Score: 89% · 32/36

Exam completed on: September 9, 2021

(dIII	questions Show all Show wr	ong or
×	Question 1 of 36	
	Select the regular expression that returns the full year 2018, and not the single numbers individually.	~
	[0-9]	
	X [2-8]+	
	[0-9]*	
	√ [0-9]+	
	Question 2 of 36	
V	Select the scenario that would not make a feature a prime candidate for transformation.	~
	few outliers	
	dramatic skew	



/

Question 3 of 36

You believe that spam text messages are longer than real ones. What is one way to determine if you are correct?

Create a feature that contains bins for the length of each text message and plot the data on a pie chart.

Create a feature that contains bins for the length of each text message and plot the data on a histogram.

Create a feature that contains bins for the length of each text message and plot the data on a scatter plot with a trend line.

/

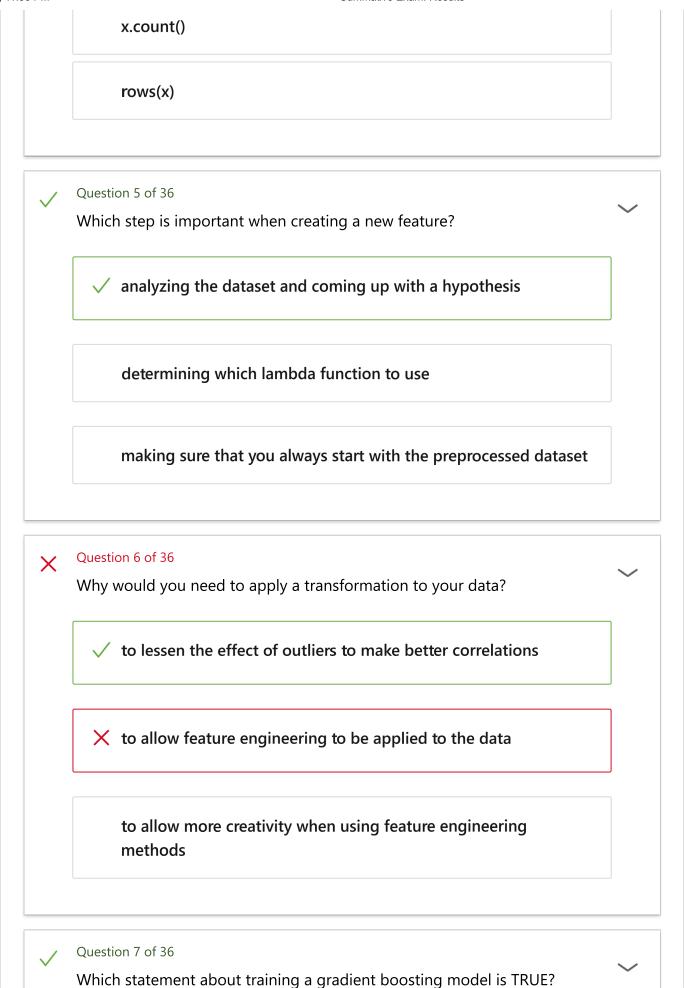
Question 4 of 36

After reading the data using pd.read_csv() into x, how can you tell the number of rows?



 \checkmark len(x)

x.rows()



Changing the learning rate parameter will have a very small effect on the overall results.

✓ An n_jobs parameter is not needed.

All decision trees will have unlimited depth.



Ouestion 8 of 36

How does gradient boosting work?

It forms a strong learner by discarding mistakes from prior test iterations.

It generates a range of input conditions for a machine learning model to test with.

✓ It combines weak learners together to form a strong learner by improving on mistakes from prior test iterations.

It generates boundary conditions for a machine learning model to test with.



Question 9 of 36

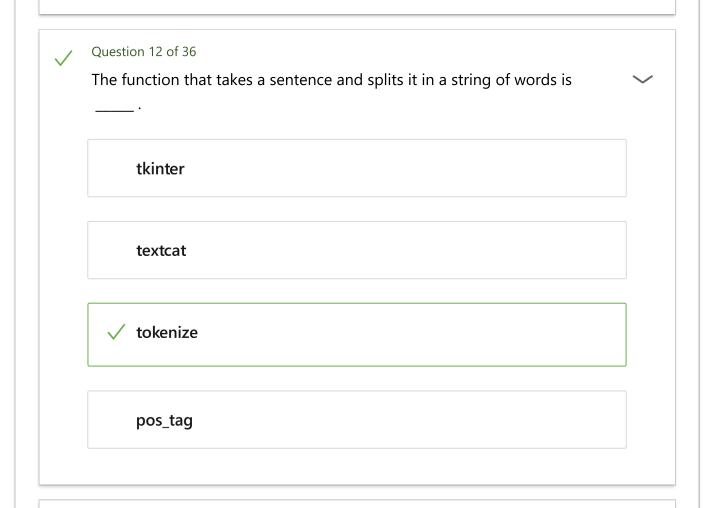
Which two concepts are combined to create a more powerful tool to tune and evaluate machine learning models?

