

CSE 6363 - MACHINE LEARNING
ASSIGNMENT 3 REPORT

SUBMITTED BY: ASFIYA MISBA (1002028239)

DECISION TREE

Implemented decision tree model for the titanic dataset with criteria misclassification, entropy and gini. The validation and test accuracy scores are as follows:

Other parameters are kept constant for comparison but can be changed during run-time.

Other parameters were kept as max_depth=5, min_sample_split=3, min_samples_leaf=1.

Criteria	Validation Accuracy	Test Accuracy
misclassification	71.43%	80.92%
gini	76.67%	79.01%
entropy	72.38%	80.92%

SAMPLE OUTPUT

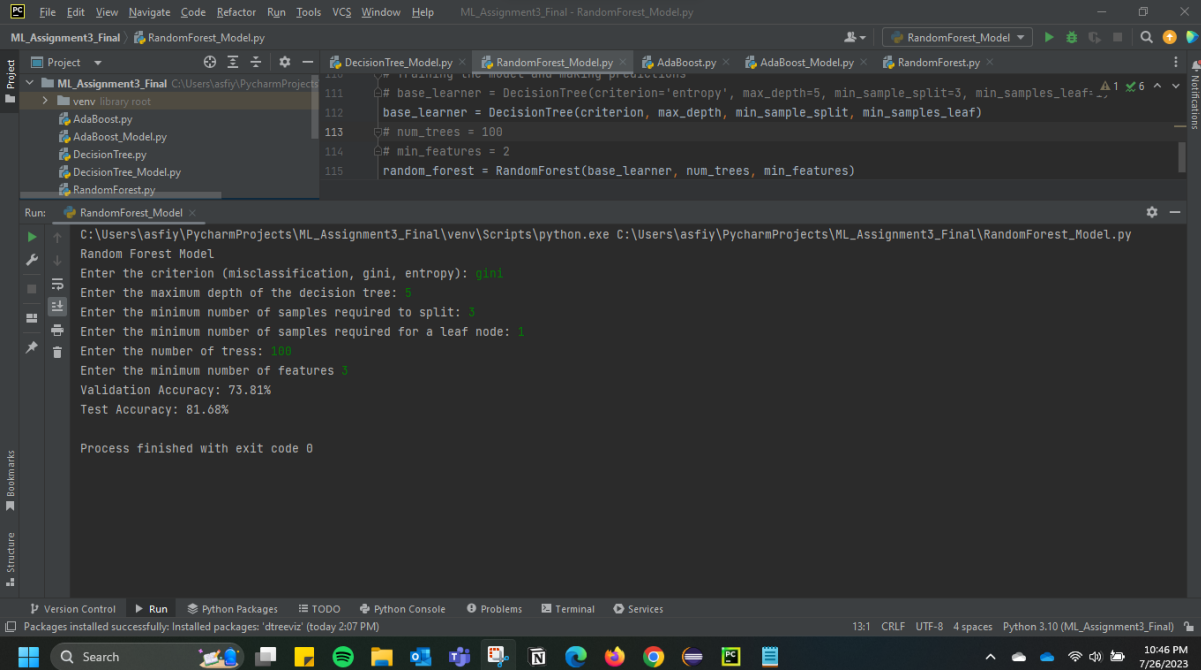
```
File Edit View Navigate Code Refactor Run Tools VCS Window Help ML_Assignment3_Final - DecisionTree_Model.py
ML_Assignment3_Final DecisionTree_Model.py
Project Run: DecisionTree_Model (1)
C:\Users\asfiy\PycharmProjects\ML_Assignment3_Final\venv\Scripts\python.exe C:\Users\asfiy\PycharmProjects\ML_Assignment3_Final\DecisionTree_Model.py
Decision Tree Model
Enter the criterion (misclassification, gini, entropy): gini
Enter the maximum depth of the decision tree: 5
Enter the minimum number of samples required to split: 3
Enter the minimum number of samples required for a leaf node: 1
Validation Accuracy: 71.90%
Test Accuracy: 81.68%
Do you want to see the decision tree? (yes/no): yes
Title:
if Title = 1:
    Pclass:
        if Pclass = 1:
            -> Class: 0
        if Pclass = 2:
            -> Class: 0
        if Pclass = 3:
            -> Class: 0
    if Title = 2:
        FamilySize:
            if FamilySize = 2:
                -> Class: 1
            if FamilySize = 3:
                -> Class: 1
            if FamilySize = 4:
                -> Class: 1
Version Control Run Python Packages TODO Python Console Problems Terminal Services
Packages installed successfully: Installed packages: 'dtreeviz' (today 2:07 PM)
65:1 CRLF UTF-8 4 spaces Python 3.10 (ML_Assignment3_Final)
10:35 PM
7/26/2023
```

RANDOM FOREST

Random forest is implemented using decision tree as the base learner.
The accuracy scores are as follows:

Criteria	Validation Accuracy	Test Accuracy
misclassification	72.86%	80.92%
gini	75.71%	80.15%
entropy	73.81%	81.68%

