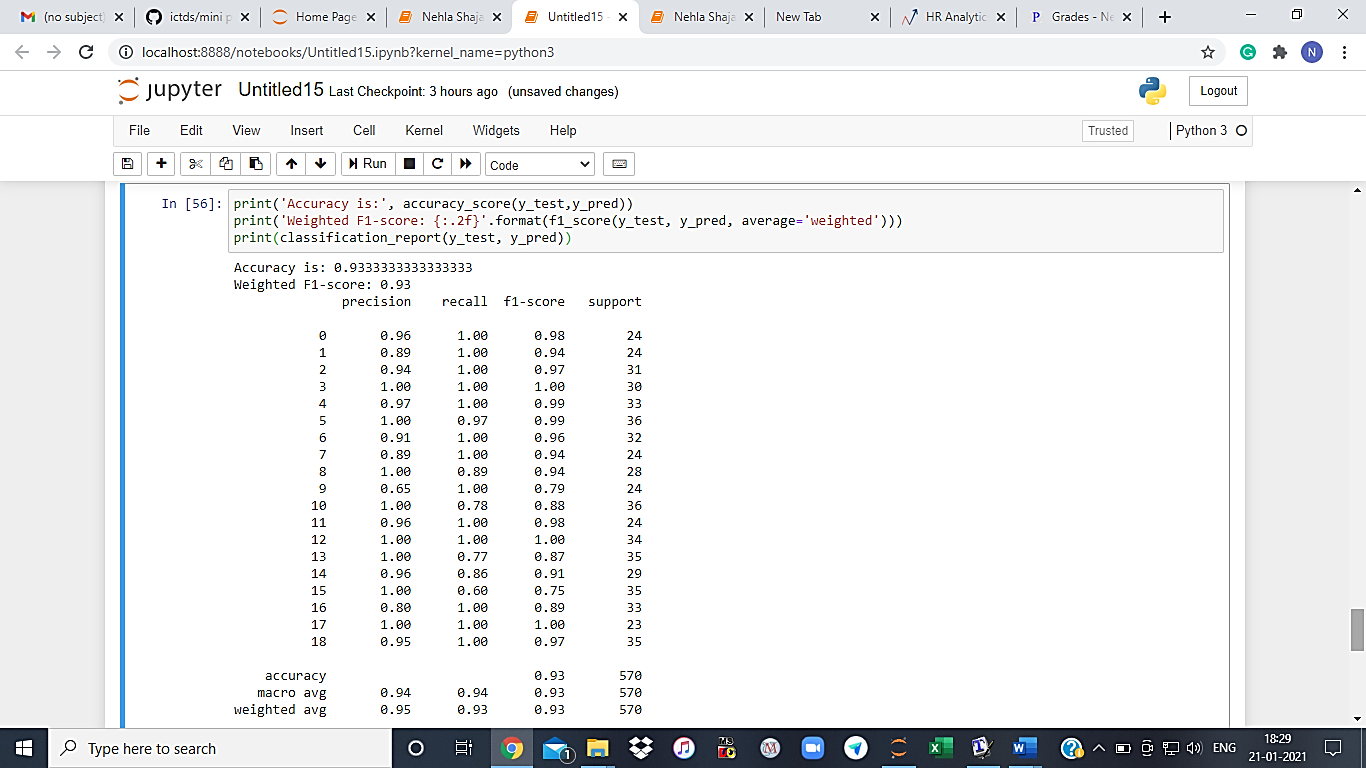


Plotting a graph between accuracy and k values will give us the k value for which there is a maximum accuracy.



