

End Semester Exam

First Half - Summer Examination 2022

M.Tech. (Artificial Intelligence)-SEM-II

Course Code: PCEC202 and Course Name: Big Data Analytics

Duration: 03 Hours

Max. Marks: 60

Instructions:

- (1) All questions are compulsory.
- (2) Draw neat diagrams wherever applicable.
- (3) Assume suitable data if required, and state it clearly.

		Max. Marks	CO	BT Level
Q-1	Solve any Six questions out of Eight .	12		
i)	Explain Decision support system (DSS) and its components.	2	CO1	Understand
ii)	What is Distributed file system? List components of HDFS.	2	CO2	Understand
iii)	List any 4 components of hadoop ecosystem.	2	CO3	Understand
iv)	Differentiate between SQL and NOSQL	2	CO4	Understand
v)	Explain Pig - Hadoop ecosystem component.	2	CO4	Understand
vi)	Give importance of Shuffle and sort phase in Mapreduce.	2	CO4	Understand
vii)	Explain significance of data privacy and ethics in data science.	2	CO5	Understand
viii)	Explain Bloom filter with example.	2	CO6	Understand
Q-2	Solve any Four questions out of Six .	16		
i)	Explain Classification, Clustering, Prediction, and Association with example.	4	CO1	Understand
ii)	Write steps of Girvan- Newman Algorithm. Explain clustering of Social-Network Graphs using GN algorithm for	4	CO2	Apply
iii)	Describe Hadoop architecture in detail with its advantages and limitations.	4	CO3	Understand
iv)	Explain Hive and HiveQL with example.	4	CO4	Understand
v)	Discuss anonymization Techniques with example.	4	CO5	Understand
vi)	Discuss Hoeffding tree with its advantages.	4	CO6	Understand
Q-3	Solve any Two questions out of Three .	16		
i)	Business Intelligence: how it is applied in real time applications? Describe with a case study of any one application.	8	CO1	Understand
ii)	Apply Matrix – Vector Multiplication using MapReduce model to solve the following example	8	CO4	Apply
	$\begin{matrix} 3 & 6 & & 2 \\ 4 & 7 & * & 3 \end{matrix}$			

iii)	Explain different types of recommendation systems. Using an example case study describe how it can be used to provide recommendations to users.	8	CO2	Understand
Q-4	Solve any Two questions out of Three .	16		
i)	Apply different measures of similarity/ distance to find a) Jaccard distance between {a, b, e, d} & {b, d, f, g} c) Cosines of the angles between (-1, 2, 3) and (3, 1, -2). d) Hamming Distance between 10011010 & 10101010 e) Edit distance between MUMBAI and BOMBAY.	2 2 2 2	CO2	Apply
ii)	What is Data Privacy? Explain different threats and privacy control techniques.	8	CO5	Understand
iii)	Investigate number of ones in a given data stream using a sliding window based on DGIM algorithm. Also Estimate the number of 1's in last k bits where k =8 and 18. Comment on in each case, how far off the correct value is your estimate? 10110010110101010010101011011	8	CO6	Apply