# PROFESSIONAL TRAINING REPORT

**at**

**Sathyabama Institute of Science and Technology (Deemed to be University)**

Submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering

By

**Name: KARTHICK B**

**(Reg .no. 39110465)**



# DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING SCHOOL OF COMPUTING

**SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY JEPPIAAR NAGAR, RAJIV GANDHI SALAI,**

**CHENNAI – 600119, TAMILNADU**

## NOVEMBER 2021

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY**

**(DEEMED TO BE UNIVERSITY)**

## Accredited with Grade “A” by NAAC

**(Established under Section 3 of UGC Act, 1956)**

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI– 600119**

[**www.sathyabamauniversity.ac.in**](http://www.sathyabamauniversity.ac.in/)

**DEPARTEMENT OF COMPUTER SCIENCE AND ENGINEERING**

**BONAFIDE CERTIFICATE**

This is to certify that this Project Report is the bonafide work of **KARTHICK (39110465)** who carried out the project entitled “**ATM SIMULATION”** under my supervision from April 2020 to June 2020.

**INTERNAL GUIDE:**

**Name: Dr.A.CHRISTY,MCA.,Ph.D**

**HEAD OF DEPARTMENT**

**Dr. L. LAKSHMANAN M.E., Ph.D.,**

**Dr.S.VIGNESHWARI, M.E., Ph.D.,**

**Submitted for Viva Voice Examination heldon**

**Internal Examiner ExternalExaminer**

# DECLARATION

**I KARTHICK** hereby declare that the Project Report entitled ‘**DRY BEAN CLASSIFICATION** done by me under the guidance of **Dr.A.CHRISTY, MCA.,Ph.D** and **Dr. L. LAKSHMANAN M.E., Ph.D., and Dr.S.VIGNESHWARI, M.E.,Ph.D.,**at Sathyabama Institute of Science and Technology is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

## DATE:

**PLACE :** CHENNAI **SIGNATURE OF THECANDIDATE**

## ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E., Ph.D.**, **Dean**, School of Computing, **Dr.S.Vigneshwari M.E., Ph.D., and Dr.L.Lakshmanan M.E., Ph.D.,** Heads of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. A. CHRISTY,MCA.,Ph.D** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and engineering** who were helpful lin many ways for the completion of the project.

# TRAINING CERTIFICATE

A picture containing text

Description automatically generated

**ABSTRACT**

In this project, I had done a machine learning model that predicts the classification of Dry bean [about 7 classes are there]by providing a dry bean dataset to the model. The input given to the model contains various features as input such as area of the bean, roundness, solidity of the bean, aspect ratio, and so on. By these attributes, we first train our model with 80% [training data] of the given input data and keep aside the remaining 20% [testing data] of the data for validation. After training the model, we must test our model to check the accuracy of the model. So, with the help of the trained model, the remaining 20% of the dataset is fitted inside the model. With the test data being trained, we are going to predict the class. Thus, we check whether the predicted value and the actual value are the same. So, by this method, we can get the accuracy of the model and can predict the dry bean classes with different data sets too. In this project, we will use some machine learning algorithms to do the classification.

## TABLE OF CONTENTS:

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Title.**  Abstract  List of Figures  List of Abbreviations | **Page No.** |
| **1.** | **INTRODUCTION**   * 1. INTRODUCTION TO DRY BEAN CLASSIFICATION   2. NEED FOR THE DRY BEAN CLASSIFICATION   3. INTRODUCTION TO MACHINE LEARNING   4. INTRODUCTION TO PYTHON | 1  1  2  2  3 |
| **2.** | **LITERATURE SURVEY**  2.1. REVIEW OF LITERATURE | 5  5 |
| **3.** | **SYSTEM SPECIFICATION**  3.1. HARDWARE REQUIREMENTS  3.2 PROPOSED DESCRIPTION  3.3 PROBLEM DESCRIPTION | 8  8  8  8 |
| 4 | **ABOUT THE DATASET AND ALGORITHM ANALYSIS**   * 1. ABOUT THE DATASET AND DESCRIPTION OF THE DATASET   2. ALGORITHM ANALYSIS | 9  9  10 |
| **5** | **SUMMARY AND CONCLUSION**  **REFERENCES**  **APPENDIX**  A. SOURCE CODE  B. SCREEN SHOTS | 12  23  13  13  17 |

## LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| **FIGURE NO.** | **FIGURE NAME.** | **PAGE NO.** |
| **5.1** | LOGISTIC REGRESSION SOURCE CODE SCREENSHOT | 17 |
| **5.2**  **5.3**  **5.4**  **5.5**  **5.6** | RANDOM FOREST CLASSIFIER SOURCE CODE SCREENSHOT    DECISION TREE CLASSIFIER SOURCE CODE SCREENSHOT  ACCURACY LINE PLOT  ACCURACY BAR CHART  DECISION TREE PLOTTING | 18  19  20  21  22 |

