

FEDERAL UNIVERSITY OF TECHNOLOGY OWERRI
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY (SICT)
DEPARTMENT OF COMPUTER SCIENCE
2019/2020 HARMATTAN SEMESTER EXAMINATIONS

CSC 513: DATA MINING AND BIG DATA ANALYSIS

Time: 2½ hours

Date: February 19, 2021

INSTRUCTION: Answer question ONE and ANY OTHER FOUR (4) questions.

- 1 (a) (i) What is the essence of data mining? 4 marks
(ii) Define the term data classification. 3 marks
(b) (i) With respect to data classification, briefly explain the term class. 4 marks
(ii) Compare and contrast training data and test data. 4 marks
(c) Explain any five (5) issues associated with data mining. 5 marks
- 2 (a) What is Predictive Analytics? 5 marks
(b) Briefly describe the two categories of Predictive Analysis Models. 10 marks
- 3(a) (i) State four requirements of Clustering in data mining. 4 marks
(ii) What is the relationship between association and correlation with respect to data mining? 3 marks
(b) Outline the operation of the k-means algorithm. 8 marks
- 4(a) Discuss the process of knowledge discovery in data mining 7 marks
(b) How do we categorize data mining? Explain any three 4 marks
(c) What are the kinds of data that can be mined? 4 marks
- 5 (a) How can data be discovered? Explain any three 6 marks
(b) Explain at least three data mining tools and how they can be applied? 9 marks
- 6 (a) Write short notes on the following
(i) Frequent patterns 3 marks
(ii) Market basket analysis 3 marks
(iii) Apriori algorithm 3 marks
- (b) The table below consists of transaction IDs and items. Find out the list of items whose minimum support is greater than 2 6 marks

TID	ITEMS
10	A, C, D
20	B, C, E
30	A, B, C, E
40	B, E

- 7 (a) What is big data? 2 Marks
(b) Explain big data in terms of the following metrics
(i) velocity 2 Marks
(ii) volume 2 Marks
(iii) variety 2 Marks
(iv) veracity 2 Marks
(v) value 2 Marks
(c) Describe the relationship between data mining, big data and business intelligence using the diaper and beer model. 3 Marks