Rice Classification

Presented By:

-Aditi Bhalawat
-Krish Lodha
-Lavni Singhal
-Mahi Kulshresth

Problem Statement

The task at hand involves classifying rice into two types, ‘Cammeo’ and ‘Osmancik’, based on their distinct physical geometric characteristics. This aims to leverage these features effectively for accurate identification and sorting of the rice grains according to their respective types.

What Is My DataSet?

-The dataset includes 3810 images of rice grains from these two varieties, each processed to extract seven key morphological features.

-This dataset is crucial for the characterization and comparison of rice varieties, supporting advancements in rice cultivation practices.

Features Present In The Dataset

1.**Area:** Number of pixels within the boundaries of a rice grain.

2.Perimeter: Circumference calculated by measuring the distance between pixels around the grain's boundaries.

3.Major Axis Length: Longest line that can be drawn through the rice grain.

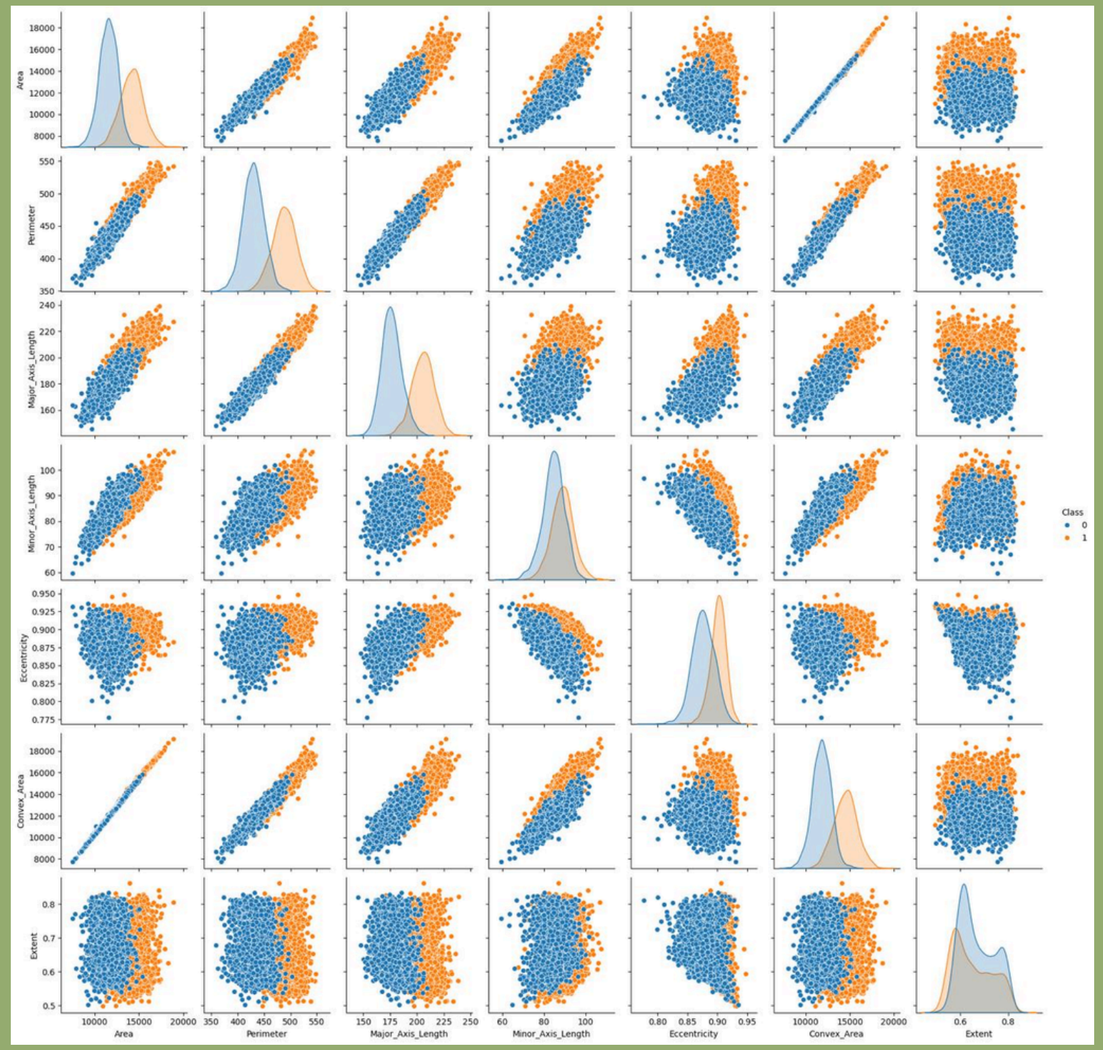
4.Minor Axis Length: Shortest line that can be drawn across the rice grain.