**CHAPTER 1**

**1.1INTRODUCTION**

* 1. **Introduction to the Dry Bean Classification:**

There is a wide range of genetic diversity of dry bean which is the most produced one among the edible legume crops in the world. Seed quality is definitely influential in crop production. Therefore, seed classification is essential for both marketing and production to provide the principles of sustainable agricultural systems. The primary objective of this study is to provide a method for obtaining uniform seed varieties from crop production, which is in the form of population, so the seeds are not certified as a sole variety. Thus, a computer vision system was developed to distinguish seven different registered varieties of dry beans with similar features in order to obtain uniform seed classification. For the classification model, images of 13,611 grains of 7 different registered dry beans were taken with a high-resolution camera. A user-friendly interface was designed using the MATLAB graphical user interface (GUI). Bean images obtained by computer vision system (CVS) were subjected to segmentation and feature extraction stages, and a total of 16 features; 12 dimension and 4 shape forms, were obtained from the grains. Multilayer perceptron (MLP), Support Vector Machine (SVM), k-Nearest Neighbors (kNN), Decision Tree (DT) classification models were created with 10-fold cross-validation and performance metrics were compared. Overall correct classification rates have been determined as 91.73%, 93.13%, 87.92% and 92.52% for MLP, SVM, kNN and DT, respectively. The SVM classification model, which has the highest accuracy results, has classified the Barbunya, Bombay, Cali, Dermason, Horoz, Seker and Sira bean varieties with 92.36%, 100.00%, 95.03%, 94.36%, 94.92%, 94.67% and 86.84%, respectively. With these results, the demands of the producers and the customers are largely met about obtaining uniform bean varieties. In this project, we are going to predict the class of the dry bean by using SOME **MACHINE LEARNING** algorithms.

.

## 1.2 NEED FOR THE DRY BEAN CLASSIFICATION:

There is a necessity for classification of dry bean as there are 7 different types of dry bean classes are present. So, it is necessary for someone to know how to differentiate them and classify them according to their class and kind. So, it is necessary to have a dry bean classification system that predicts which class the Bean belong to given its features as inputs.

## 1.3 INTRODUCTION TO MACHINE LEARNING:

Machine Learning is a field in computer science that learns from experience without being programmed. It is a part of Artificial Intelligence, or we can say that machine learning is a sub-topic of Artificial Intelligence.

The science behind ML is to make computers perform actions by themselves. A Machine Learning algorithm is a generic program that will understand the data, and build models with that data. These models are available for the end-users to carry out tasks.

Machine Learning is used anywhere from automating mundane tasks to offering intelligent insights, industries in every sector try to benefit from it. You may already be using a device that utilizes it. For example, a wearable fitness tracker like Fitbit, or an intelligent home assistant like Google Home. But there are much more examples of ML in use.

* Prediction — Machine learning can also be used in the prediction systems. Considering the loan example, to compute the probability of a fault, the system will need to classify the available data in groups.
* Image recognition — Machine learning can be used for face detection in an image as well. There is a separate category for each person in a database of several people.
* Speech Recognition — It is the translation of spoken words into the text. It is used in voice searches and more. Voice user interfaces include voice dialing, call routing, and appliance control. It can also be used a simple data entry and the preparation of structured documents.
* Medical diagnoses — ML is trained to recognize cancerous tissues.
* Financial industry and trading — companies use ML in fraud investigations and credit checks.
  1. **INTRODUCTION TO PYTHON:**

Python is a high-level, dynamic programming language. Python3.4 version was used as it is a mature, versatile and robust programming language. It is an interpreted language that makes the testing and debugging extremely quickly as there is no compilation step. There are extensive open-source libraries available for this version of python and a large community of users. Python is a simple yet powerful, interpreted, and dynamic programming language, which is well known for its functionality of processing natural language data, i.e. spoken English using NLTK. Other high level programming languages such as ―R and ―MATLAB were considered because they have many benefits such as ease of use but they do not offer the same flexibility and freedom that Python can deliver.

Pandas is an open-source Python library providing high-performance data manipulation and analysis tool using its powerful data structures. The name Pandas is derived from the word Panel Data – an Econometrics from Multidimensional data. In 2008, developer Wes McKinney started developing pandas when in need of high performance, flexible tools for analysis of data. Prior to Pandas, Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data—load, prepare, manipulate, model

and analyze. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

Key Features of Pandas :

Fast and efficient Data Frame object with default and customized indexing.

