GUJARAT TECHNOLOGICAL UNIVERSITY
MCA - SEMESTER– V • EXAMINATION – WINTER 2019

Subject Code: 4659306   **Date: 16/11/2019**
Subject Name: Big Data Analytics
Time:10.30 am to 01.00 pm   **Total Marks: 70**
Instructions:
1. **Attempt all questions.**
2. **Make suitable assumptions wherever necessary.**
3. **Figures to the right indicate full marks.**

Q.1	**(a)**	Explain Terms :	
		i) Big Data and BI	**04**
		ii) Clique, Link Prediction, Corpus	**03**
	(b)	Do as directed.	**07**
		i) Say True or False: Smoothing methods are data driven.	
		ii) Say True or False: Collaborative filtering are supervised methods.	
		iii) Say True or False: K-means clustering algorithm automatically finds the best value of k as part of its normal operation.	
		iv) Say True or False: Edge width measures the strength of the relationships between two connected nodes.	
		v) What is Text mining?	
		vi) List any two key skill sets of data scientist.	
		vii) State any one use of Social network analytics.	
Q.2	**(a)**	List phases of Data Analytics Life cycle. Explain data preparation and model building phase with its importance.	**07**
	(b)	i) Define four types of data. Give one example of each.	**04**
		ii) What is Euclidean distance, Manhattan distance and maximum coordinate distance?	**03**

OR

	(b)	i) Discuss drivers of Big Data.	**04**
		ii) Differentiate between Directed vs. Undirected Networks	**03**
Q.3	**(a)**	Compare user-based collaborative filtering and item-based collaborative filtering.	**07**
	(b)	When Association rule is said to be strong? Why?	**07**

OR

Q.3	**(a)**	Explain how Apriori algorithm is used to find Frequent item sets.	**07**
	(b)	Differentiate between	**07**
		i) Collaborative filtering and Association Rules	
		ii) Hierarchical Clustering and Non-Hierarchical Clustering.	
Q.4	**(a)**	State package names available in R to collect data from Twitter and Facebook. What is the importance of social network data in business? List key challenges of it.	**07**
	(b)	i) Consider following 3 statements and Represents words in a term document matrix:	**04**
		S1: this is the first sentence.	
		S2: this is the second sentence.	
		S3: this is the third and final sentence.	
		ii) Why considerable thinking and preprocessing is required in text mining?	**03**

OR

| Q.4 | (a) | What are social networks? How to Visualizing and Analyzing Networks. | 07 |
| | (b) | i) Consider following 3 statements and Explain Tokenization: | 04 |

S1: this is the first sentence.
S2: this is the second sentence.
S3: this is the third and final sentence

03

ii) What is bag-of words approach?

Q.5	(a)	How regression model is used for time series forecasting?	07
	(b)	i) Differentiate between predictive and descriptive time series forecasting	04
		ii) What is an Advanced Exponential Smoothing?	03

OR

Q.5	(a)	Explain moving average method of time series forecasting.	07
	(b)	i) What is autocorrelation and ARIMA Model?	04
		ii) What is difference between data partitioning in cross-sectional and time series data?	03

****