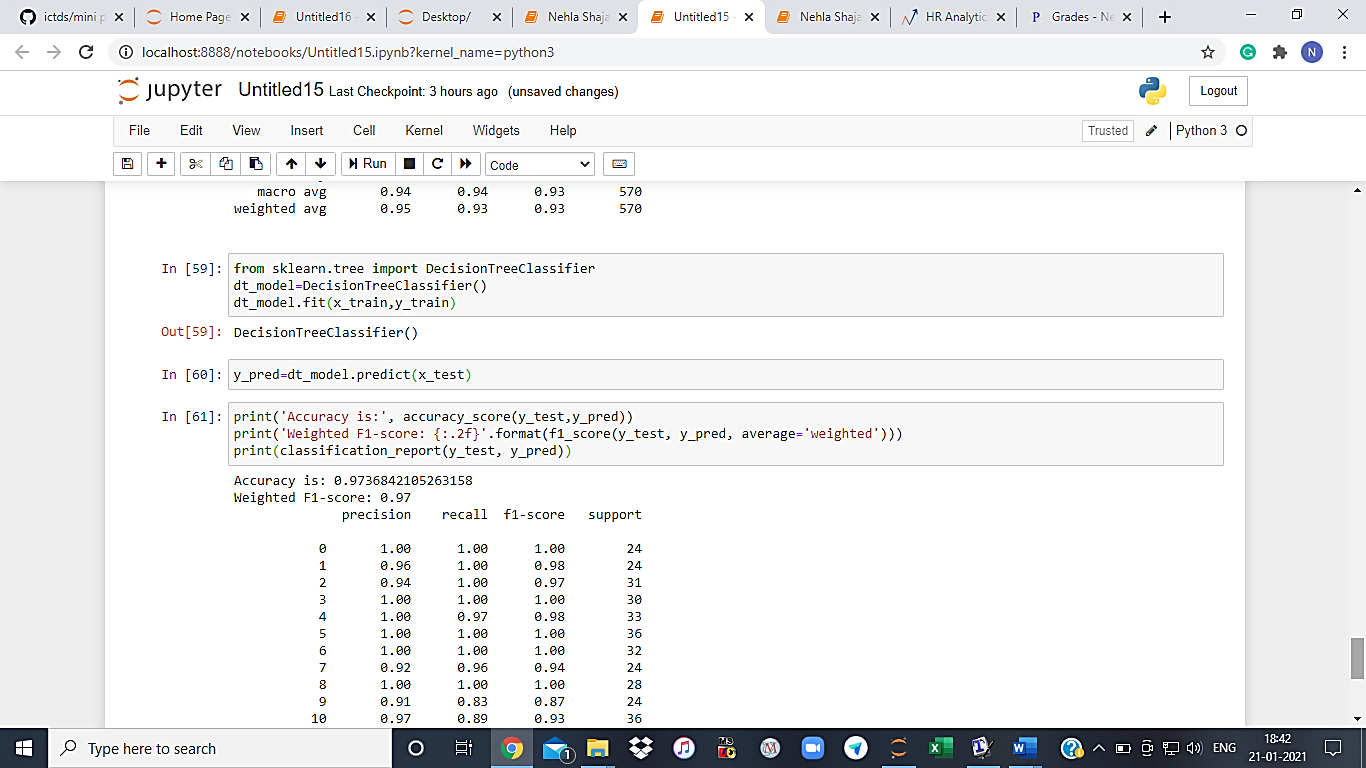
Here k=3 gives a maximum accuracy and hence the optimum value of k is taken as 3.

Now evaluating the model by finding out accuracy, f1\_score and the classification report.