Tools for loading data into in-memory data objects from different file formats.

* Data alignment and integrated handling of missing data.

Reshaping and pivoting of data sets.

Label-based slicing, indexing and sub setting of large data sets.

Columns from a data structure can be deleted or inserted.

Group by data for aggregation and transformations.

High performance merging and joining of data.

Time Series functionality.

## CHAPTER 2

**REVIEW OF LITERATURE**

1. **Variety Classification of Lactuca Sativa Seeds Using Single-Kernel RGB Images and Spectro-Textural-Morphological Feature-Based Machine Learning**

*R. Concepcion II, Sandy C. Lauguico, Khamsoy Siphengphet, Jonnel D. Alejandrino, E. Dadios, A. Bandala*

2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)

2020

Growing lettuce become popular now and the use of specific seeds on a constraint environment relies on the proper phenotypic classification of seed germplasm. Lettuce cultivars are usually differentiated based on leaf characteristics when it is matured because its seeds are characterized by almost the same spectro–textural–morphological signatures. Visual inspection of small lettuce seeds leads to the subjective classification that is unideal for seed phenotyping. To overcome this agro–industrial challenge, computer vision was incorporated with computational intelligence. In this study, two types of Lactuca Sativa L. cultivars were used, namely grand rapid and Chinese loose–leaf lettuce seeds. A consumer–grade Huawei Nova 5T mobile phone camera was used to capture single–kernel RGB images totaling to 100 samples for each variant. RGB color space thresholding was used in seed vegetation. 22 spectro–textural–morphological features were extracted and 4 were selected using feature importance with extra trees classifier (FI–ETC). KNN, decision tree for classification (DTC), Naïve Bayes (NB), and SVM with color, texture, and morphological seed features as inputs were configured to classify the lettuce seed cultivar. DTC and SVM bested other machine learning models in classifying lettuce

**2. A Deep Learning-Based Seed Classification with Mobile Application**

*Yusuf Başol, S. Toklu, Computer Science, Turkish Journal of Mathematics and Computer Science, 2021.*

TLDR

In this study, a mobile application has been developed that quickly detects and classifies seed images with high accuracy using CNN, one of the deep learning techniques.

Abstract

Seed quality is an essential factor in agricultural production. Some seeds are inherently small so it is difficult to identify and classify differences between species. In the traditional method, these differences are classified by experts considering the morphological structure, texture and color. This method involves a classification process that is costly, subjective and time confusing, what makes it necessary to develop a process that can automatically detect the type of seeds. In this study, a mobile application has been developed that quickly detects and classifies seed images with high accuracy using CNN, one of the deep learning techniques

**3. Color and texture for corn seed classification by machine vision**

*K. Kiratiratanapruk, W. Sinthupinyo, Computer Science ,International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS), 2011.*

TLDR

A method to classify more than ten categories of seed defects by using color, texture features and support vector machine (SVM) type classifier is proposed and an image capturing machine that is able to support large volume of seed samples is developed.

Abstract

Machine vision has been applied to various food materials inspection process of agricultural industry in order to achieve fast and accurate operation. In this paper, we proposed a method to classify more than ten categories of seed defects by using color, texture features and support vector machine (SVM) type classifier. We also developed an image capturing machine that is able to support large volume of seed samples. The image capturing machine was designed to control uncertainty light level, reflection and shadow appeared on seed samples. Therefore, quality of images that can be obtained from the machine in term of accurate color and exposure is high. In addition, the designed image capturing machine also provides support to background subtraction and touching object segmentation processes. In the image classification part, color histograms in RGB and HSV color space together with texture based on Grey level co-occurrence matrix (GLCM) and Local binary pattern (LBP) is adopted as features. The proposed systems were evaluated from more than 10,000 sample images. The obtained accuracies are 95.6% for normal seed type and 80.6% for group of defect seed types. The preliminary results of this study are useful information for future development of the quality control technique in practical usage

**4. Weed seeds identification by machine vision**

*P. Granitto, H. Navone, P. F. Verdes, H. Ceccatto, Engineering ,2002.*

Abstract The implementation of new methods for reliable and fast identification and classification of seeds is of major technical and economical importance in the agricultural industry. As in ocular inspection, the automatic classification of seeds should be based on knowledge of seed size, shape, color and texture. In this work, we assess the discriminating power of these characteristics for the unique identification of seeds of 57 weed species. Using the performance of a naive Bayes classifier as selection criterion, we identified a nearly optimal set of 12 (six morphological+four color+two textural) seed characteristics to be used as classification parameters. We found that, as expected, size and shape characteristics have larger discriminating power than color and textural ones. However, all these features are required to reach an identification performance acceptable for practical applications. In spite of its simplicity, the naive Bayes classifier reveals itself surprisingly good for the identification of seed species. This might be due to the careful selection of the feature set, leading to nearly independent parameters as assumed by this method. We also found that, using the same feature set, a more sophisticated classifier based on an artificial neural network committee performs only slightly better than this simple Bayesian approach.

