Classification Quiz

Last Updated: 2022-05-24 | Time to Complete: 90 min

miatta.sinayoko@gmail.com Switch account

Draft saved

* Indicates required question

Email *

☑ Record **miatta.sinayoko@gmail.com** as the email to be included with my response

What is your cohort name? *

Quintela

1. Of the following types of questions, which one can be answered best by a   8 points
Classification model?

☐ How many or how much of something

☐ What should we expect to happen next?

☐ Is this new observation in the predefined group A or B (or C, D, etc)?

☐ Is this observation weird?

☐ What are the groupings that exist in the data already

2. Which of the following algorithms are used to predict categorical outcomes based on historical outcomes? (Select all that apply)   7 points

☐ Decision Tree

☐ K-Means

☐ K-Nearest Neighbors

☐ Linear Regression

☐ Logistic Regression

☐ Random Forest

3. Which of the following are true regarding test, validate and train datasets? 8 points (Select all that apply)

☐ A "Test" data set is not a sample from the original dataset, but is instead gathered later, after having initially built the model on all of the original data.

☐ A "Train" data set is the generally largest of the 3 samples.

☐ One purpose of splitting the data into 3 samples is to avoid overfitting your model to one sample.

☐ The primary purpose of a "Test" data set is to test many different algorithms and see which performs best.

☐ Only the best model should be tested on the "Test" dataset.

4. When determining which independent variable(s) will help predict a target   8 points
variable, i.e. what are key drivers of a target variable, what method(s) should
be used? (Select all that apply)

☐   statistical tests

☐   confusion matrix

☐   data visualization

☐   baseline prediction

5. What method(s) can be used to avoid overfitting? Select all that apply.   8 points

☐   adjust hyperparameters

☐   fit model on unseen data

☐   split train dataset into validate and test

☐   skip cleaning the raw data

6. A Decision Tree algorithm learns patterns within sets of labeled points,   7 points
the classifier produces a _____ in order to classify new incoming data.

○   test rule

○   fact rule

○   decision rule

○   choice rule

7. How does the K-Nearest Neighbor algorithm make its predictions?   7 points

○   It makes predictions based on how many features a new data point has compared to
     a known data point.

○   It makes predictions based on coefficients that weight each input variable.

○   It makes predictions based on a sequence of rules applied to the new data point.

○   It makes predictions based on how close a new data point is to known data points.

8. The Logistic Regression algorithm transforms a(n) _____ into a logistic   7 points
model using a logistic function.

○   Linear model

○   Discrete model

○   Exponential model

○   Strategic model

9. Put the following steps in the correct order. 1 is first, 5 is last.   15 points

	1	2	3	4	5
Evaluate the model using the validate data frame	○	○	○	○	○
Explore, visualize, and test relationships between variables	○	○	○	○	○
Split the data into train, validate, and test data frames	○	○	○	○	○
Fit and evaluate the model on the training dataframe	○	○	○	○	○
Evaluate the best model on test data frame	○	○	○	○	○

10. Given the following labeled sample, we want to predict whether or not a   7 points
customer will churn in the next month (1-yes or 0-no).

has_churned = [1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0], where churning is the positive
case, represented by a 1.

What should our baseline prediction be?

○ 0

○ .27

○ .5

○ .73

○ 1

11. Given the above scenario (we want to predict whether or not a customer   8 points
will churn in the next month), answer the following 2 questions related to the
confusion matrix that would result from your baseline predictions. 1. Fill out
blocks A-D using your baseline prediction NOTE: One value may apply to
multiple cells. Remember: has_churned = [1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0]

CONFUSION MATRIX	Has Churned (1)	Has Not Churned (0)	
Predicted Churn (1)	A	B	E
Predicted Not Churn (0)	C	D	F
	G	H	I

	0	8	11	3
A	☐	☐	☐	☐
B	☐	☐	☐	☐
C	☐	☐	☐	☐
D	☐	☐	☐	☐

12. (continuing from question above, relating to the confusion matrix that   10 points
would result from your baseline predictions). 2. Match the metric VALUE
with the cell in the confusion matrix. NOTE: One value may apply to
multiple cells. Remember: has_churned = [1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0]

CONFUSION MATRIX	Has Churned (1)	Has Not Churned (0)	
Predicted Churn (1)	A	B	E
Predicted Not Churn (0)	C	D	F
	G	H	I

	0	.73	1
E	☐	☐	☐
F	☐	☐	☐
G	☐	☐	☐
H	☐	☐	☐
I	☐	☐	☐

A copy of your responses will be emailed to miatta.sinayoko@gmail.com.

Page 1 of 1

Submit   Clear form

Never submit passwords through Google Forms.

This form was created inside of Codeup. Report Abuse

Google Forms