5.Eccentricity: Measures the roundness of the grain compared to an ellipse with the same moments.

6.Convex Area: Pixel count of the smallest convex shell that encompasses the grain.

7.Extent: Ratio of the rice grain area to the bounding box area.

Scatter Plots

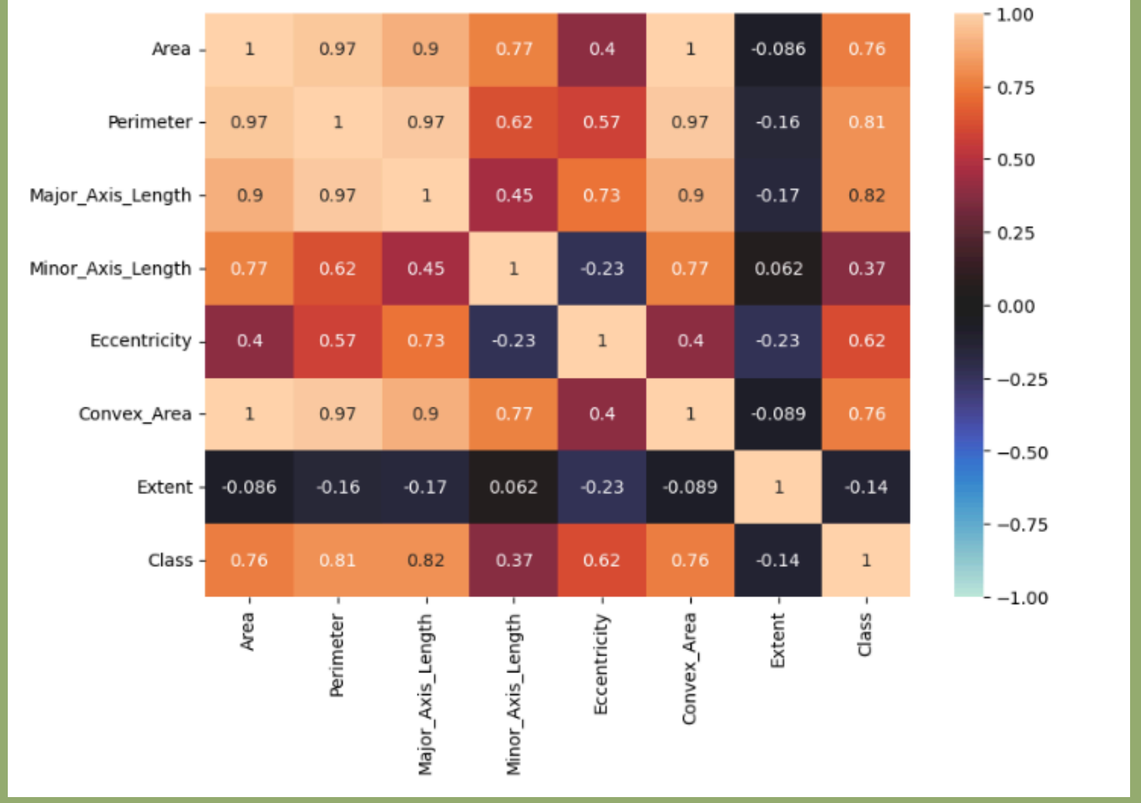


Overview Of The Dataset

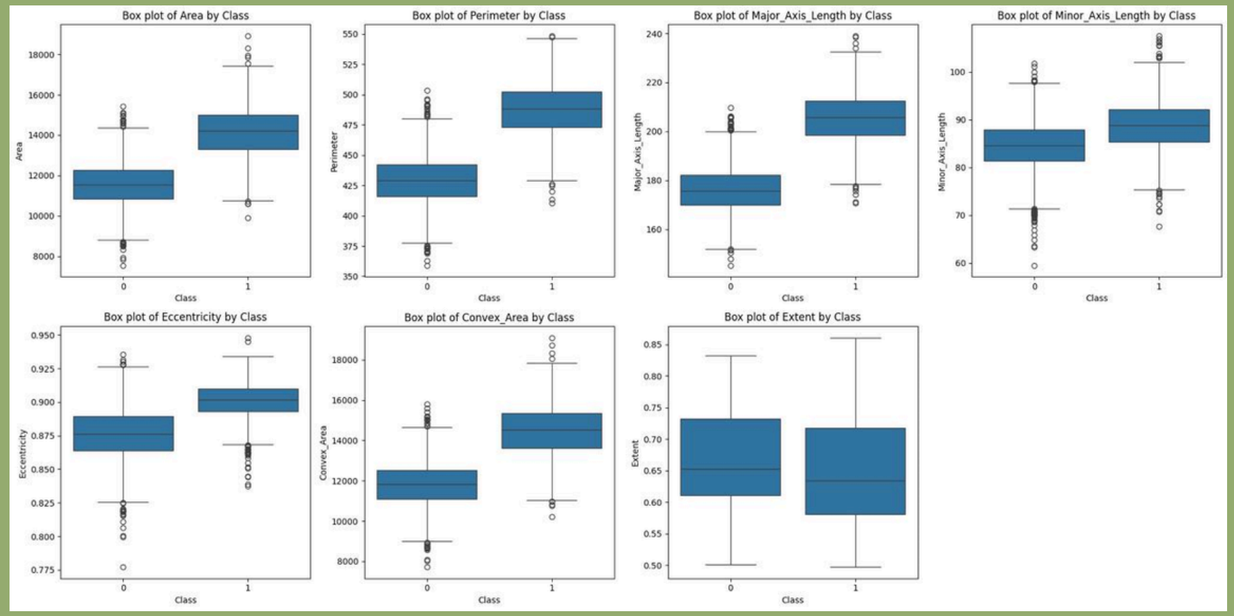
	Area	Perimeter	Major_Axis_Length	Minor_Axis_Length	Eccentricity	Convex_Area	Extent	Class
0	15231	525.578979	229.749878	85.093788	0.928882	15617	0.572896	Cammeo
1	14656	494.311005	206.020065	91.730972	0.895405	15072	0.615436	Cammeo
2	14634	501.122009	214.106781	87.768288	0.912118	14954	0.693259	Cammeo
3	13176	458.342987	193.337387	87.448395	0.891861	13368	0.640669	Cammeo
4	14688	507.166992	211.743378	89.312454	0.906691	15262	0.646024	Cammeo