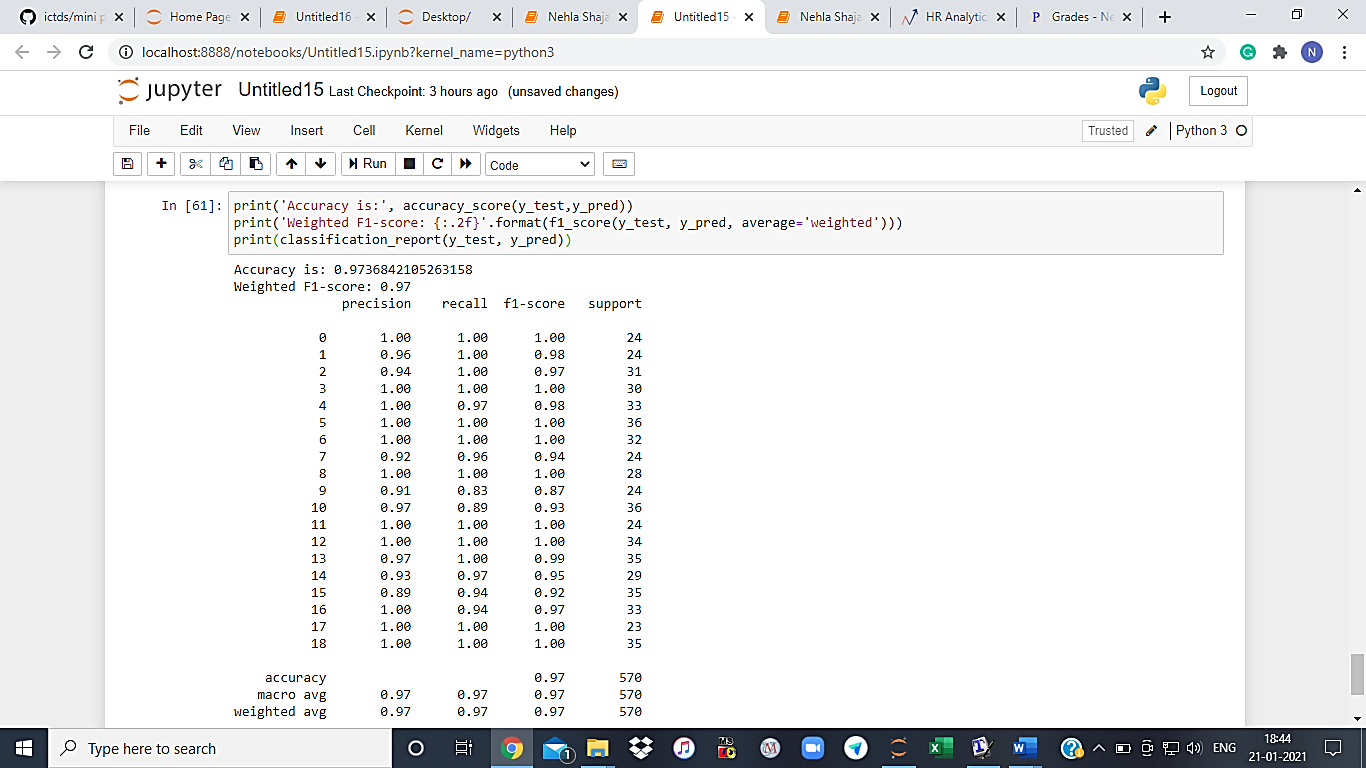


KNN model gives: Accuracy = 93.3%, Weighted f1 score = 0.93

*2.Building a decision tree model*



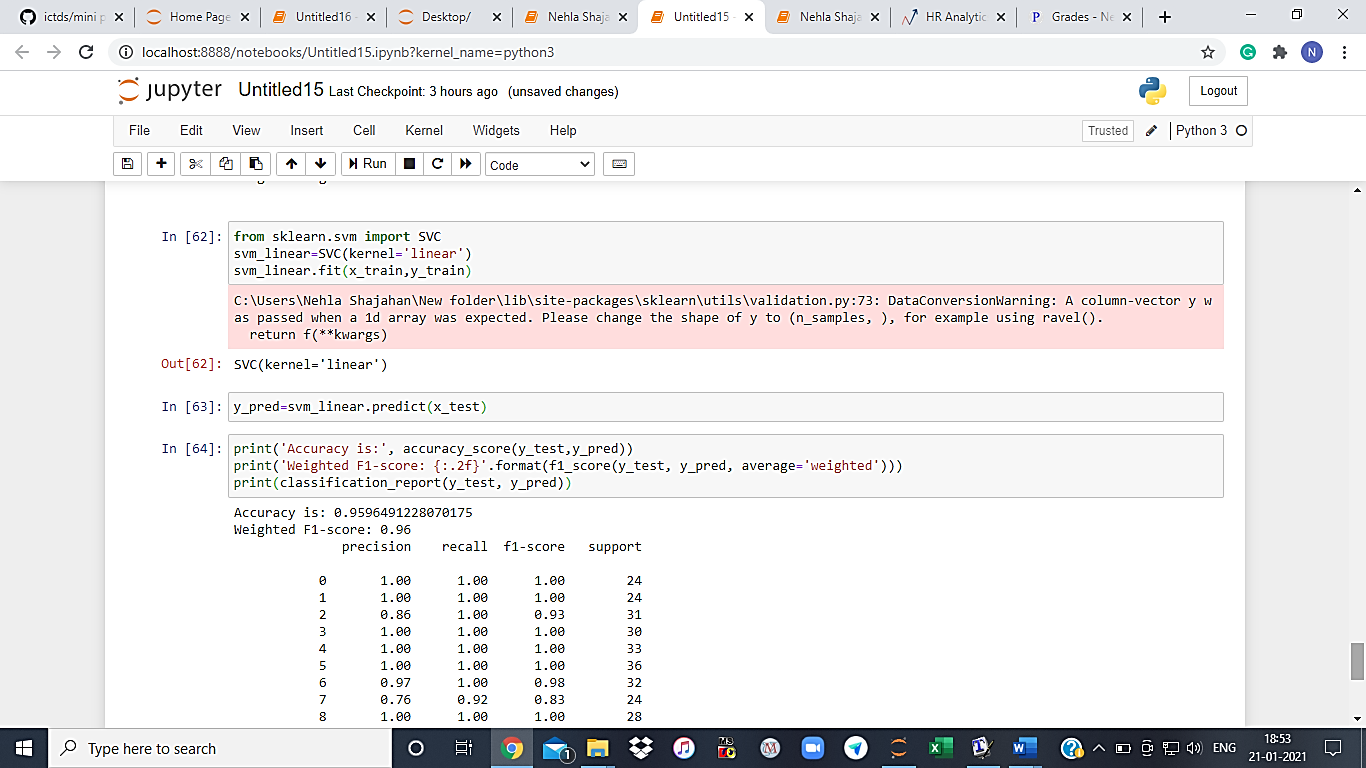
Evaluating the model performance.



