B.Tech. (SEM V) THEORY EXAMINATION 2022-23 **DATA ANALYTICS**

Time: 3 Hours

Note: Attempt all Sections. If require any missing data; then choose suitably.

SECTION A

1. Attempt all questions in brief.

- (a) What are the common open source tools for the model building phase?
- (b) What is decision tree?
- (c) What is learning rate?
- (d) What is rule induction?
- (e) How can you deal with uncertainty?
- (f) What is DSMS?
- (g) Write names of two visualization tools.
- (h) Explain the principle behind Hierarchical clustering technique.
- (i) Differentiate between Pig and SQL.
- (j) Write a short note on R graphical user interface.

SECTION B

2. Attempt any three of the following:

- (a) Compare and contrast traditional analytics structure to modern analytics architecture.
- (b) Explain multivariate analysis and Bayesian network.
- (c) Explain Datar-Gionis-Indyk-Motwani (DGIM) algorithm for counting oneness in a window.
- (d) Why PCY algorithm is preferred over Apriori algorithm?
- (e) How RDBS is different from NoSQL?

SECTION

Attempt any one part of the following: 3.

- (a) Explain Apache Hadoop, KNIME & Open refine in detail.
- (b) Explain various phases of Data Analytics Life Cycle.

4. Attempt any one part of the following:

- (a) Differentiate between Crisp logic and Fuzzy logic.
- (b) What are the different kernel methods of Data Analytics?

$2 \ge 10 = 20$

Total Marks: 100

10x1 = 10

10x1 = 10

Sub Code:KCS051 Roll No.

5. Attempt any one part of the following:

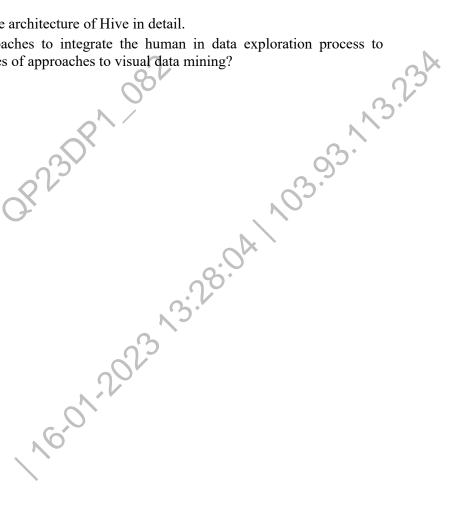
- (a) Explain Bernoulli sampling with its algorithm.
- (b) What are the different components of a general stream processing model? List few sources of streaming data .

6. Attempt any one part of the following:

- (a) What is Prediction error ? With the help of suitable example explain prediction error in classification and regression.
- Explain SON algorithm to find all or most frequent item sets using at most two (b) passes.

7. Attempt any one part of the following:

- Draw and discuss the architecture of Hive in detail. (a)
- What are the approaches to integrate the human in data exploration process to (b) realize different types of approaches to visual data mining?



10x1 = 10

10x1 = 10

10x1 = 10