**General Questions**

1. Tell the category of model considered for the given problem statement. Explain.

**Ans:** We should use the classification model for the given problem statement, because the table considered for the model has an alphabetical dataset.

1. Differentiate between regression model and classification model.

|  |  |
| --- | --- |
| **Regression model** | **Classification model** |
| Considered if the dataset has only numerical values | Considered if the dataset has alphabetical data |
| Preferred for supervised learning | Can be used for supervised learning and unsupervised learning |
| Depicts minimum error in the model | Depicts confusion matrix of the model |
| Easy to understand | Needs prior knowledge about the structure of the confusion matrix |
| Output is simple and easy to understand | Classification report is challenging to understand without prior knowledge |

1. Can you use the classification model for more than one output?

**Ans:** Yes

1. Differentiate between type 1 error and type 2 error.

|  |  |
| --- | --- |
| **Type 1 error** | **Type 2 error** |
| Depicts false negative value in the confusion matrix | Depicts false positive value in the confusion matrix |
| Does not cause a significant setback in the result | Causes a significant setback in the result |
| Preferable to have minimum value | Most preferable to have nil value |

1. List the variables in the given problem statement

| **User ID** | **Age** | **Estimated Salary** | **Purchased** | **Gender** |
| --- | --- | --- | --- | --- |

1. Give the maximum count of any one variable in the problem.

**Ans:** 400

1. Are there any redundant variables in the given problem statement?

**Ans:** Yes. The variable “**User ID”** is a redundant variable. It does not offer any importance in the analysis. Removing this variable will not cause any significant setback during the analysis.

1. Identify the independent variables in the given problem statement.

**Ans:** 'Age', 'EstimatedSalary', 'Gender\_Male'

1. Which of the given variables is the dependent variable?

**Ans:** 'Purchased'

1. Is the dependent variable categorised under binary dataset?

**Ans:** Yes.

1. Give the count of individual variables in the output.

**Ans:** 400

1. Can you perform ensemble learning during classification analysis?

**Ans:** Yes.

1. Identify the algorithms that can be used to train the ML model for the given problem statement.

**Ans:** Support Vector Machine, Decision Tree, Random Forest.

1. What do you mean by the term support?

**Ans:** The term support describes the count of the class of variables during the analysis.

1. Define the term accuracy.

**Ans:** Accuracy describes the percentage of correct classification of both classes in a dependent variable to the total input of the test set in the problem.

1. Define the term recall.

**Ans:** Recall describes the percentage of correct classification of any one of the classes in a dependent variable to the total input of the same class in the test set in the problem

1. Write the meaning of precision.

**Ans:** Precision describes the percentage of correct classification of any one of the classes in a dependent variable, with the account considering wrong classification of the other class in the test set in the problem

1. The overall performance of any one of the classifiers in the model is called as \_\_\_\_\_\_\_\_\_\_

**Ans:** F1 measure

1. Define macro averaging.

**Ans:** Macro averaging describes the average performance of the following: recall, precision, F1 measure

1. Define weighted average.

**Ans:** Weighted average describes the sum of the product of the proportion, i.e., the weight of each class in the problem statement.

**Classification Codes Using SVM, DT and RF**

1. Give the confusion matrix for the SVM algorithm for the given problem statement.
2. The count of true positive in the model is \_\_\_\_\_\_\_\_\_\_
3. The count of true negative in the model is \_\_\_\_\_\_\_\_\_\_
4. The count of type 1 error in the model is \_\_\_\_\_\_\_\_\_\_
5. The count of type 2 error in the model is \_\_\_\_\_\_\_\_\_\_
6. Mention the accuracy of the model.
7. Tell the percentage of correct classification of the purchase.
8. Give out the percentage of correct classification of the purchase.
9. The precision of purchased value is \_\_\_\_\_\_\_\_\_
10. \_\_\_\_\_\_\_\_\_ is the precision of the not purchased dataset
11. The overall performance of purchased value is \_\_\_\_\_\_\_\_\_
12. Give the overall performance of the not purchased value.
13. The average performance of recall is \_\_\_\_\_\_\_\_\_
14. The average performance of precision is \_\_\_\_\_\_\_\_\_
15. \_\_\_\_\_\_\_ is the average performance of F1 measure
16. The weighted average of recall is \_\_\_\_\_\_\_\_\_
17. The weighted average of precision is \_\_\_\_\_\_\_\_\_
18. The weighted average of F1 measure is \_\_\_\_\_\_\_\_\_
19. Give the support values of purchased
20. Give the support values of not-purchased

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Q. No.** | **Question** | **Unit** | **SVM** | **DT** | **RF** |
| 1 | Confusion matrix |  |  |  |  |
| 2 | True positive | no | 77 | 71 | 257 |
| 3 | True negative | no | 18 | 38 | 142 |
| 4 | Type 1 error | no | 2 | 8 | 0 |
| 5 | Type 2 error | no | 23 | 3 | 1 |
| 6 | Model accuracy | no | 79 | 91 | 100 |
| 7 | Recall purchased | % | 44 | 93 | 99 |
| 8 | Recall not purchased | % | 97 | 90 | 100 |
| 9 | Precision of purchased | % | 90 | 83 | 100 |
| 10 | Precision of not purchased | % | 77 | 96 | 100 |
| 11 | F1 measure of purchased | % | 59 | 87 | 100 |
| 12 | F1 measure of not purchased | % | 86 | 93 | 100 |
| 13 | Macro average of recall | % | 71 | 91 | 100 |
| 14 | Macro average of precision | % | 83 | 89 | 100 |
| 15 | Macro average of F1 measure | % | 73 | 90 | 100 |
| 16 | Weighted average of recall | % | 79 | 91 | 100 |
| 17 | Weighted average of precision | % | 81 | 91 | 100 |
| 18 | Weighted average of F1 | % | 77 | 91 | 100 |
| 19 | Support value of purchased | no | 41 | 41 | 143 |
| 20 | Support value of not purchased | no | 79 | 79 | 257 |