**CHAPTER 3**

**SYSTEM SPECIFICATION**

## 3. SYSTEM SPECIFICATION

**Hardware Requirements:**

1. Processor – Pentium4
2. RAM – 1 GB
3. Hard Disk –40GB
4. Mouse – Standard Mouse
5. Keyboard – Logitech Keyboard
6. Processor Speed –2.4GHZ

## Display Mode:

1. Color Quality – Highest[32bit]
2. Screen Resolution – 1024 by 768Pixels

## PROJECT DESCRIPTION Need for The Software:

Nowadays everyone is very tired of classifying the beans manually. By seeing them through the naked eye and saying whether a particular bean is of this class is very tough and not accurate and very time-consuming too. So with the use of this system, the machine can predict whether this is of the particular class or not in no time. This makes our job much more easier and efficient with very good accuracy and at a low cost .

## PROBLEM DESCRIPTION:

Seven different types of dry beans were used in this research, considering the features such as form, shape, type, and structure by the market situation. A computer vision system was developed to distinguish seven different registered varieties of dry beans with similar features in order to obtain uniform seed classification. For the classification model, images of 13,611 grains of 7 different registered dry beans were taken with a high-resolution camera. Bean images obtained by computer vision system were subjected to segmentation and feature extraction stages, and a total of 16 features; 12 dimensions and 4 shape forms, were obtained from the grains.

**CHAPTER 4**

**ABOUT THE DATASET AND ANALYSIS OF ALGOITHM**

**4.1 About the dataset:**

|  |  |
| --- | --- |
| **Abstract**: Images of 13,611 grains of 7 different registered dry beans were taken with a high-resolution camera. A total of 16 features; 12 dimensions and 4 shape forms, were obtained from the grains. |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Characteristics:** | Multivariate | **Number of Instances:** | 13611 | **Area:** | Computer |
| **Attribute Characteristics:** | Integer, Real | **Number of Attributes:** | 17 | **Date Donated** | 2020-09-14 |
| **Associated Tasks:** | Classification | **Missing Values?** | N/A | **Number of Web Hits:** | 61143 |

**4.2 ATTRIBUTE INFORMATION:**

* 1. Area (A): The area of a bean zone and the number of pixels within its boundaries.  
     2.) Perimeter (P): Bean circumference is defined as the length of its border.  
     3.) Major axis length (L): The distance between the ends of the longest line that can be drawn from a bean.  
     4.) Minor axis length (l): The longest line that can be drawn from the bean while standing perpendicular to the main axis.  
     5.) Aspect ratio (K): Defines the relationship between L and l.  
     6.) Eccentricity (Ec): Eccentricity of the ellipse having the same moments as the region.  
     7.) Convex area (C): Number of pixels in the smallest convex polygon that can contain the area of a bean seed.  
     8.) Equivalent diameter (Ed): The diameter of a circle having the same area as a bean seed area.  
     9.) Extent (Ex): The ratio of the pixels in the bounding box to the bean area.  
     10.)Solidity (S): Also known as convexity. The ratio of the pixels in the convex shell to those found in beans.  
     11.)Roundness (R): Calculated with the following formula: (4piA)/(P^2)  
     12.)Compactness (CO): Measures the roundness of an object: Ed/L  
     13.)ShapeFactor1 (SF1)  
     14.)ShapeFactor2 (SF2)  
     15.)ShapeFactor3 (SF3)  
     16.)ShapeFactor4 (SF4)  
     17.)Class (Seker, Barbunya, Bombay, Cali, Dermosan, Horoz and Sira)

**4.3 ALGORITHM ANALYSIS:**

The following steps are required to build the system

Step 1; Import the libraries that are required to build this model as these libraries have some files that are useful for our system.

Step 2: Load the dataset as we need to read the data to pandas data frame.

Step 3: We need to separate the output value[labels] from the input value in the given dataset by assigning the output values as y and input as x.

Step 4: Need to perform test train split from the sklearn library in order to segregate the training data and validation data

Step 5: Next step is to define the model i.e.. decision tree classifier and then fit the model with the training dataset.

Step 5: Now we have the trained model. Thus we can make predictions using this model and we can check this by making predictions with test data.

Step 6: Can also check the accuracy of this model by comparing the predicted output and actual output of the test data.

Step 7: Should import some files for determining the confusion matrix and classification report.

Step 8: With the help of a classification report, it is easy to find out the accuracy, precision, recall, fi score of this model.