SAMPLE OUTPUT



The screenshot displays the PyCharm IDE interface. The top pane shows the code for `RandomForest_Model.py`, which includes the following Python code:

```
111 base_learner = DecisionTree(criterion='entropy', max_depth=5, min_sample_split=3, min_samples_leaf=1)
112 base_learner = DecisionTree(criterion, max_depth, min_sample_split, min_samples_leaf)
113 num_trees = 100
114 min_features = 2
115 random_forest = RandomForest(base_learner, num_trees, min_features)
```

The bottom pane shows the output of the program, which prompts the user to enter various parameters and displays the resulting accuracy scores:

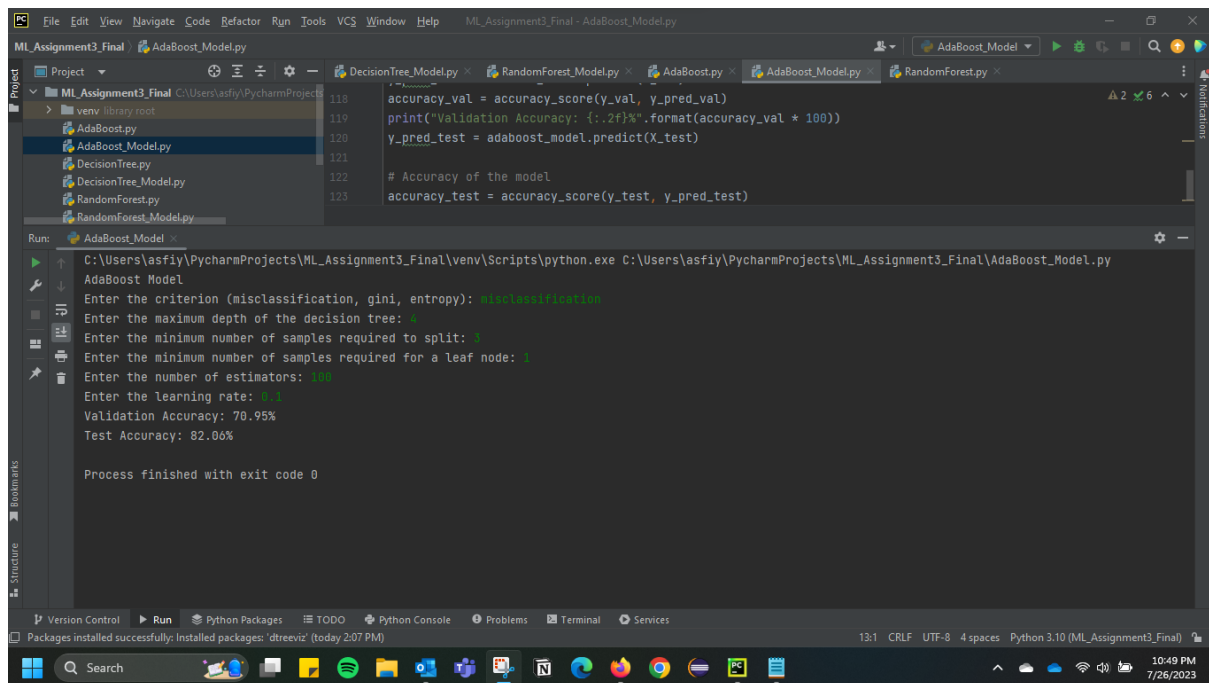
```
Random Forest Model
Enter the criterion (misclassification, gini, entropy): gini
Enter the maximum depth of the decision tree: 5
Enter the minimum number of samples required to split: 3
Enter the minimum number of samples required for a leaf node: 1
Enter the number of trees: 100
Enter the minimum number of features: 2
Validation Accuracy: 73.81%
Test Accuracy: 81.68%
Process finished with exit code 0
```

ADABOOST

Adaboost is implemented using decision tree as the base learner.
The accuracy scores are as follows:

Criteria	Validation Accuracy	Test Accuracy
misclassification	72.86%	78.63%
gini	75.24%	80.92%
entropy	75.24%	80.53%

SAMPLE OUTPUT



```
File Edit View Navigate Code Refactor Run Tools VCS Window Help ML_Assignment3_Final - AdaBoost_Model.py
ML_Assignment3_Final C:\Users\asfiy\PycharmProjects
venv library root
AdaBoost.py
AdaBoost_Model.py
DecisionTree.py
DecisionTree_Model.py
RandomForest.py
RandomForest_Model.py
AdaBoost_Model.py
118 accuracy_val = accuracy_score(y_val, y_pred_val)
119 print("Validation Accuracy: {:.2f}%".format(accuracy_val * 100))
120 y_pred_test = adaboost_model.predict(X_test)
121
122 # Accuracy of the model
123 accuracy_test = accuracy_score(y_test, y_pred_test)

Run: AdaBoost_Model
C:\Users\asfiy\PycharmProjects\ML_Assignment3_Final\venv\Scripts\python.exe C:\Users\asfiy\PycharmProjects\ML_Assignment3_Final\AdaBoost_Model.py
AdaBoost Model
Enter the criterion (misclassification, gini, entropy): misclassification
Enter the maximum depth of the decision tree: 4
Enter the minimum number of samples required to split: 1
Enter the minimum number of samples required for a leaf node: 1
Enter the number of estimators: 100
Enter the learning rate: 0.1
Validation Accuracy: 70.95%
Test Accuracy: 82.06%
Process finished with exit code 0
```