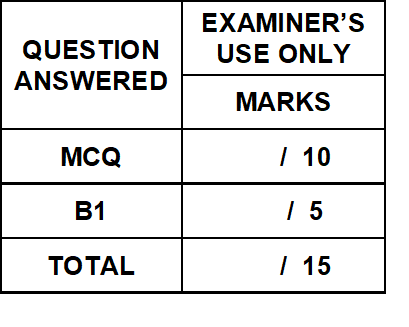
|  |  |  |  |
| --- | --- | --- | --- |
| **Specialist Diploma In Industrial Internet of Things**  **Engineering Analytics & Machine Learning (ECSE202)** | | | Description: C:\Users\rajahk\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.Outlook\Z7A7LSNK\big (2).jpg |
|  | | |  |
| **Name:** | **Class: PE** | | |
| **Admin No:** | **Marks:** | | |

**Written Test Set B**

**Instructions to Candidates:**

* The duration of the written test is 30 minutes.
* This paper consists of two sections, A and B.
* Section A is worth 10 marks and Section B is worth 10 marks
* Answer all questions in both sections.
* You are to write all your answers on this set of question paper.

**Section A:** This section consists of 10 multiple choice questions. Each question is worth 1 mark.

* 1. Which of the following statement(s) is (are) **TRUE** about the underlying format of date time for Microsoft Excel?

1. 0.25 is 06:00 AM.
2. It is stored in a 32-bit floating point.
3. Time is stored as 64-bit float point.
4. The length of a day equals 0.995.

( )

* 1. Which of the following is a characteristic of data collected from sensors?

1. Highly structured
2. Usually unstructured
3. No missing data
4. Mostly received through software API (Application Programming Interface)

( )

* 1. A sensor is used to detect whether the temperature of a cooker is low, medium or high. Below is the data collected over various cooking sessions:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Low | Medium | High |
| Occurrence | 500 | 325 | 244 |

What is the empirical probability of a High temperature occurring?

1. 0.304
2. 0.478
3. 0.532
4. 0.228

( )

* 1. In a normal distribution, 68% of values are:

1. within 2 mean of the standard deviation
2. within 1 standard deviation of the mean
3. within 2 standard deviation of the mean
4. within 1 mean of the standard deviation

( )

* 1. A random variable X sample space is defined by. What is the probability of P(X=1U X=8)? Note that the probability of any two sets of outcomes is mutual exclusive.

1. 3/8
2. 3/6
3. 2/8
4. 2/7

( )

* 1. Which of the following statement(s) is (are) **TRUE** about Unix Epoch (or Unix Timestamp)?

1. Unix Epoch is the number of days have elapsed since1970 (midnight UTC/GMT).
2. Unix Epoch is the number of minutes have elapsed since 1970 (midnight UTC/GMT).
3. Unix Epoch is the number of seconds have elapsed since 1970 (midnight UTC/GMT).
4. Unix Epoch is the number of seconds have elapsed since 1900(midnight UTC/GMT).

( )

* 1. Which of the following statement(s) is (are) **False** about data cleaning.

1. Remove or compensate for mean that skews the data unrealistically.
2. To produce clean data for further analysis.
3. To handle missing data.
4. All attributes of the right type or format.

( )

* 1. Data points with standard transformation (z-transform)

1. have values range from 0 to 1
2. have values with mean of 1 and standard deviation of 1
3. have values with mean of 0 and standard deviation of 1
4. have values range from -1 to 1

( )

* 1. Why data collection is responsible for the quality of analysis?

1. Data is always collected from sensors
2. Data collection is usually tedious.
3. Data is easily collected.
4. Unreliable input data would led to wrong analysis result.

( )

* 1. Data Analytics process consist of these processes:

1. Data Cleaning
2. Data Collection
3. Modelling/Prediction
4. Transformation/feature extraction
5. Visualization

Which of the following is the correct sequence of processes?

1. (2),(1),(3),(4),(5)
2. (2),(1),(4),(5),(3)
3. (1),(3),(5),(4),(2)
4. (1),(5)(3),(4),(5)

( )

10

**Section B:** This section consists of a short question worth 5 marks.

1. Company WXY had started trial run production of a new 200m steel bar by 4 types of processes. A technician was assigned to performance on a sample of steel bar on each process. The technician reported data as shown below:

|  |  |  |
| --- | --- | --- |
| Process | Mean (m) | Standard Deviation |
| A | 201 | 1.5 |
| B | 199 | 0.5 |
| C | 199 | 0.07 |
| D | 202 | 0.01 |

Assuming the data is normally distributed. Which process would be ready for mass production that required a failure rate of only 5% and tolerance of +/- 0.1m? Show workings to justify your recommendation.

5