Step 10: This model which I have done has got an accuracy of 89%,92% and 93% for decision tree , logististic regression, and random forest respectively.

Step 10: Can also plot the accuracies and create a beautiful grap and compare the accuracies.

**CHAPTER 5**

**SUMMARY, CONCLUSION AND FUTURE SCOPE**

We have successfully completed our mini-project ‘DRY BEAN CLASSIFICATION ‘in python. We got accuracies like 89%, 92% and 93%. And we can conclude that Now on, we can say what type of dry bean is given.

**Future Enhancements**

* Can be provided to farmers to help them classify easily.
* Can use computer vision in the future, through that we can classify it through cameras that can be used in manufacturing plants which results in less time consumption for segregation of bean types while packing.
* Can make this as a giant seed classifying system that can classify any type of seed, whether it is an apple seed or any seed.
* Some upgrades can be done to this system via checking with other algorithms such as Support Vector Machine (SVM),and so on,

All good things must come to an end, so does our DRY BEAN CLASSIFICATION SYSTEM.

**APPENDIX**

**CODING**

**DECISION TREE CLASSIFIER**

#importing all the modules as required

Import pandas as pd

import matplotlib.pyplot as plt

from sklearn import tree

from sklearn.tree import DecisionTreeClassifier

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

#extracting the dataset

data\_set = pd.read\_excel("C:/Users/Admin/Desktop/New folder/DryBeanDataset/Dry\_Bean\_Dataset.xlsx")

data\_set.head()

# For y, We assigned the target variable[what we are going to predict].

y = data\_set['Class']

# For feature inputs, except label [Class]

data\_set.drop('Class', axis=1, inplace=True)

# Split train data into train and validation data.

x\_train, x\_valid, y\_train, y\_valid = train\_test\_split(data\_set, y, test\_size=0.2, shuffle=True)

# defining the mode

model = DecisionTreeClassifier()

# Fitting the model with the training dataset.

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

# Making predictions with test data.

predictions = model.predict(x\_valid)

# importing the required module for confusion matrix and classification\_report.

from sklearn.metrics import classification\_report

from sklearn.metrics import confusion\_matrix

print(classification\_report(y\_valid, predictions))

# for checking the accuracy, recall, precision etc.

print(confusion\_matrix(y\_valid, predictions))

# To Reduce Complexity Take Max\_Depth As 2

model2 = DecisionTreeClassifier(max\_depth = 2)

model2.fit(x\_train, y\_train)

# Plotting a tree

plt.figure(figsize = (20,12))

tree.plot\_tree(model2,fontsize = 15)

plt.show()

**RANDOM FOREST CLASSIFIER**

#importing all the modules as required

Import pandas as pd

from sklearn import tree

from sklearn.ensemble import RandomForestClassifier

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

#extracting the dataset

data\_set = pd.read\_excel("C:/Users/Admin/Desktop/New folder/DryBeanDataset/Dry\_Bean\_Dataset.xlsx")

data\_set.head()

# For y, We assigned the target variable[what we are going to predict].

y = data\_set['Class']

# For feature inputs, except label [Class]

data\_set.drop('Class', axis=1, inplace=True)

# Split train data into train and validation data.

x\_train, x\_valid, y\_train, y\_valid = train\_test\_split(data\_set, y, test\_size=0.2, shuffle=True)

# defining the mode

model = RandomForestClassifier()

# Fitting the model with the training dataset.

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

# Making predictions with test data.

predictions = model.predict(x\_valid)

# importing the required module for confusion matrix and classification\_report.

from sklearn.metrics import classification\_report

from sklearn.metrics import confusion\_matrix

print(classification\_report(y\_valid, predictions))

# for checking the accuracy, recall, precision etc.

print(confusion\_matrix(y\_valid, predictions))

**LOGISTIC REGRESSION**

#importing all the modules as required

Import pandas as pd

from sklearn import tree

from sklearn

from sklearn.linear\_model import LogisticRegression

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

#extracting the dataset

data\_set = pd.read\_excel("C:/Users/Admin/Desktop/New folder/DryBeanDataset/Dry\_Bean\_Dataset.xlsx")

data\_set.head()

# For y, We assigned the target variable[what we are going to predict].

y = data\_set['Class']

# For feature inputs, except label [Class]

data\_set.drop('Class', axis=1, inplace=True)

# Split train data into train and validation data.

x\_train, x\_valid, y\_train, y\_valid = train\_test\_split(data\_set, y, test\_size=0.2, shuffle=True)

# defining the mode

model = LogisticRegression(max\_iter=100000)

# Fitting the model with the training dataset.

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

# Making predictions with test data.

predictions = model.predict(x\_valid)

