HOPE-AI | CLASSIFIER ASSIGNMENT

CLASSIFICATION REPORT OF SOCIAL NETWORK ADVERTISEMENTS

Confusion Matrix:

PREDICTED		ACTUAL CLASS	
	N-test dataset count	**Purchased**	**Not Purchased**
	Purchased	True Purchased-T(P) (Correctly classified)	False Purchased-F(P) (Wrongly classified)
	Not Purchased	False Not purchased-F(N) (Wrongly classified)	True not purchased-T(N) (Correctly classified)

1. **Random Forest Classification:**

i) **Support: Individual count of the class in test set**

Q: What is the count of the class (purchased),(not purchased) in the classification of the test data set in the classification?

A: Purchased: 22
Not Purchased: 58

ii) **Accuracy: Overall performance of the model**

Q: What is the percentage of the correct classification of both (purchased, not purchased) to the total input of the test set?

A: Accuracy= T(P)+T(N) / T(P)+F(P)+T(N)+F(N)
Accuracy: 0.93 = 93%

iii) **Recall: Correctly classified class**

Q1: What is the percentage of correct classification of (purchased) to the total input of the (purchased) in the test set?

A1: Recall (purchased)=T(P) / T(P)+F(P) = 0.91

Q2: What is the percentage of correct classification of (not purchased) to the total input of the (not purchased) in the test set?

A2: Recall (not purchased)=T(N) / T(N)+F(N) = 0.93

iv) **Precision: Correctly & wrongly classification of class**

Q1: What is the percentage of correct classification of (purchased) to the sum of correctly classified as (purchased) and wrongly classified as (not purchased) in the test set?

A1: Precision (purchased) = T(P) / T(P)+F(N) = 0.83

Q2: What is the percentage of correct classification of () to the sum of correctly classified as () and wrongly classified as () in the test set?

A2: Precision= T(P) / T(P)+F(N) = 0.96

v) **F1-Score/F1-measure: Overall performance of individual class**

If recall is high, Precision is low, by seeing the value of F1-measure, can validate the model performance

Q1: What is the overall performance of the (purchased)?

A1: F1-score (purchased)=2*[recall*precision] / [recall+precision] = 0.87

Q2: What is the overall performance of the (not purchased)?

A2: F1-score (not purchased)=2*[recall*precision] / [recall+precision] = 0.95

vi) **Macro-average: Overall average**

Q: What is the average performance of precision (correctly & wrongly classified)

A: Recall = R(P)+R(N) / 2 = 0.92

Precision = P(P)+P(N) / 2 = 0.90

F1-score = F1(P)+F2(N) / 2 = 0.91

vii) **Weighted-average:**

Q: What is the sum of the product of proportion rate (weight) of each class?

A: Recall: R(P)*[total count of () in test set / total count of test set] + R(N) = 0.93

Precision : R(P)*[total count of () in test set / total count of test set] + R(N) = 0.93

F1-score : R(P)*[total count of () in test set / total count of test set] + R(N) = 0.93

<u>Random Forest Classification report Tabulation:</u>

		precision	recall	f1-score	support
	0	0.96	0.93	0.95	58
	1	0.83	0.91	0.87	22
	accuracy	0.93			80
	macro avg	0.90	0.92	0.91	80
	weighted avg	0.93	0.93	0.93	80

2. **Decision Tree Classification:**

i) **Support: Individual count of the class in test set**

Q: What is the count of the class (purchased),(not purchased) in the classification of the test data set in the classification?

A: Purchased: 22
Not Purchased: 58

ii) **Accuracy: Overall performance of the model**

Q: What is the percentage of the correct classification of both (purchased, not purchased) to the total input of the test set?

A: Accuracy= T(P)+T(N) / T(P)+F(P)+T(N)+F(N)
Accuracy: 0.91 = 91%

iii) **Recall: Correctly classified class**

Q1: What is the percentage of correct classification of (purchased) to the total input of the (purchased) in the test set?

A1: Recall (purchased)=T(P) / T(P)+F(P) = 0.86

Q2: What is the percentage of correct classification of (not purchased) to the total input of the (not purchased) in the test set?

A2: Recall (not purchased)=T(N) / T(N)+F(N) = 0.93

iv) **Precision: Correctly & wrongly classification of class**

Q1: What is the percentage of correct classification of (purchased) to the sum of correctly classified as (purchased) and wrongly classified as (not purchased) in the test set?

A1: Precision (purchased) = T(P) / T(P)+F(N) = 0.83

Q2: What is the percentage of correct classification of () to the sum of correctly classified as () and wrongly classified as () in the test set?

