**Classification Models**

**Evaluation Metrics**

**Test set :**

|  |  |  |  |
| --- | --- | --- | --- |
| **N: Test set** |  | **Actual** | |
| **100** |  | **Purchased(P)** | **Non-Purchased (NP)** |
| **predicted** | **Purchased(P)** | **TRUE –PURCHASED**  **(True Positive)** | **FALSE – PURCHASED**  **(False Positive)** |
| **Non-Purchased (NP)** | **FALSE- NON PURCHASED**  **(False Negative)** | **TRUE-NON PURCHASED**  **(True Negative)** |

**SVM - Support Vector Machine(Classifier)**

The report for SVM Classification is: '

' precision recall f1-score support

0 0.90 0.94 0.92 79

1 0.87 0.80 0.84 41

accuracy 0.89 120

macro avg 0.89 0.87 0.88 120

weighted avg 0.89 0.89 0.89 120

|  |  |  |  |
| --- | --- | --- | --- |
| **N: Test set** | **P – 79** | **Actual** | |
| **120** | **NP - 41** | **Purchased(P)** | **Non-Purchased (NP)** |
| **predicted** | **Purchased(P)** | **74** | **5** |
| **Non-Purchased (NP)** | **8** | **33** |

**Questions:**

1. What is the Overall Performance of the SVM Classification model?

% of correct classification of purchase and non-purchase in the total input of the Test set: Accuracy.

Accuracy of SVM Classifier is 0.89

Formula: (T(P)+T(NP) ) / (T(P)+T(NP)+F(P)+F(NP))

(74+33) / (74+5+8+33) = 79/120 = 0.89

Accuracy = **89%**

//Apples currently classified as apples. (True,Positive)

1. What is the percentage of correct classification of **Purchase** in the **total input of Purchase** in the Test set?

Recall of SVM classifier where purchase =1

Recall is 0.94

Formula: (T(P) ) / (T(P)+ F(P)) = 74/79 = 0.936

Recall = **94%**

//Oranges currently classified as oranges. (True,Positive)

1. What is the percentage of correct classification of **Non-Purchase** in the **total input of Non-Purchase** in the Test set?
   * 1. Sample : ‘0’
   1. Recall , 0.81

Formula: (T(NP) ) / (T(NP)+F(NP)) = 33 / (33+8) = 0.80

Recall = **80%**

1. What is the percentage of correctly classified ‘**purchased**’ over sum of correct and incorrect classified ‘purchased’
   * 1. Sample : ‘1’
   1. Precision where purchase =1
   2. Precision is 0.90

Formula: (T(P) ) / (T(P)+F(NP)) = 74/ (74+8) = 0.90

Precision = **90%**

1. What is the percentage of correctly classified ‘**non purchased** over sum of correct and incorrect classified ‘non purchased’
   * 1. Sample : ‘0’
   1. Precision where purchase =0
   2. Precision is 0.86

Formula: (T(NP) ) / (T(NP)+F(P)) = 33 / (33 + 5) = 0.868

Precision = **87%**

1. What is the F1 score of ‘ Purchased? 0.86
   1. Formula: (2 \* (Recall \* Precision)) / (Recall + Precision) = 2\* (94\*90) / (94+90 ) = 91.9

F1 score = **92%**

1. What is the F1 score of ‘non purchased’? 0.84
   1. Formula: (2 \* (Recall \* Precision)) / (Recall + Precision)= 2\* (80\*87) / (80+87)=83.35

F1 score = **84%**

1. What is the weighted average of precision? 0.89
   1. (Precision (P) \* (P/(P+NP)) ) + (Precision (NP) \* (NP/(P+NP)) )

= 90 \* (79/120) + 87 \* (41/120) = 59.25 + 29.725 =88.98

Weighted average (Precision) = **89%**

1. What is the weighted average of Recall? 0.89
   1. (Recall (P) \* (P/(P+NP)) ) + (Recall (NP) \* (NP/(P+NP)) )
      1. 94 \* (79/120) + 80 \* (41/120) = 61.88 + 27.33 = 89.21
      2. Weighted average(Recall) = **89%**
2. What is the weighted average of F1 score? 0.89
   1. (F1 score (P) \* (P/(P+NP)) ) + (F1 score (NP) \* (NP/(P+NP)) )
   2. 92 \* (79/120) + 84 \* (41/120) = 60.56 + 28.7 = 89.26
   3. Weighted average(F1 score) = **89%**
3. What is the macro average of precision? 0.87
   1. (Precision(P) + precision(NP) )/ 2
   2. ( 90+87) / 2 = 88.5
4. What is the macro average of Recall?
   1. (Recall(P) + Recall (NP) ) / 2
   2. ( 94+80) / 2 = 87
   3. Macro average of (recall) = 87%
5. What is the macro average of F1 Score? 0.88
   1. (F1(P) + F1(NP) ) / 2
   2. ( 92+84) / 2 = 88
   3. Macro average of (F1 score) = 88%

