IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 7, July 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007029 |

Classification of Dry Beans Using Machine Learning

Shantanu R. Desai[1], Mrs. Shilpa K. C[2]

B.E Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India[1]

Assistant Professor, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India[2]

ABSTRACT: From different parts of the world, you can find many different kinds of dry beans. Now, it's hard work to put them into groups based on their different properties. But classifying them is very important, especially when it comes to selling and renting them to people who are interested. By taking pictures of the dry beans, we hope to solve this problem of putting them into different groups. When the data from these images was taken out, 16 features or attributes and 4 shapes were found. ML models were used to test these features, and the results were different accuracy scores. The bean varieties were further sorted by the model that gave them the highest score into the following groups: Seker, Barbunya, Bombay, Cali, Dermosan, Horoz, and Sira. These results will be a big help to the people who needed them and to the shop owners who sell dry beans every day.

.

KEYWORDS: Classification Seker, Barbunya, Bombay, Cali, Dermosan, Horoz and Sira.

I. INTRODUCTION

It is a branch of AI (Artificial Intelligence) called machine learning, which focuses on designing systems that can learn from data and improve their accuracy over time without being taught.

An algorithm consists of a series of statistical processing operations. In machine learning, algorithms are taught to detect patterns and characteristics in vast volumes of data in order to make predictions and judgments. The most critical stage is to choose the best algorithm since the accuracy of the predictions or judgments will be determined by this.

Steps in the training process of an artificial intelligence (AI) model may be summarised as follows:

1. Selecting and preparing thedata
2. Choosing a suitable algorithm to run thedataset
3. Training the algorithm to create amodel
4. Using and improving themodel

Different ML techniques will be utilised in this work to estimate the model's accuracy, as previously indicated. Once the dried bean dataset has been classified based on different attributes, this is done. In addition, a high degree of accuracy was achieved by taking into account a number of other factors.

II. PROBLEM STATEMENT

Develop a machine learning model to distinguish between seven different registered Varieties of dry beans with similar features in order to obtain uniform seed classification.

Objective

- Acquiring and cleaning of data.
- To classify the dry beans based on the various features given.
- This will further be of use to the concerned department or authorities to determine the factors based on which the dry beans are classified into different classes and its significance. They can accordingly set the prices of each class of dry beans, make marketing plots and strategies. The model can also help the concerned authorities understand or analyze the factors of pricing affecting the new market.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007029 |

III. LITERATURE REVIEW

Classification of plant diseases using deep learning methods is the focus of this section. For the categorization of olive peacock spot and aculeus olearius illnesses, Uuz et al. [7] developed a deep convolutional neural network (DCNN). Olive peacock spot and aculeus olearius disease were found in 3400 samples of olive leaves from healthy plants. The DCNN-based VGG16 and VGG19 architectures are in use. The RMS Prop optimization technique and Stochastic gradient descent helped to increase the network's performance. Convolutional neural networks (CNN)-based multi-task systems were proposed by Esgario et al. [8] to identify the intensity of stress produced by biotic agents on coffee leaves. Computational experiments led to the development of a detection system that is both more precise and more resilient. Using a multi-task system-based CNN with ResNet50 architecture, we were able to achieve a severity estimate accuracy of 86.51 percent and a biotic stress classification accuracy of 95.24 percent. Thus, this technique is suitable for identifying and quantifying biotic pressures in coffee plantations. Esgario et al. [9] developed a pre-trained Convolutional Neural Network (CNN) for agricultural disease diagnosis. CNN models like DenseNet201, ResNet101, VGG19, Visual Geometry Group 16 (VGG16), and AlexNet are included in this pre-trained network. Leafhopper, yellow vein mosaic virus, citrus canker, citrus Hindu mite, brown spot, Cercospora leaf spot, two-spotted spider mite, Epilachna beetle, tiny leaf disease, and Tobacco Mosaic Virus TMV were all included in the author's list of 10 illnesses that affect four crops. In real time, they analysed the prediction scores and categorised photos for each illness type. For the identification of plant diseases, Hernández et al. [10] developed a Bayesian Deep Learning approach. Bayesian inference improves classification performance above current best practises. This plant disease detection problem's posterior density was estimated using the predictions' uncertainty, which could be calculated.