A2: Precision= T(P) / T(P)+F(N) = 0.95

v) **F1-Score/F1-measure: Overall performance of individual class**

If recall is high, Precision is low, by seeing the value of F1-measure, can validate the model performance

Q1: What is the overall performance of the (purchased)?

A1: F1-score (purchased)=2*[recall*precision] / [recall+precision] = 0.84

Q2: What is the overall performance of the (not purchased)?

A2: F1-score (not purchased)=2*[recall*precision] / [recall+precision] = 0.94

vi) **Macro-average: Overall average**

Q: What is the average performance of precision (correctly & wrongly classified)

A: Recall = R(P)+R(N) / 2 = 0.89
Precision = P(P)+P(N) / 2 = 0.90
F1-score = F1(P)+F2(N) / 2 = 0.89

vii) **Weighted-average:**

Q: What is the sum of the product of proportion rate (weight) of each class?

A: Recall: R(P)*[total count of () in test set / total count of test set] + R(N) = 0.91

Precision : R(P)*[total count of () in test set / total count of test set] + R(N) = 0.91

F1-score : R(P)*[total count of () in test set / total count of test set] + R(N) = 0.91

Decision Tree Classification report Tabulation:

		precision	recall	f1-score	support
	0	0.95	0.93	0.94	58
	1	0.83	0.86	0.84	22
	accuracy	0.91			80
	macro avg	0.89	0.90	0.89	80
	weighted avg	0.91	0.91	0.91	80

3. **Support Vector Machine:**

i) **Support: Individual count of the class in test set**

Q: What is the count of the class (purchased),(not purchased) in the classification of the test data set in the classification?

A: Purchased: 22
Not Purchased: 58

ii) **Accuracy: Overall performance of the model**

Q: What is the percentage of the correct classification of both (purchased, not purchased) to the total input of the test set?

A: Accuracy= T(P)+T(N) / T(P)+F(P)+T(N)+F(N)
Accuracy: 0.93 = 93%

iii) **Recall: Correctly classified class**

Q1: What is the percentage of correct classification of (purchased) to the total input of the (purchased) in the test set?

A1: Recall (purchased)=T(P) / T(P)+F(P) = 0.00

Q2: What is the percentage of correct classification of (not purchased) to the total input of the (not purchased) in the test set?

A2: Recall (not purchased)=T(N) / T(N)+F(N) = 1.00

iv) **Precision: Correctly & wrongly classification of class**

Q1: What is the percentage of correct classification of (purchased) to the sum of correctly classified as (purchased) and wrongly classified as (not purchased) in the test set?

A1: Precision (purchased) = T(P) / T(P)+F(N) = 0.28

Q2: What is the percentage of correct classification of () to the sum of correctly classified as () and wrongly classified as () in the test set?

A2: Precision= T(P) / T(P)+F(N) = 0.00

v) **F1-Score/F1-measure: Overall performance of individual class**

If recall is high, Precision is low, by seeing the value of F1-measure, can validate the model performance

Q1: What is the overall performance of the (purchased)?

A1: F1-score (purchased)=2*[recall*precision] / [recall+precision] = 0.43

Q2: What is the overall performance of the (not purchased)?

A2: F1-score (not purchased)=2*[recall*precision] / [recall+precision] = 0.00

vi) **Macro-average: Overall average**

Q: What is the average performance of precision (correctly & wrongly classified)

A: Recall = R(P)+R(N) / 2 = 0.50
Precision = P(P)+P(N) / 2 = 0.14
F1-score = F1(P)+F2(N) / 2 = 0.22

vii) **Weighted-average:**

Q: What is the sum of the product of proportion rate (weight) of each class?

A: Recall: R(P)*[total count of () in test set / total count of test set] + R(N) = 0.28

Precision : R(P)*[total count of () in test set / total count of test set] + R(N) = 0.08

F1-score : R(P)*[total count of () in test set / total count of test set] + R(N) = 0.12

Support Vector Machine Classification report Tabulation:

		precision	recall	f1-score	support
	0	0.00	0.00	0.00	58
	1	0.28	1.00	0.43	22
	accuracy	0.28			80
	macro avg	0.14	0.50	0.22	80
	weighted avg	0.08	0.28	0.12	80

Conclusion:

From the above three classifications (Random Forest, Decision Tree, Support Virtual Machine), Random Forest classification model has produced 93% of the accuracy, hence for the Social Network Advertisement dataset, we shall create a model with Random forest classification for best results in the deployment phase.