# importing the required module for confusion matrix and classification\_report.

from sklearn.metrics import classification\_report

from sklearn.metrics import confusion\_matrix

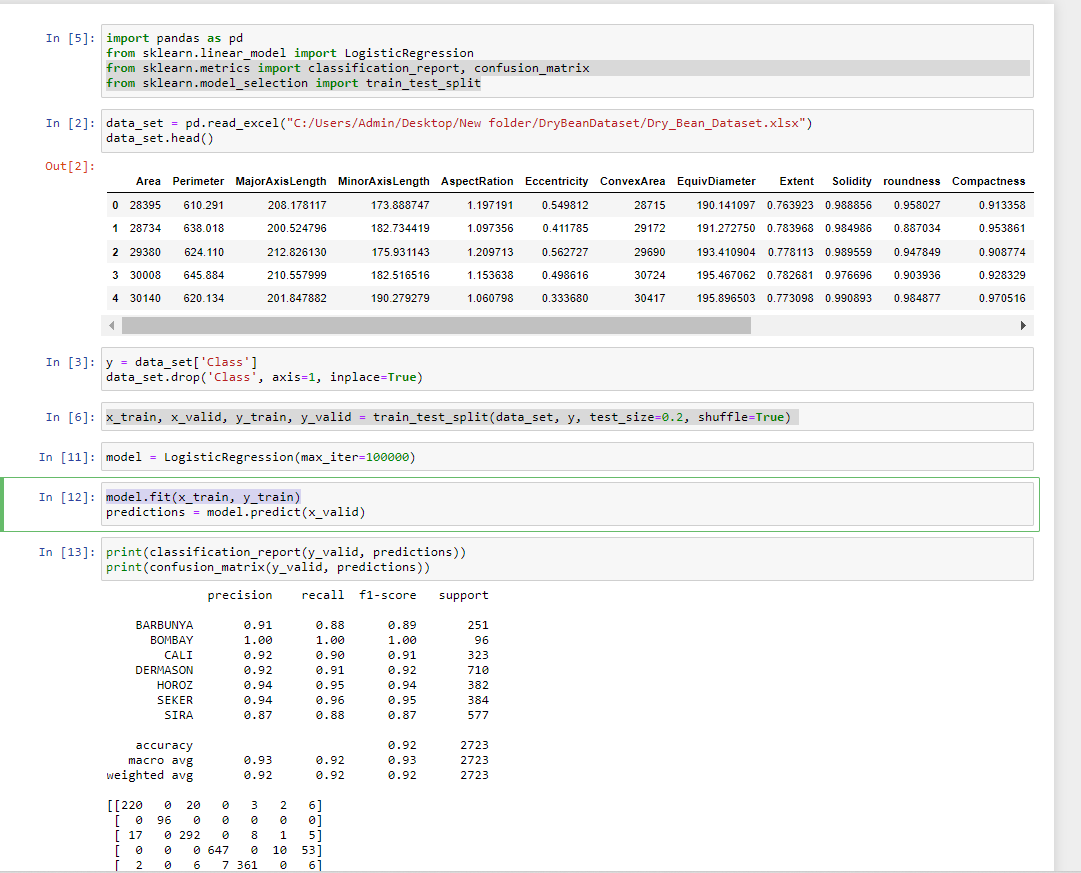
print(classification\_report(y\_valid, predictions))