cross validation and transformation cross validation and vectorizing ✓ grid search and cross validation grid search and transformation Question 10 of 36 How would you define a stop word? a frequently used word that appears before a period a frequently used word that appears at the end of a sentence ✓ a frequently used word that doesn't contribute to the meaning of the sentence a frequently used word that contributes to the meaning of the sentence Question 11 of 36 X What is the purpose of a grid search? to determine which hyperparameter has the greatest effect on a

model's accuracy

to determine which hyperparameter has the least effect on a model's accuracy

- to apply different combinations of hyperparameter settings to see if test set performance can be improved
- × to apply different combinations of hyperparameter settings to see if the training performance can be improved



Question 13 of 36

What is true about TF-IDF?

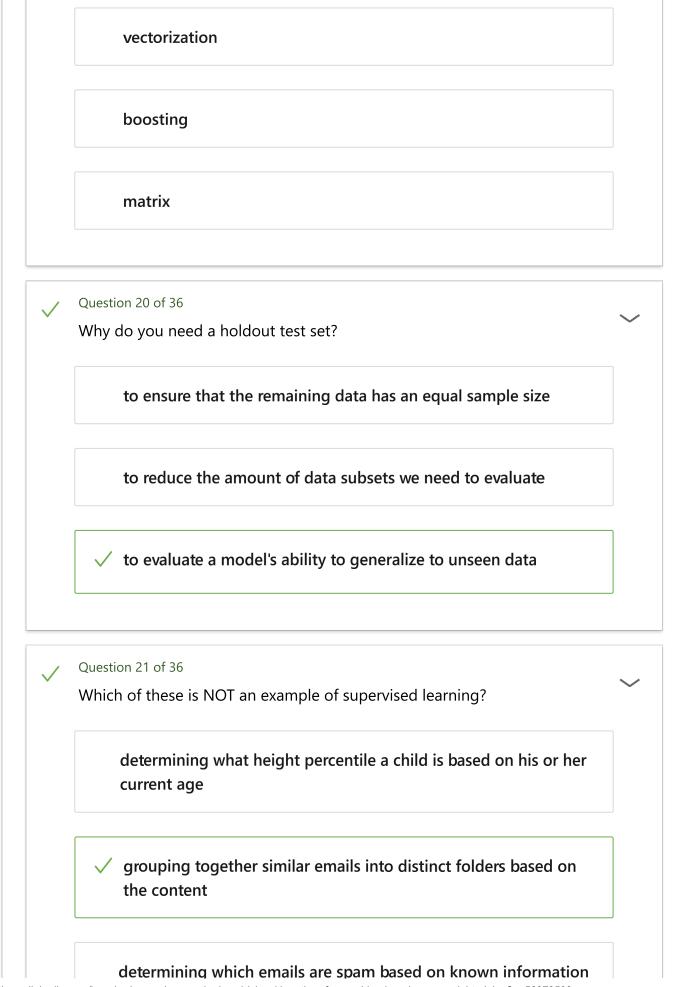
	√ all of these answers	
	The cells represent a weighting,	
	There is still one row per text message.	
	The columns still represent single unique terms.	
✓	Question 14 of 36 NLP is a field concerned with the ability of a computer to human language.	~
	manipulate	
	understand	
	analyze	
	✓ all of these answers	
✓	Question 15 of 36	~
	What does an n-gram represent?	
	all the words of length n in a string	

all the sentences of length n inside a data file all the words that appear n times inside a data file √ all combinations of n adjacent words Question 16 of 36 In bigram, how many tokens will be generated for the statement "NLP is an interesting topic"? 5 **√** 4 2 3 Question 17 of 36 Why is lemmatizing more accurate in finding word variations that have the same meaning? ✓ It has a database of nouns, verbs, adjectives, and adverbs that are grouped together as sets of synonyms. It uses context clues from other words in the sentence.