IV. RESULTS AND DISCUSSION

The data of dry beans is studied by analyzing the various features given in the dataset. The dry beans are classified into 7 different classes, label encoded as:

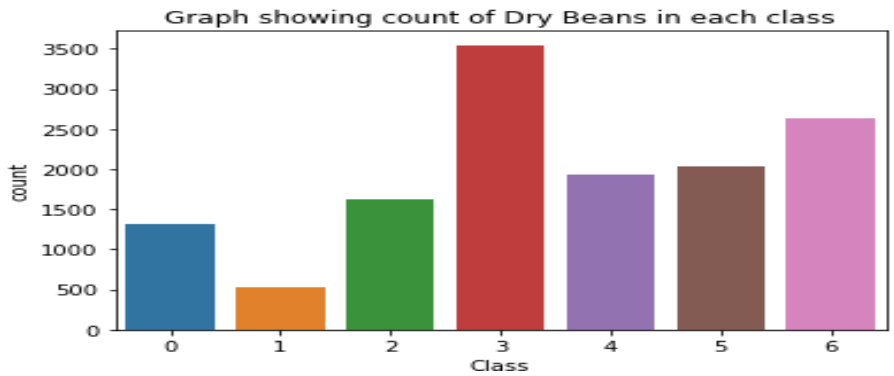
1. BARBUNYA
2. BOMBAY
3. CALI
4. DERMASON
5. HOROZ
6. SEKER
7. SIRA

Below is the plot of the number of dry beans in each class.

Count Plot

fig : Count of dry beans in each class

```
sns.countplot(x=d.Class)
#plt.ylabel='Count'
plt.title('Graph showing count of Dry Beans in each class')
plt.show()
```



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

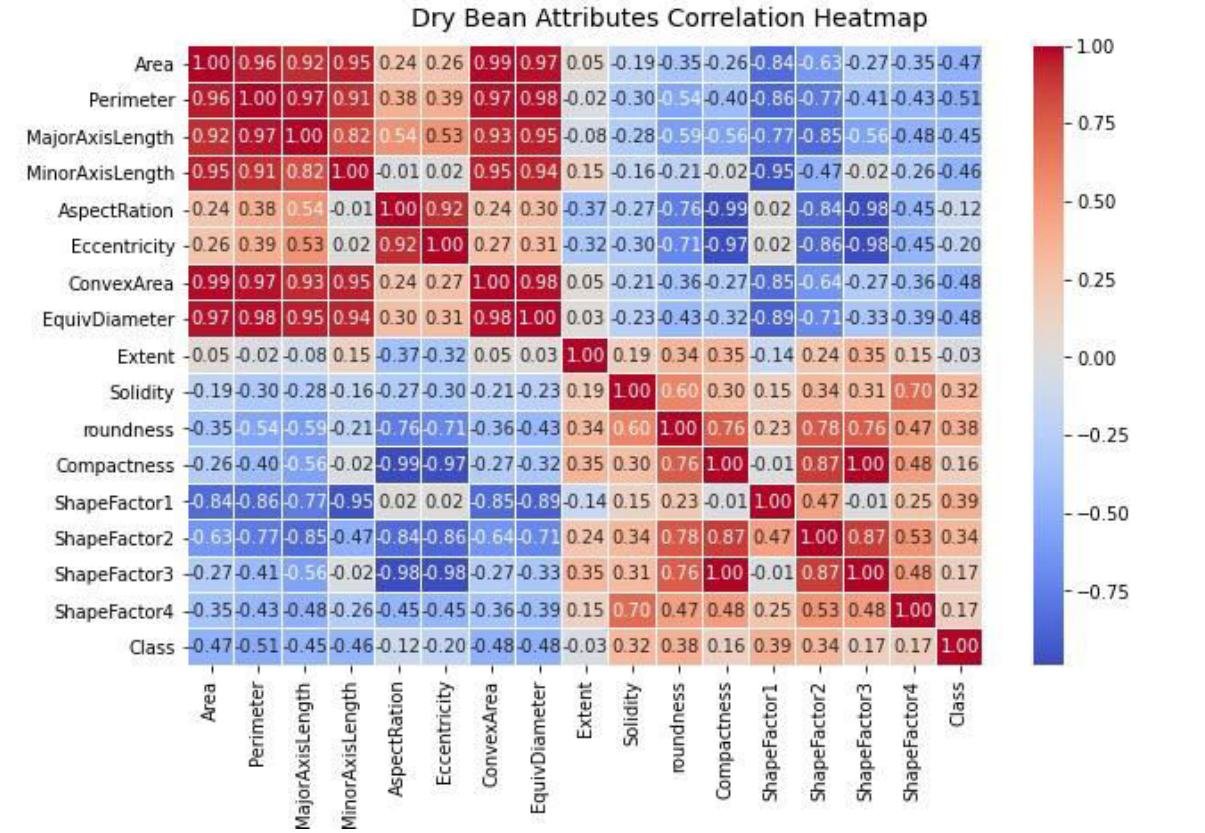
|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007029 |