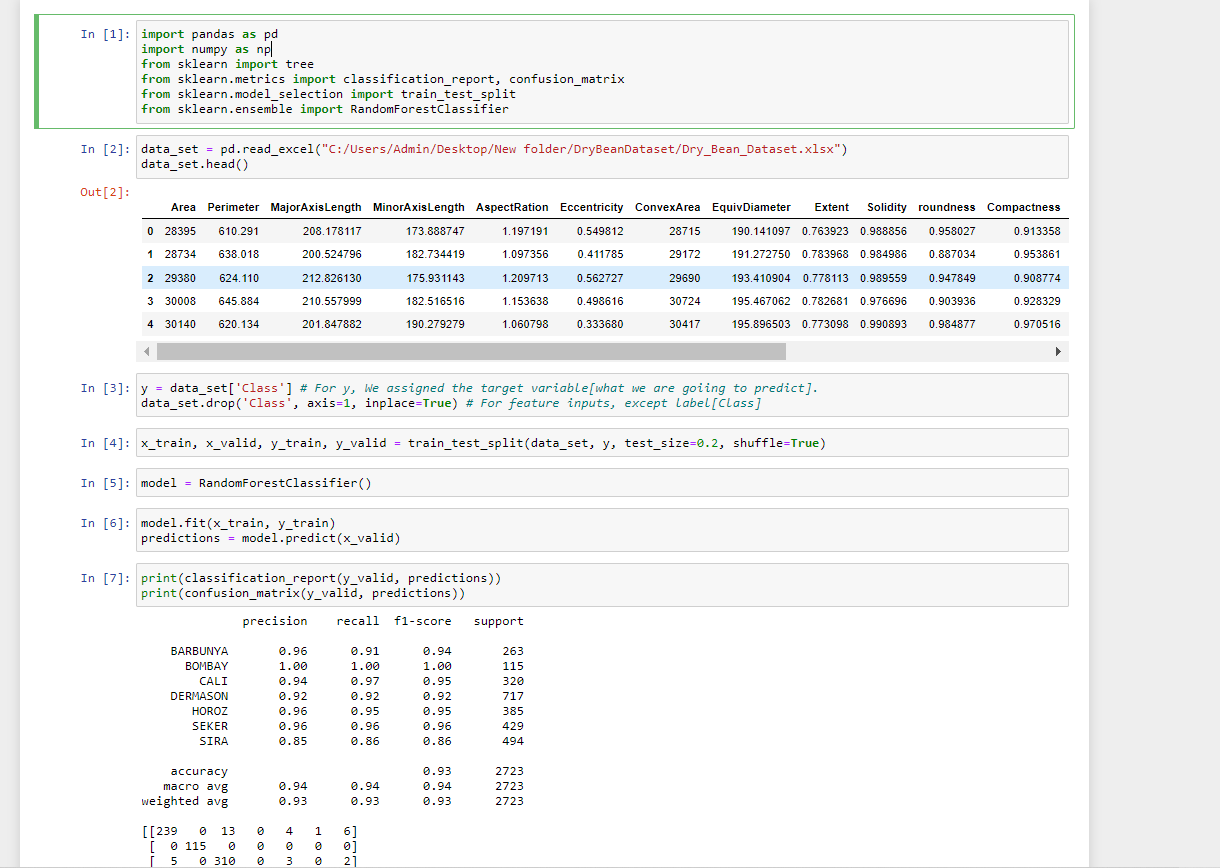
# for checking the accuracy, recall, precision etc.

print(confusion\_matrix(y\_valid, predictions))