Correlation Heatmap

```
corr = rice.corr('spearman')
plt.figure(figsize=(9,6))
sns.heatmap(corr, vmin=-1, center=0, vmax=1, annot=True);
```



BOX-PLOTS



MODEL USED

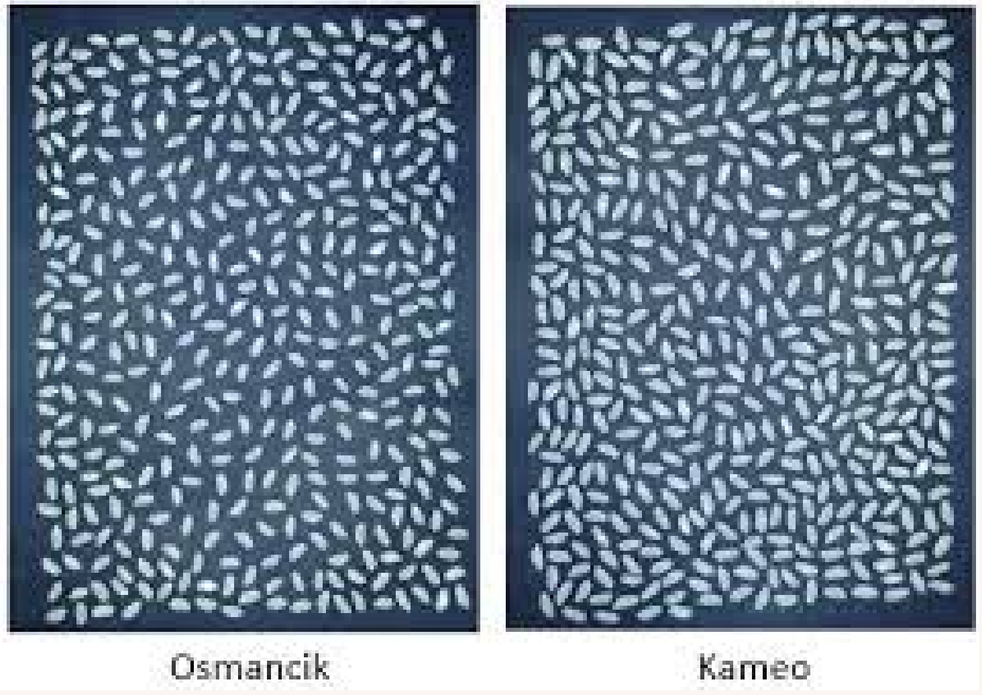
Random Forest Classifier

-Ensemble Learning: Random Forest combines multiple decision trees to improve accuracy and robustness in classifying rice grain types based on diverse features.

- Outlier Robustness: By aggregating predictions from multiple trees, Random Forest inherently reduces the impact of outliers in the dataset, ensuring more reliable classification results.

- Feature Importance: It identifies significant features affecting rice grain classification, aiding in understanding key factors like area, perimeter, and shape metrics.