- It can be observed that the class 3, “DERMASON” has the highest number of dry beans, slightly above3500.
- The class 1, “BOMBAY” has the least number of dry beans, slightly above500.

HeatMap

```
f, ax = plt.subplots(figsize=(10, 6))
corr = d.corr()
hm = sns.heatmap(round(corr,2), annot=True, ax=ax, cmap="coolwarm",fmt='.2f',
                 linewidths=.05)
f.subplots_adjust(top=0.93)
t= f.suptitle('Dry Bean Attributes Correlation Heatmap', fontsize=14)
```



- Each square in the plot shows the correlation between the variables on eachaxis. Correlation ranges from -1 to+1.
- The values close to +1 means that they are more POSITIVELY correlated.
- It can be observed from the plot that, Shape Factor 3 and Compactness arehighlyCorrelated implying there is a linear trend between them.
- Eccentricity has a correlation of 0.92 with aspect ratio. It is morepositively correlated and directlyproportional.
- Aspect ratio and Eccentricity have good correlation of0.92.
- Convex area and EquivDiameter have a high correlation with Area,Perimeter, Major and Minor Axis Length.(>0.90)
- The features in the upper left triangle seem to have a goodcorrelation.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007029 |

Preparing the Machine Learning models:
This data is multivariate and linear. This requires CLASSIFICATION to be done.

- **Splitting the data for training andtesting:**

```
from sklearn.model_selection import train_test_split

x = df.drop('Class',axis=1)
y = df.Class
X_train, X_test, y_train, y_test = train_test_split(x,y,test_size=0.3,random_state=42)
```

fig*: *heatmap to find correlation

- The data is split accordingly to two variables, namely x and y, in order to be further split for training and testing of the machine learningmodels.
- After the process of splitting the data, the different Classification algorithms used in this project are trained and tested foraccuracy.

V. CONCLUSION

As seen earlier, different algorithms and models were used to arrive at the accuracy rate that is achieved through this task. 2 different ML algorithms to inspect each and every aspect that would give some information about the dry beans. Each dry bean variety has been classified depending on various factors and plots were created of the same. Each model was trained in such way that, any input given in correlation with the already existing attributes, the model was able to give a score which made it easy to compare the various dry beans from the given data set. The task is identified as classification and various ML models like Support Vector Classifier and Random Forest Classifier, were Used and the algorithm with the best accuracy score is identified. The outcomes of the models used are noted and compared in order for the selection of the best model suited for this task. Therefore, this task was implemented to classify the dry bean varieties with the help of the dataset set that was made available. The best suited model is identified and classification of dry bean task can be relatively easily accomplished.

REFERENCES

- **Journals**

[1] Mengyu Huang, “Theory and Implementation of linear regression”,IEEE, Chongqing, China, Jul 2020.

- **Links**

[2] https://en.wikipedia.org/wiki/Machine_learning
[3] http://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/
[4] https://www.javatpoint.com/applications-of-machine-learning
[5] https://en.wikipedia.org/wiki/Anaconda_(Python_distribution)
[6] https://towardsdatascience.com/everything-you-need-to-know-about-jupyter-Notebooks-10770719952b
[7] https://www.geeksforgeeks.org/ml-linear-regression/

- **Textbook**

[8] Alberto Boschetti, Luca Massaron,” Python Data Science Essentials - Third Edition”, Packt Publishing, ISBN: 978178953786
[9] Al Sweigart,“Automate the Boring Stuff with Python-1st Edition”, No Starch Press, 2015.ISBN: 978-1593275990

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details