**DT - Decision Tree (Classifier)**

Report:

The report for Decision Tree Classification is: '

' precision recall f1-score support

0 0.89 0.92 0.91 79

1 0.84 0.78 0.81 41

accuracy 0.88 120

macro avg 0.87 0.85 0.86 120

weighted avg 0.87 0.88 0.87 120

|  |  |  |  |
| --- | --- | --- | --- |
| **N: Test set** | **P – 79** | **Actual** | |
| **120** | **NP - 41** | **Purchased(P)** | **Non-Purchased (NP)** |
| **predicted** | **Purchased(P)** | **73** | **6** |
| **Non-Purchased (NP)** | **9** | **32** |

**Questions:**

1. What is the Overall Performance of the Decision Tree Classification model?

% of correct classification of purchase and non-purchase in the total input of the Test set: Accuracy.

Accuracy of Decision Tree Classifier is 0.88

Formula: (T(P)+T(NP) ) / (T(P)+T(NP)+F(P)+F(NP))

(73+32) / (73+6+9+32) = 105/111 = 0.875

**Accuracy = 88%**

//Apples currently classified as apples. (True,Positive)

1. What is the percentage of correct classification of **Purchase** in the **total input of Purchase** in the Test set?

Recall of Decision Tree classifier where purchase =1

Recall is 0.92

Formula: (T(P) ) / (T(P)+ F(P)) = 73/79 = 0.9240

Recall = 92%

//Oranges currently classified as oranges. (True,Positive)

1. What is the percentage of correct classification of **Non-Purchase** in the **total input of Non-Purchase** in the Test set?

Formula: (T(NP) ) / (T(NP)+F(NP)) = 32 / (32+9) = 32/41 = 0.78

Recall = 78%

1. What is the percentage of correctly classified ‘**purchased**’ over sum of correct and incorrect classified ‘purchased’
   * 1. Sample : ‘1’
   1. Precision where purchase =1
   2. Precision is 0.89

Formula: (T(P) ) / (T(P)+F(NP)) = 73/ (73+9) = 73/ 82 = 0.890

Precision = 89%

1. What is the percentage of correctly classified ‘**non purchased** over sum of correct and incorrect classified ‘non purchased’
   * 1. Sample : ‘0’
   1. Precision where purchase =0
   2. Precision is 0.84

Formula: (T(NP) ) / (T(NP)+F(P)) = 32 / (32 + 6) = 32 / 38 = 0.8421

Precision = 84%

1. What is the F1 score of ‘ Purchased? 0.91
   1. Formula: (2 \* (Recall \* Precision)) / (Recall + Precision) = 2\* (89\*92) / (89+92 ) = 90.47

F1 score = 90.5

1. What is the F1 score of ‘non purchased’? 0.81
   1. Formula: (2 \* (Recall \* Precision)) / (Recall + Precision)=
   2. 2\* (78\*84) / (78+84)= 80.8

F1 score = 80.8

1. What is the weighted average of precision? 0.87
   1. (Precision (P) \* (P/(P+NP)) ) + (Precision (NP) \* (NP/(P+NP)) )

= 89 \* (79/120) + 84 \* (41/120)

= (0.89\* 0.658) + (0.84 \* 0.34)

= 0.585 + 0.28

=0.865

Weighted average (Precision) = **87%**

1. What is the weighted average of Recall? 0.87
   1. (Recall (P) \* (P/(P+NP)) ) + (Recall (NP) \* (NP/(P+NP)) )
      1. (0.92\* 0.658) + (0.78 \* 0.34)
      2. =0.605 + 0.265
      3. =0.87
      4. Weighted average(Recall) = **87%**
2. What is the weighted average of F1 score? 0.87
   1. (F1 score (P) \* (P/(P+NP)) ) + (F1 score (NP) \* (NP/(P+NP)) )
   2. (0.90 \* 0.658) + (0.81 \* 0.34)
   3. = 0.5922 + 0.275
   4. = 0.867
   5. Weighted average(F1 score) = **87%**
3. What is the macro average of precision? 0.87
   1. (Precision(P) + precision(NP) )/ 2
   2. ( 89+84) / 2 = 86.5
   3. Macro average of (Precision) = 86%
4. What is the macro average of Recall? 0.85
   1. (Recall(P) + Recall (NP) ) / 2
   2. ( 92+78) / 2 = 85
   3. Macro average of (recall) = 85%
5. What is the macro average of F1 Score? 0.86
   1. (F1(P) + F1(NP) ) / 2
   2. ( 91+81) / 2 = 86
   3. Macro average of (F1 score) = 86%

