

CATHOLIC UNIVERSITY OF EASTERN AFRICA
FACULTY OF SCIENCE
DEPARTMENT OF COMPUTER AND INFORMATION SCIENCE

CMT 434 PYTHON FOR DATA SCIENCE

CONTINUOUS ASSESSMENT TEST

Date: 31/10/2024

Time: 10AM-11:30AM

Instructions:

Attempt all questions

All codes to be written in PYTHON.

Indicate your and admission number on the answer sheet

Submit your work by uploading to Link provided in ODEL

Question 1 (10 marks)

- a) Explain how the following terms as used in data science: [3 marks]
- i. Artificial intelligence analysis
 - ii. Data analysis
 - iii. Big Data
- b) Data science derived insights from huge and diverse sets of data that may be structured, unstructured or semi structured.
Required:
- i. Explain the underlined types of data. [3 marks]
 - ii. State the single terms that refers to huge and diverse sets of data [1 mark]
- c) Explain how variety of big data has influenced the development of new types of Database Management Systems. [4 marks]

Question 2 (10 marks)

- a) Use a diagram to illustrate the various components of Hadoop framework. [6 marks]
- b) State any FOUR reasons why Python is preferred language for Data Science. [4 marks]

Question 3 (20 marks)

- a) Numpy library/module is important because it is used to create NumPy arrays. Why is necessary to so yet we can create arrays using python lists? [3 marks]
- b) Read the narration below and answer the questions that follow:

Prof. Geoffrey Hinton has raised ethical issues related to the use AI and machine learning. After he pioneered techniques such as back propagation, algorithms like the Restricted Boltzmann Machine (RBM) and having worked in projects like Google Brains, Geoffrey is now switching his efforts towards creating awareness on the dangers of the same thing he

helped create. He admits that after mechanizing thinking in expert systems, we have now automated it in machine learning and in its variants like deep learning. We have created a tool that can solve many problems for humans but at the same time, a tool we might not control and therein lies a big risk. GPT4 has helped draw attention to this reality. Required:

- i. Explain the term machine learning. [1 mark]
- ii. What is RBM and why is it said to be restricted? [4 marks]
- iii. What is back propagation? [1 mark]
- iv. Differentiate machine learning-based systems from Knowledge based systems. [4 marks]
- v. Write a program that reads the above narration from the keyboard and writes it into a text file named hinton.txt that is in the same directory as the program. [4 marks]
- vi. Write a program that reads hinton.txt and displays the sentence in it each in a separate line. The lines should be serialised/numbered with the serial number appearing at the beginning. [6 marks]