|  |  |  |
| --- | --- | --- |
| جامعة فيلادلفيا*cid:image001.png@01CFF76F.9CE975D0*Philadelphia University | اسم النموذج: نموذج الامتحانPaper Examination | QFO-AP-FI-015 |
| الجهة المصدرة: كلية تكنولوجيا المعلومات |  رقم الاصدار : 2 Revision 2  |
| الجهة المدققة: عمادة ضمان الجودة |  التاريخ :10/11/2018 |
| عدد صفحات النموذج: 1 |

Lecturers: Dr. Belal Alifan

Coordinator: Dr. Belal Alifan

Internal Examiner: Dr. Mohammad Aloudat

Data Science Fundamentals (74114100) Mid Exam First Semester 2023-2024

Student Name: Date: 26-11-2023 Time: 60 minutes

Information for Candidates

1. *This examination paper contains 7 questions, totaling 30 marks.*
2. *The marks for parts of questions are shown in round brackets.*

Advice to Candidates

 *1. You should attempt all questions.*

 *2. You should write your answers clearly.*

I. Basic Notions

*Objectives: The aim of the questions in this part is to evaluate the required minimal student knowledge and skills. Answers in the pass category represent the minimum acceptable standard.*

*Question One:* Choose the correct answers [9 Marks]

1. Which of the following is **NOT** a Characteristics of Big Data?
2. Velocity B. Volume C. Variety D. Volatility
3. A ---------------------is constructed by **integrating** data from multiple **heterogeneous** sources.
4. data warehouse B. Big Data C. Data D. Relational Data
5. The characteristic of big data that refers to the Quality of the data is:

A. Volume B. Velocity C. Variety D**.** Veracity

1. \_\_\_\_\_\_\_\_\_\_\_ is any data that is expensive to manage and hard to extract value from

A. Data B. Big Data C. Data warehouse D. Information

1. \_\_\_\_\_\_\_\_: Data that has no inherent structure, often with multiple formats.

A. Structured data B. Semi-Structured C. Quasi-Structured D. Unstructured

1. Which type of the following data examples can be analyzed directly?
2. JSON files B. Videos C. Audios D. Bank statement
3. \_\_\_\_\_\_\_\_\_refers to how useful the data is in decision making.
4. Value B. Velocity C. Volume D. Variety
5. The aim of data science is to:
6. Reveal the features of complicated natural, human, and social phenomena.
7. Study the data
8. Understand what is the meaning of data
9. None of the above.
10. Data that is used to teach machines is called\_\_\_\_\_\_
11. Testing data B. Information C. Training data D. Machine data

*Question Two*: Classify the following as structured, unstructured or semi-structured data

[3 Marks]

* YouTube videos.

* Philadelphia University student information system’s data (registration).

* Weather data set stored in a JSON file

II. Familiar Problems Solving

Objectives*: The aim of the questions in this part is to evaluate that the student has some basic knowledge of the key aspects of the lecture material and can attempt to solve familiar problems.*

*Question Three*: Chose the correct words to fill in the blanks below. [4 marks]

 Data reduction Data pre-processing Data Integration

Data discretization  Artificial Intelligence Data Science

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_converts a large number of data values into smaller once, so that data evaluation and data management becomes very easy.
2. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_is an interdisciplinary field of scientific methods, processes, algorithms and systems to extract knowledge or insights from data in various forms?
3. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_is a process to reduce the large data into smaller once in such a way that data can be easily transformed further.
4. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_is about giving machines the capability of mimicking human behavior, particularly cognitive functions.

*Question four:* Communicating Your Results is the seventh stage of the data Science life cycle stages, Fill in the blank with the name of the other stages in the right order: [3 Marks]

*Question Five:*  Feature engineering stage is the process of transforming raw data into ‘features’ that better represent the underlying structure and patterns in the data.

Give two examples of feature engineering techniques. [2 marks]

*Question 6:*

Binning is a smoothing method that data scientists use to treat noisy data, considering the following ordered data values which are some product prices in US dollars [6 marks]

Prices (Data values): 11,13,13,15,15,16,19,20,20,20,21,21,22,23,24,30,40,45,45,45,71,72,73,75

Perform smoothing according to the following methods

a) Smoothing by bin mean

b) Smoothing by bin median

c) Smoothing by bin boundaries

1. Unfamiliar Problems Solving

*Objectives. The aim of the questions in this part is to evaluate that the student can solve familiar problems with ease and can make progress towards the solution of unfamiliar problems, and can set out reasoning and explanation in a clear and coherent manner.*

*Question 7:*

Assume that the Ministry of education and higher education in Jordan would like to implement a full-fledged e-learning system for the use of schools, universities and the continuous education in Jordan. The system should be able to handle different e-learning contents including videos, images, documents, … as well as various e-learning activities such as assignments, discussion boards, comments, and feedback etc.

Do you recommend the use of big data in this case, **why? If not, why not? (Justify your answer)** [3 Marks]

 Good Luck