

LOGISTIC REGRESSION ASSIGNMENT QUIZ

How many types of occupations do we have?

(Note: The give dataset contains missing values “?”, Thus impute missing values with NaN and perform all the questions accordingly)

Marked Answer :

14

Correct Answer :

14

MARKS OBTAINED ✓ 1 TOTAL MARKS : 1

How many people are working as tech support and have an annual income greater than 50k?

Marked Answer :

934

Correct Answer :

278

MARKS OBTAINED ✗ 0 TOTAL MARKS : 1

How many total missing values are present in the dataset?

Marked Answer :

4262

Correct Answer :

4262

MARKS OBTAINED ✓ 1 TOTAL MARKS : 1

If there are missing values in the Marital Status column, which option among the following should be used for replacing the missing values:

Marked Answer :

Mode

Correct Answer :

Mode

MARKS OBTAINED ✓ 1 TOTAL MARKS : 1

How many people are having private work classes and are not from the United States of America?

Marked Answer :

2151

Correct Answer :

2151

MARKS OBTAINED ✓ 1 TOTAL MARKS : 1

How many people are either having Annual Income(last column) less than or equal to 50k or their working hours is greater than or equal to 40 hrs:

Marked Answer :

25903

Correct Answer :

23448

MARKS OBTAINED ✗ 0 TOTAL MARKS : 1

Which of the following methods can you use for handling outliers?

Marked Answer :

Both of the above methods

Correct Answer :

Both of the above methods

MARKS OBTAINED ✓ 1 TOTAL MARKS : 1

Chi-square is used to analyze:

Marked Answer :

both 1 and 2

Correct Answer :

both 1 and 2

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

What is VIF?

Marked Answer :

It can detect multicollinearity

Correct Answer :

It can detect multicollinearity

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

What predict_proba will tell you?

Marked Answer :

It will predict the class probabilities

Correct Answer :

It will predict the class probabilities

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

Logistic regression is useful for regression problems:

Marked Answer :

False

Correct Answer :

False

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

In logistic regression, if the predicted logit is 0, what's the transformed probability?

Marked Answer :

0.5

Correct Answer :

0.5

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

Which variant of logistic regression is recommended when you have a categorical dependent variable with more than two values?

Marked Answer :

Multinomial logistic regression

Correct Answer :

Multinomial logistic regression

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

Perform the following tasks for answering the remaining questions

- Rename the last column as Annual Income
- Remove the missing values from the dataset
- Change the labels of categorical data into numerical data using Label Encoder.
- Split the dataset into a train and test of proportions 70:30 and set the random state to 0.
- Build a Logistic Regression Model on the data.

Answer the following questions with the help of the above-created model.

What is the accuracy score of the above model?

Marked Answer :

0.70 to 0.85

Correct Answer :

0.70 to 0.85

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

Perform the following tasks for answering the remaining questions

- Rename the last column as Annual Income
- Remove the missing values from the dataset
- Change the labels of categorical data into numerical data using Label Encoder.
- Split the dataset into a train and test of proportions 70:30 and set the random state to 0.
- Build a Logistic Regression Model on the data.

Answer the following questions with the help of the above-created model.

What is the specificity of the above model?

Marked Answer :

None of the above

Correct Answer :

None of the above

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

Perform the following tasks for answering the remaining questions

- Rename the last column as Annual Income
- Remove the missing values from the dataset
- Change the labels of categorical data into numerical data using Label Encoder.
- Split the dataset into a train and test of proportions 70:30 and set the random state to 0.
- Build a Logistic Regression Model on the data.

Answer the following questions with the help of the above-created model.

What is the model's precision when the target is False?

Marked Answer :

0.70 to 0.80

Correct Answer :

0.70 to 0.80

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

Perform the following tasks for answering the remaining questions

- Rename the last column as Annual Income
- Remove the missing values from the dataset
- Change the labels of categorical data into numerical data using Label Encoder.
- Split the dataset into a train and test of proportions 70:30 and set the random state to 0.
- Build a Logistic Regression Model on the data.

Answer the following questions with the help of the above-created model.


What is the total support value from the above model?

Marked Answer :

9049

Correct Answer :

9049

MARKS OBTAINED  1

TOTAL MARKS : 1

Perform the following tasks for answering the remaining questions

- Rename the last column as Annual Income
- Remove the missing values from the dataset
- Change the labels of categorical data into numerical data using Label Encoder.
- Split the dataset into a train and test of proportions 70:30 and set the random state to 0.
- Build a Logistic Regression Model on the data.

Answer the following questions with the help of the above-created model.


What is the f1 score of the above model when the target is True?

Marked Answer :

0.30 to 0.40

Correct Answer :

0.30 to 0.40

MARKS OBTAINED  1

TOTAL MARKS : 1

Perform the following tasks for answering the remaining questions

- Rename the last column as Annual Income
- Remove the missing values from the dataset
- Change the labels of categorical data into numerical data using Label Encoder.
- Split the dataset into a train and test of proportions 70:30 and set the random state to 0.
- Build a Logistic Regression Model on the data.

Answer the following questions with the help of the above-created model.


How many records are correctly classified by the model?

Marked Answer :

7173

Correct Answer :

7173

MARKS OBTAINED  1

TOTAL MARKS : 1


Choose the code which is used to perform reset indexes in the original data frame.

Marked Answer :

data.reset_index(inplace=True)

Correct Answer :

data.reset_index(inplace=True)

MARKS OBTAINED  1

TOTAL MARKS : 1

What is the cost function used in Binary Logistic Regression?

Marked Answer :

Log Loss

Correct Answer :

Log Loss

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

Which below method is used to find the best fitting line in logistic regression

Marked Answer :

Maximum Likelihood

Correct Answer :

Maximum Likelihood

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

What is the relation between Sensitivity and Specificity?

Marked Answer :

Sensitivity Increases Specificity Decreases

Correct Answer :

Sensitivity Increases Specificity Decreases

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

Use the creditworthiness data for the following questions, Don't perform any Data cleaning or pre-processing before answering the questions.

What is the entropy of the column Htype using the formula $(-1 * np.sum(np.log2(probs) * probs))$ where probs is the list of probability of each unique item in the column?

Marked Answer :

1.14

Correct Answer :

1.14

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

Use the creditworthiness data for the following questions, Don't perform any Data cleaning or pre-processing before answering the questions.

What is the total number of columns obtained post running the command pandas get_dummies in the raw data when the Initial columns count is 21?

Marked Answer :

65

Correct Answer :

65

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

What is the X & Y axis for drawing AUC ROC plot?

Marked Answer :

FPR & TPR

Correct Answer :

FPR & TPR

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

Use the creditworthiness data for the following questions, Don't perform any Data cleaning or pre-processing before answering the questions.

How many records present with the following conditions age >= 25 and age <= 50, and inplans equals 'bank' and Cpur is either 'Business' or 'electronics' and foreign as 'yes'?

Marked Answer :

0

Correct Answer :

0

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

Calculate the Recall for the below confusion matrix

Confusion Matrix		Predicted	
		Good	Bad
Actual	Good	300	200
	Bad	100	400

Marked Answer :

4/10

Correct Answer :

4/7

MARKS OBTAINED ❌ 0

TOTAL MARKS : 1

Calculate the F1 score for the below confusion matrix?

Confusion Matrix		Predicted	
		Good	Bad
Actual	Good	300	200
	Bad	100	400

Marked Answer :

0.67

Correct Answer :

0.67

MARKS OBTAINED ✔️ 1

TOTAL MARKS : 1

Total Marks

26 / 29