It has a centralized database of nouns, verbs, adjectives, and adverbs that are constantly updated by its userbase.

It specializes in finding synonyms for common slang words found in SMS messages.

/	Question 18 of 36 Lemmatizing 'meanness' and 'meaning' results in	•
	✓ meanness meaning	
	mean mean	
	meaning meanness	
	meanness meanness	
	Question 19 of 36 Random forest is an example of what type of method?	•
	✓ ensemble	



about the sender



Question 22 of 36

How does the Box-Cox power transformation work?

It provides the equation of a line that has the best correlation with the actual data points.

It applies a range of logarithms to your data points to determine if the result fits closely to a normal distribution.

✓ It applies a range of exponents to your data points to determine if the result fits closely to a normal distribution.

It provides a mapping of the actual data points to data points that would be found in a normal distribution.



Question 23 of 36

If X is -2, which one of the following is a correct transformation?



√ 1/y^2

x^-2

x^2

y^2

Question 24 of 36

Which statement best describes a bimodal distribution graph?

a curve with one spike and two long tails on each side of the spike

a curve with two large dips in different locations

✓ a curve with two large spikes in different locations

a curve with a long tail to the right

/ Question 25 of 36

Lemmatizing _____.

is faster than stemming

chops the end of the word using heuristics

✓ is slower than stemming

does not understand the context in which the word is used



Question 26 of 36

Why is it necessary to analyze the output of a stemming process?



Stemming algorithms are not perfect, as they can stumble on slang words and certain root words.

Stemming algorithms can inadvertently remove certain tokens.

Stemming algorithms can crash on long words and possibly return incomplete output.



Question 27 of 36

What is the result of running the stemmer against 'run', 'running', and 'runner'?



V

run

run

runner

runner

run

runner

running running runner

run run run Question 28 of 36 X Why is stemming important? X It captures variations of the same root word to help an NLP algorithm learn more words. It captures variations of the same root word to produce a larger number of tokens. It indexes all variations of a word to populate a NLP word database. It reduces variations of the same root word to produce a smaller number of tokens. Question 29 of 36 Stemming Meanness/meaning will result in _____. Meanness Mean Mea

meaning



Question 30 of 36

Using the NLTK, which package can be called to get a pre-defined list of stop words?

nltk.stopwords.words

√ nltk.corpus.stopwords.words

nltk.corpus.words.stopwords

nltk.stopwords

/

Question 31 of 36

What will be displayed on the screen when the following code runs?

```
tokenized = ['test','in','the','rest','of','for','new','last']
result = [word for word in tokenized if word not in
['in','on','the','of','for']]
print(result)
```

```
['test', 'rest', 'new', 'for']
```

```
 ('test', 'rest', 'new', 'last']
```

```
['test', 'for', 'the', 'rest', 'new', 'last']
```

['test', 'new', 'last']



Question 32 of 36

When writing your own tokenization function, which essential step must your function be able to do?

Check that your function doesn't read past the end of the data file.

Check that the correct punctuation is used for any sentences found in the data file.

LEARNING

Download certificate

Retake exam

Return to course

data file.

Return the number of tokens found in the data file.



Question 33 of 36

What will be printed with the following statement: print(re.split('\W+',"some of the-words are+combined"))



✓ ['some', 'of', 'the', 'words', 'are', 'combined']

```
Summative Exam: Results

['some', 'of', 'the-words', 'are+combined']

['', '', '', '-', '', '+', '']

['some of the', 'words are', 'combined']

Question 34 of 36

What stop word that will be removed from this sentence: "This is a test for the man to be successful in their lives"?

the
```

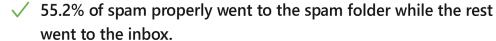
for

√ all of these answers



Question 35 of 36

In the spam-ham example, what does a recall rate of 55.2% mean?



55.2% of all emails were identified as spam.

Any email, whether it was spam or not, was correctly identified 55.2% of the time.



Question 36 of 36

What does it mean when a random forest has $max_depth=none$ and $n_estimators=10$?

the random forest will have 10 decision trees with a minimum depth of zero

the random forest will have unlimited trees with a depth of 10

√ the random forest will have 10 decision trees of unlimited depth

Return to course

English (English) ▼ • About • Become an Instructor • Help • Privacy & Terms ∧ • Accessibility • Apps ∧

Linkedin Corporation © 2021