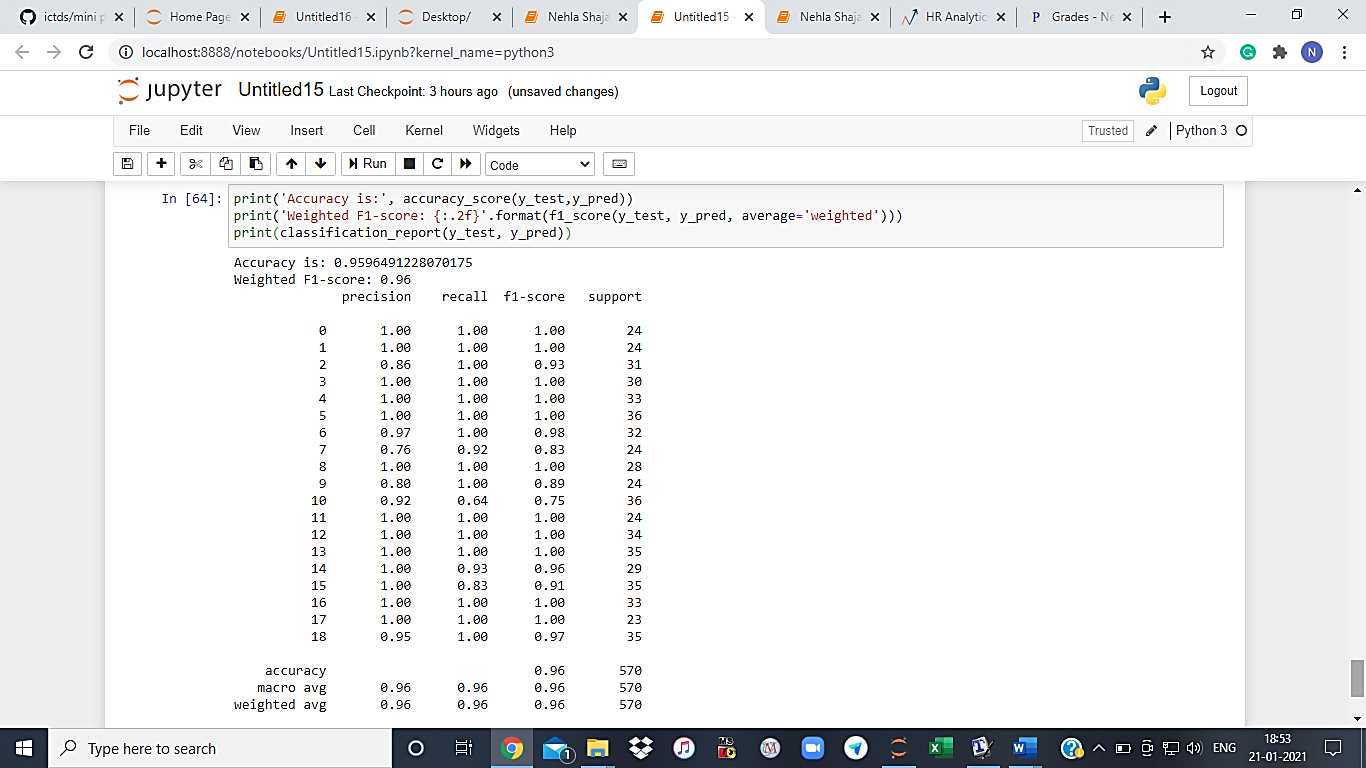
Decision Tree model gives:

