Tenth Assignment

Performance summary	
Maximum number of attempts	Unlimited
Number of attempts	1
Your score	61
Status	Not available

•	Results

Course	Big Data
Test	10_Assignment_BD

This are your test results

Duration	4h 7m 45s
Answered	7 of 7 questions (100%)
Your score	61 of 100 points (61%)



Knowledge Questions 4



Practice 1



Naive Bayes 2





& Knowledge Questions 35 of 46 points (76%)

i≡ True/false		
Status	Answered	
Your score	15 / 21	71%

Response

Please tick whether the following statements are right or wrong?

Note:

Maximum score → 21 points

Minimum overall score -> 0 points

Wrong answer -> -3 points

Unanswered -> 0 points

Unanswered	Right	Wrong	
0	0	•	Stream processing allows for much higher throughput than batch processing.
0	•	0	With watermarks the output can be delayed.
0	0	0	With triggers, it can be defined when the streaming application outputs the data.
0	0	0	Spark ensures that events are processed in the order of their creation.
0	0	0	Linear regression can only be used for discrete response variables.
0	•	0	Normalization can be done with an estimator.
0	•	0	One hot encoding can be done while iterating the DataFrame once.

▼ Solution

Please tick whether the following statements are right or wrong?

Note:

Maximum score -> 21 points

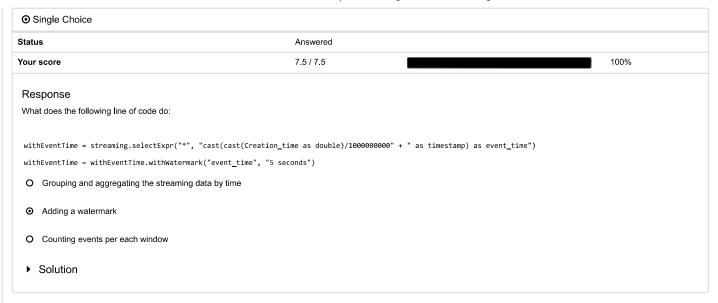
Minimum overall score -> 0 points

Wrong answer -> -3 points

Unanswered -> 0 points

Unanswered	Right	Wrong	
0	0	•	Stream processing allows for much higher throughput than batch processing.
0	•	0	With watermarks the output can be delayed.
0	⊚	0	With triggers, it can be defined when the streaming application outputs the data.
0	0	•	Spark ensures that events are processed in the order of their creation.
0	0	⊚	Linear regression can only be used for discrete response variables.
0	•	0	Normalization can be done with an estimator.
0	0	•	One hot encoding can be done while iterating the DataFrame once.

Status	Answered	
Your score		E09/
tour score	5 / 10	50%
Response		
A training function is a homomorphism	if it is	
☐ distributive		
✓ identity		
□ commutative		
▼ Solution		
A training function is a homomorphism	if it is	
☐ distributive		
□ identity		
S		
✓ commutative		
⊙ Single Choice		
Status	Answered	
Your score	7.5 / 7.5	100%
Response		
What does the following line of code do	n'	
import pyspark.sql.functions as f	•	
	f.window(f.column("event_time"), "10 minutes", "5 minutes")).count()	
	,,,,	
O Casting the data into an EventTin	ne.	
O Writing the results of the query in	to memory.	
O Counting the events per each wir	dow.	
► Solution		



∢ go back to overview

Representation of 26 points (23%)

Correction : Second last value in the time column is 8:35 instead of 8:30.

Windowing		
Status	Answered	
Your score	6 / 26	23%

Response

A zoo wants to keep track of how many visitors there are. Therefore, at all entrances and exits sensors are installed that can tell how many people enter or leave the zoo. The following table lists all events in the order that they are processed. Each event causes the trigger to start the computation.

Time	Entering/Leaving	Number of Persons
8:00	Entering	1
8:30	Leaving	3
8:10	Entering	5
8:15	Entering	1
8:25	Leaving	1
8:20	Leaving	2
8:30	Entering	4
8:40	Leaving	3

Assume you have a window of size 15 minutes that slides every 10 minutes. How does the number of people change during each window? For each window, give the start time, the end time, and the change of the number of people (+1 if a person enters, and -1 if a person leaves).

Notes:

- If an event happens exactly at the end time of a window, it does not count for this window anymore.
- The last window starts at 8:40

Start	End	Change
8:00	8:10	+1
8:10	8:20	+6
8:20	8:30	-2
8:30	8:40	0
8:40	8:50	+1

▶ Solution

∢ go back to overview

Naive Bayes 20 of 28 points (71%)

The following DataFrame contains information about fruits. It stores the color of the pulp, its taste, and whether it is edible. Use a Simple Naive Bayes classifier to predict whether a fruit with yellow pulp that tastes sour is edible based on the given table.

PulpColor	Taste	Edible
red	sour	yes
red	sour	no
red	bitter	no
yellow	sweet	yes
yellow	bitter	no
green	sour	yes
green	sweet	yes
green	bitter	no



⊙ Test period was over at 6/30/2021, 8:00 PM.