JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST-III (EXAMINATION - JUNE 2023)

M.Tech. - II Semester (CSE)

COURSE CODE (CREDITS): 22M11CI213 (3)   MAX. MARKS: 35

COURSE NAME: Big Data Analytics

COURSE INSTRUCTOR: Deepak Gupta   MAX. TIME: 2 Hrs.

Note: 1) All questions are compulsory. Marks and COs are indicated against each question. 2) Attempt questions in the given sequence. 3) Be precise in your answers. 4) Write neatly.

Q 1.   How does Apache Spark process the data streams from sources like Kafka, Flume, etc.?   [5]   CO1

Q 2.   Explain the usages of SparkContext object on a Spark cluster by taking one example for each.   [5]   CO1

Q 3.   (a) Differentiate between hard and soft clustering with suitable big data examples.   [2.5   CO5
         (b) Discuss the benefits and limitations of clustering for streams and parallelism taking   *2]
             suitable big data examples.

Q 4.   Discuss any five techniques used for direct discovery of communities in social mining graphs with suitable examples.   [5]   CO4

Q 5.   How does the knowledge-based recommendation system differ from collaborative and content-based recommendation system?   [5]   CO4

Q 6.   (a) How matrix factorization is utilized in a recommender system?   [2.5   CO3
         (b) Consider one use case to discuss the usage of probabilistic models to identify   *2]   &
             overlapping communities in social network graphs.   CO4

Q 7.   (a) What are the key AI techniques used in Amazon and Netflix recommendation system?   [2.5   CO3
         (b) Discuss various big data applications of hierarchical clustering.   *2]   &
             CO5