Accuracy = 97%, weighted f1 score =0.97

*3. Building a linear SVM model*



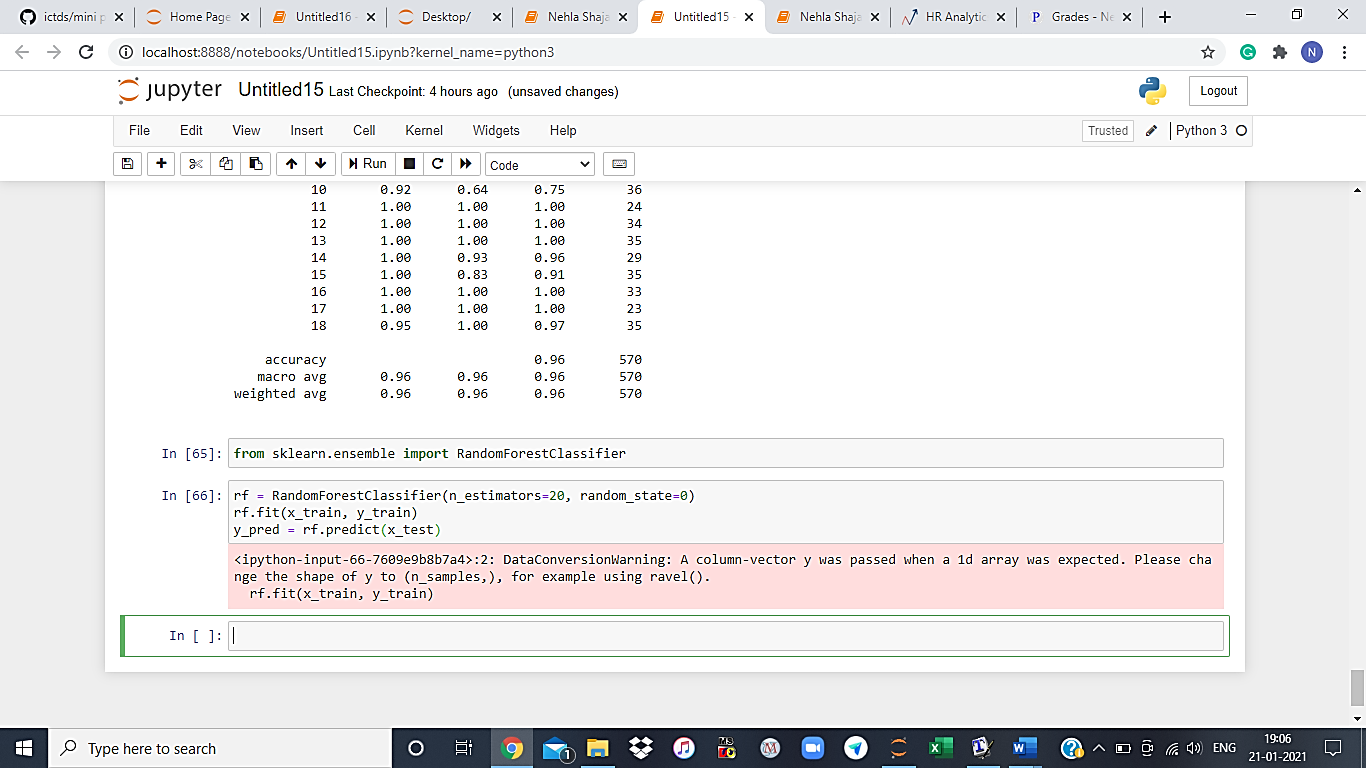
Evaluating the model performance



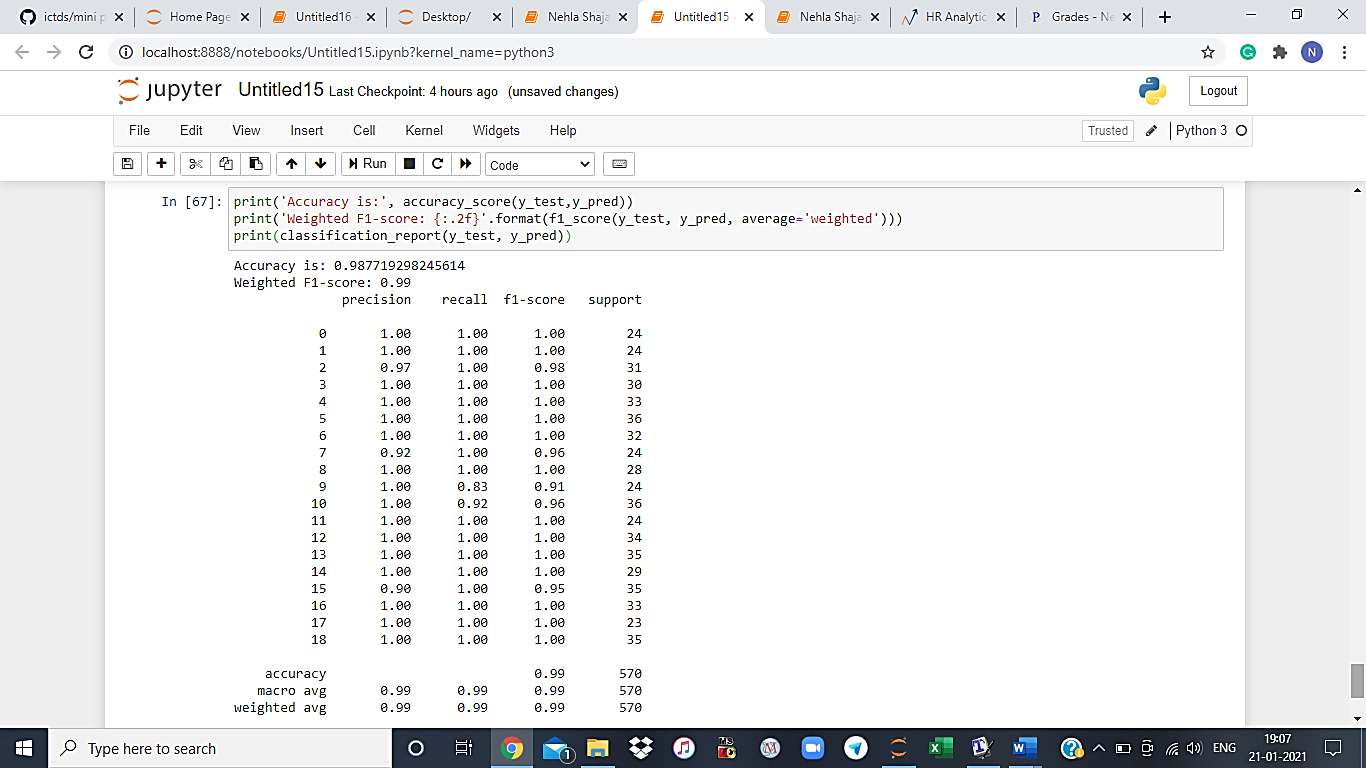
Linear SVM model gives us:

Accuracy = 95.9% ; weighted f1 score= 0.96

*4. Building a random forest model*



Evaluating the model performance



Random Forest model gave us:

Accuracy = 98.77%

Weighted f1\_score = 0.99

**CONCLUSION: RESULT ANALYSIS**

The proposed project takes the soil N, P, K, and pH values, temperature, humidity, rainfall into consideration and determines which are the best productive crops that can be grown in that appropriate soil conditions. Since the system lists all potential crops it helps the farmer determine which crop to be grown in their area. This system thus helps the farmer to decide on the maximum profitable crop and also helps in finding new crops that can be cultivated which have not been cultivated till that time by the farmer.

From the above analysis, we can very well conclude that each of the ML model gave us a relatively good accuracy suggesting that the process we carried out and the model made is good and can be used for recommending the ideal crop for a given plot.

Among the 4 models used, Random Forest model gave us the highest accuracy of almost 99% and therefore can be recommended for making the ML model in this case.