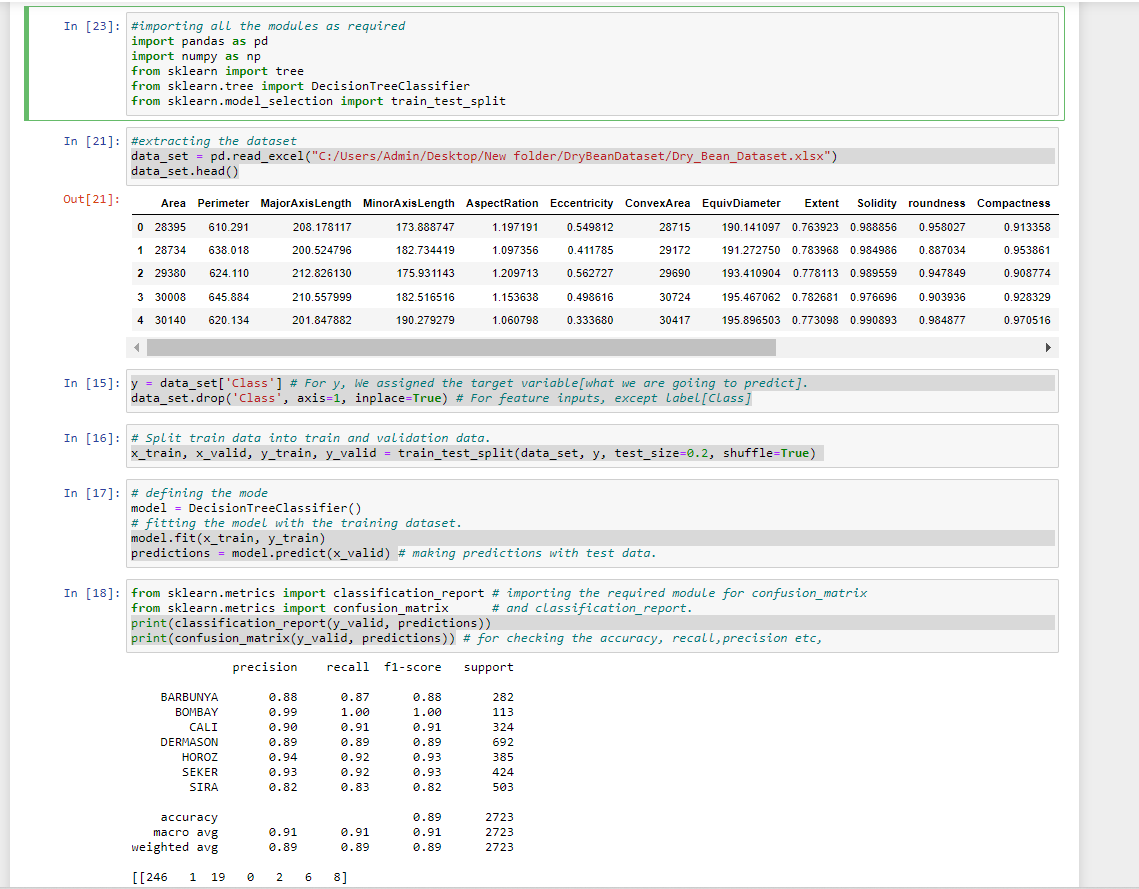
**B.SOURCE CODE**

**Fig. 5.1 Logistic regression source code screenshot**

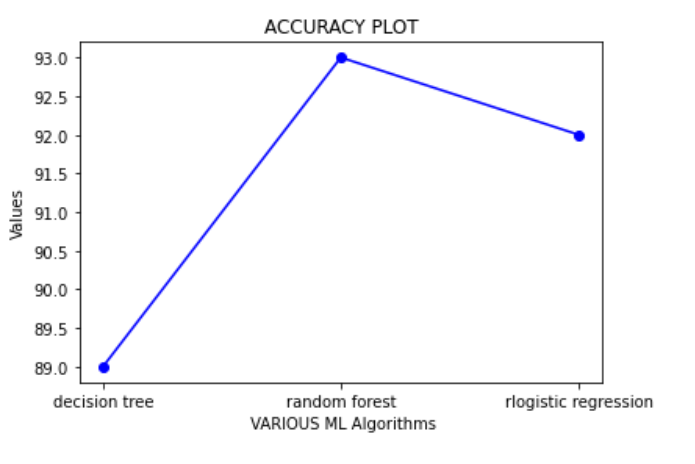




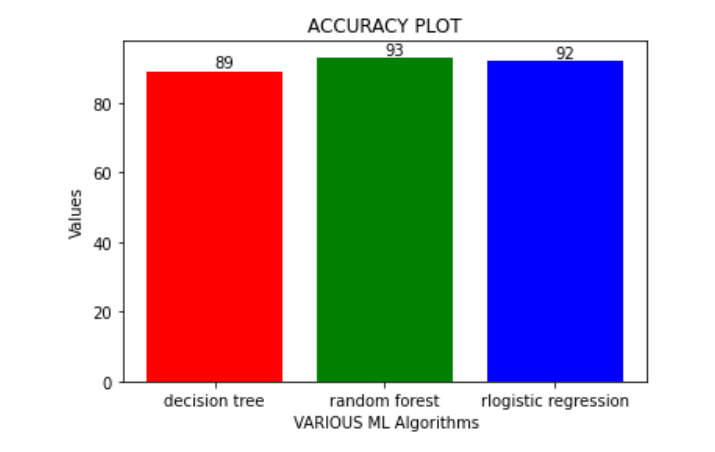
**Fig . 5.2 Random forest Classifier source code Screen shot**

****

**Fig. 5.3 Decision Tree Classifier source code Screen shot**

****

**Fig. 5.4 Accuracy Line Plot**

****

**Fig. 5. 5 Accuracy Bar Plot**

**Diagram

Description automatically generated**

**Fig.5.6 Plotting the decision tree**

**REFERENCES**

1]KOKLU, M. and OZKAN, I.A., (2020), â€œMulticlass Classification of Dry Beans Using Computer Vision and Machine Learning Techniques.â€ Computers and Electronics in Agriculture, 174, 105507.  
DOI: [[Web Link]](https://doi.org/10.1016/j.compag.2020.105507)

2] A. Vibhute, S. Bodhe Computer Science (2012) Applications of Image Processing in Agriculture:

3] Jun Sun, Shuying Jiang, H. Mao, Xiaohong Wu, Qinglin Li Materials Science (2016) Classification of Black Beans Using Visible and Near Infrared Hyperspectral Imaging.

4] T. Mitchell, ―Machine Learning‖, McGraw Hill, 1997.

5] Quinlan, J.R, ―Induction of decision trees‖. Journal of Machine Learning 1(1986) 81-106.

6] 4] Breiman, Friedman, Olshen, and Stone. ―Classification and Regression Trees‖, Wadsworth, 1984., Mezzovico, Switzerland.

7] ] My Chau Tu, Dongil Shin, Dongkyoo Shin, ―A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms‖, DASC '09 Proceedings of the 2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, IEEE Computer Society Washington, DC, USA ©2009