

Day 56

DIY

Q1. Problem Statement: Boosting Models with AdaBoost

The 'seeds.csv' dataset contains the data about the wheat seeds, the 'Type' column consist of three unique values, 1, 2, 3, which are classified based on the characteristics of seeds entailing in other columns.

Load the 'seeds.csv' dataset into a DataFrame and considering the 'Type' column as target, perform the following tasks:

1. Analyze the target column by printing it's unique values
2. Separate the feature vectors and the target variable
3. Split the dataset into train and test sets in a 70:30 ratio
4. Build a Decision Tree Classifier and a GaussianNB model and print their accuracy scores
5. For the Decision Tree Classifier and a GaussianNB models boost the accuracy using AdaBoost Classifier and compare the accuracy scores with original models using a bar plot

Dataset:

	Area	Perimeter	Compactness	Kernel.Length	Kernel.Width	Asymmetry.Coeff	Kernel.Groove	Type
0	15.26	14.84	0.8710	5.763	3.312	2.221	5.220	1
1	14.88	14.57	0.8811	5.554	3.333	1.018	4.956	1
2	14.29	14.09	0.9050	5.291	3.337	2.699	4.825	1
3	13.84	13.94	0.8955	5.324	3.379	2.259	4.805	1
4	16.14	14.99	0.9034	5.658	3.562	1.355	5.175	1

Sample Output:

1. Considering the 'Type' column as target, analyze the target column by printing the unique values

	Counts	Percentage
2	68	0.341709
1	66	0.331658
3	65	0.326633

2. Separate the feature vectors and the target variable

	Area	Perimeter	Compactness	Kernel.Length	Kernel.Width	Asymmetry.Coeff	Kernel.Groove
0	15.26	14.84	0.8710	5.763	3.312	2.221	5.220
1	14.88	14.57	0.8811	5.554	3.333	1.018	4.956
2	14.29	14.09	0.9050	5.291	3.337	2.699	4.825
3	13.84	13.94	0.8955	5.324	3.379	2.259	4.805
4	16.14	14.99	0.9034	5.658	3.562	1.355	5.175

```
0    1
1    1
2    1
3    1
4    1
Name: Type, dtype: int64
```

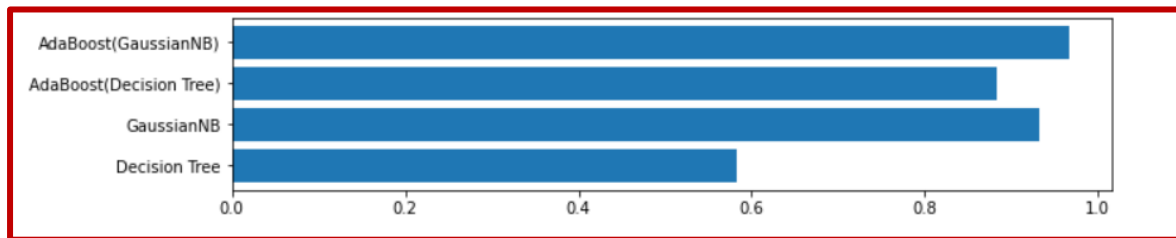
3. Split the dataset into train and test sets in a 70:30 ratio
4. Build a Decision Tree Classifier and a GaussianNB model and print their accuracy scores

```
Accuracy score of the Decision tree model is:
0.583
```

```
Accuracy score of the GaussianNB model is:
0.933
```

5. For the Decision Tree Classifier and a GaussianNB models boost the accuracy using AdaBoost Classifier and compare the accuracy scores

with original models using a par plot



edureka!
a Veranda Enterprise