**RF – Random Forest Classifier**

The report for Decision Tree Classification is: '

' precision recall f1-score support

0 0.96 0.92 0.94 79

1 0.86 0.93 0.89 41

accuracy 0.93 120

macro avg 0.91 0.93 0.92 120

weighted avg 0.93 0.93 0.93 120

|  |  |  |  |
| --- | --- | --- | --- |
| **N: Test set** | **P – 79** | **Actual** | |
| **120** | **NP - 41** | **Purchased(P)** | **Non-Purchased (NP)** |
| **predicted** | **Purchased(P)** | **73** | **6** |
| **Non-Purchased (NP)** | **3** | **38** |

**Questions:**

1. What is the Overall Performance of the Random Forest Classification model?

% of correct classification of purchase and non-purchase in the total input of the Test set: Accuracy.

Accuracy of SVM Classifier is 0.93

Formula: (T(P)+T(NP) ) / (T(P)+T(NP)+F(P)+F(NP))

(73+38) / (73+6+3+38) = 0.925

Accuracy = **93%**

//Apples currently classified as apples. (True,Positive)

1. What is the percentage of correct classification of **Purchase** in the **total input of Purchase** in the Test set?

Recall is 0.92

Formula: (T(P) ) / (T(P)+ F(P)) = 73/79 = 0.924

Recall = **92%**

//Oranges currently classified as oranges. (True,Positive)

1. What is the percentage of correct classification of **Non-Purchase** in the **total input of Non-Purchase** in the Test set?
   1. Recall is 0.93

Formula: (T(NP) ) / (T(NP)+F(NP)) = 38 / 41 = 0.9268

Recall = **93%**

1. What is the percentage of correctly classified ‘**purchased**’ over sum of correct and incorrect classified ‘purchased’
   1. Precision where purchase =1
   2. Precision is 0.96

Formula: (T(P) ) / (T(P)+F(NP)) = 73/76 = 0.960

Precision = **96%**

1. What is the percentage of correctly classified ‘**non purchased** over sum of correct and incorrect classified ‘non purchased’
   * 1. Sample : ‘0’
   1. Precision where purchase =0
   2. Precision is 0.86

Formula: (T(NP) ) / (T(NP)+F(P)) = 38/44 = 0.863

Precision = **86%**

1. What is the F1 score of ‘ Purchased? 0.94
   1. Formula: (2 \* (Recall \* Precision)) / (Recall + Precision)

= 2\* ( 0.92\*0.96 ) / ( 0.92+0.96 )

= 1.764 / 1.88

= 0.9382

F1 score = **94%**

1. What is the F1 score of ‘non purchased’? 0.89
   1. Formula: (2 \* (Recall \* Precision)) / (Recall + Precision)=

= 2\* (0.93\*0.86) / (0.93+0.86)

=1.5996 / 1.79

= 0.893

F1 score = **89%**

1. What is the weighted average of precision? 0.93
   1. (Precision (P) \* (P/(P+NP)) ) + (Precision (NP) \* (NP/(P+NP)) )

= 0.96 \* (79/120) + 0.86 \* (41/120)

= 0.96 \* 0.658 + 0.86 \* 0.342

= 0.632 + 0.294

= 0.926

Weighted average (Precision) = **93%**

1. What is the weighted average of Recall? 0.93
   1. (Recall (P) \* (P/(P+NP)) ) + (Recall (NP) \* (NP/(P+NP)) )

=0.92 \* 0.658 + 0.93 \* 0.342

= 0.605 + 0.320

=0.9258

Weighted average(Recall) = **93%**

1. What is the weighted average of F1 score? 0.92
   1. (F1 score (P) \* (P/(P+NP)) ) + (F1 score (NP) \* (NP/(P+NP)) )

=0.94 \* 0.658 + 0.89 \* 0.342

= 0.6185 + 0.3043

= 0.923

* 1. Weighted average(F1 score) = **92%**

1. What is the macro average of precision?
   1. (Precision(P) + precision(NP) )/ 2
   2. ( 96+86) / 2 = 91
   3. Macro average of (precision) = 91%
2. What is the macro average of Recall?
   1. (Recall(P) + Recall (NP) ) / 2
   2. ( 92+93) / 2 = 92.5
   3. Macro average of (recall) = 93%
3. What is the macro average of F1 Score?
   1. (F1(P) + F1(NP) ) / 2
   2. ( 94+89) / 2 = 91.5
   3. Macro average of (F1 score) = 92%

**The best model : Random Forest : 93%**