Random Forest is ideal for our dataset as it efficiently handles outliers, enhances classification accuracy through ensemble learning, and provides insights into feature relevance critical for accurate rice grain classification.



Performance of Model

Accuracy

Measures the overall correctness of the model's predictions, indicating it correctly identifies 93.18% of rice grain types.

Precision

Reflects the proportion of correctly predicted positive instances among all predicted positive instances, showing 94.12% precision in rice grain classification.

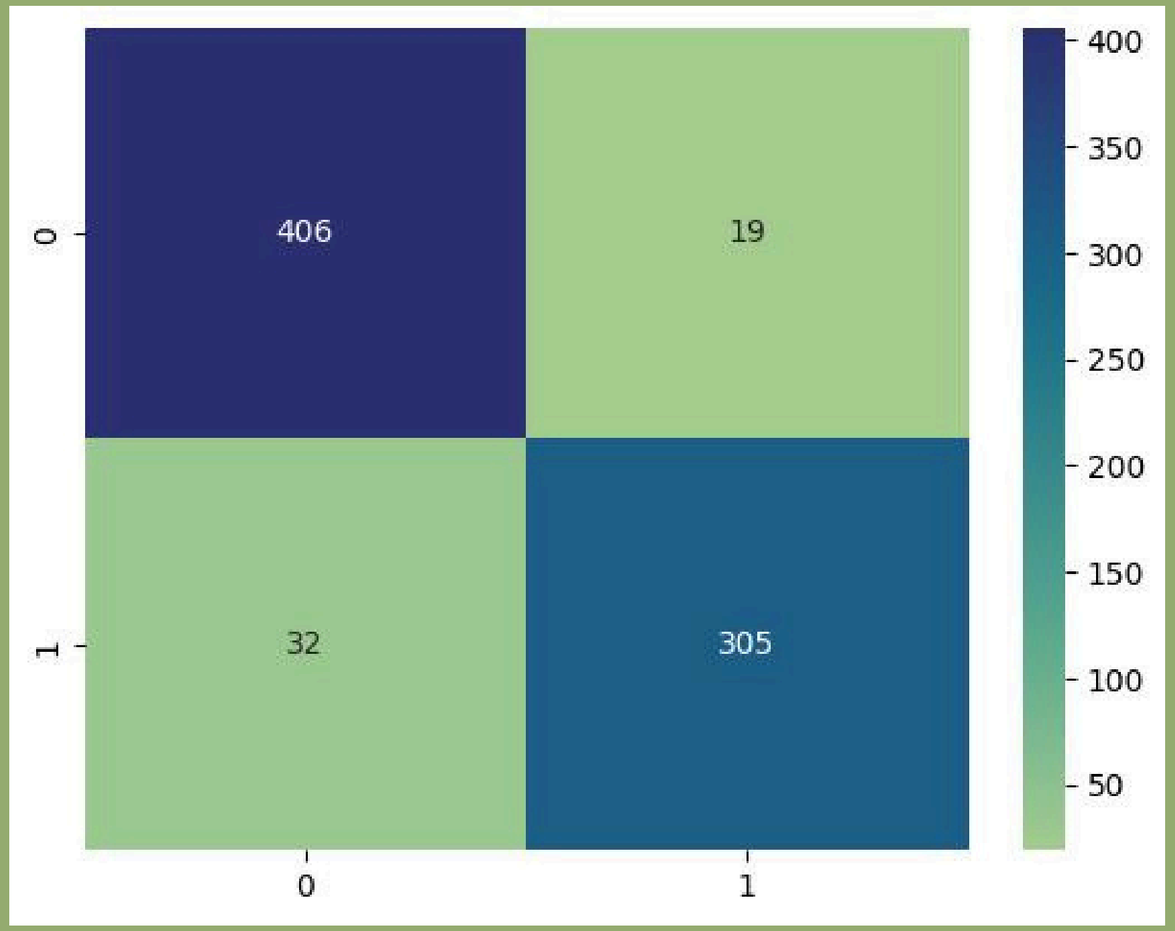
F1 Score

Harmonic mean of precision and recall, providing a balanced measure of model performance with an F1 score of 92.12%.

Recall

Indicates the proportion of actual positive instances correctly predicted by the model, achieving a recall rate of 90.21% for rice grain types.

Confusion Matrix



Shortcomings of our approach and how we would like to address them

***FEATURE CORRELATION AND REDUNDANCY :**

-The major issue faced was the highly correlated or redundant features can lead to decreased model performance and interpretability, as the model may give undue importance to these features.

***OUTLIER DETECTION:**

-The process for finding out the outliers in the dataset was also a bit cumbersome and tough but we tackled it with the help of the IQR or Inter Quartile Range.

Thank you very much!