Name: **Enrolment No:**	UPES

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2018

Programme Name: BT-CSE_LLB_CL **Semester : VII**
Course Name : Data warehouse and Data mining **Time : 03 hrs**
Course Code : CSEG411 **Max. Marks : 100**
Nos. of page(s) :2
Instructions:

SECTION A

S. No.		Marks	CO
Q 1	What does subject-oriented data warehouse signify?	**4**	**CO1**
Q 2	How is a data warehouse different from a database?	**4**	**CO2**
Q 3	What is data mining? What kind of data can be mined?	**4**	**CO5**
Q 4	Explain the difference between Data Mining and Data Warehousing?	**4**	**CO2**
Q 5	What is support and confidence in Apriori algorithm.	**4**	**CO4**
	SECTION B		
Q6.	Design a **Snowflake** schema to track the sale for a School, the following dimension tables are: 1.Students 2. Department 3.Account 4.Promotion. List the possible attributes for each of the dimension table	**10**	**CO3**
Q7.	Consider 30 evenly spaced points (with integer features 1 to 30) in 1D and 10 clusters. The current cluster assignments are: points 1,2,3 belong to cluster 1, points 4,5,6 belong to cluster 2, and so on until points 28,29,30 belong to cluster 10. The current cluster centers are c1=1, c2=4, c3=7,…, c10=28. Note these centers are not optimal. Run k-means until convergence.	**10**	**CO4, CO5**
Q8.	What is Web mining? Explain difference between web content mining and web usage mining?	**10**	**CO5**
Q9.	Difference between supervised and unsupervised learning with example.	**10**	**CO4**
	OR		
	Explain some key issues to be considered while planning for a data warehouse.		
	SECTION-C		
Q10	What is classification? Explain KNN algorithm with suitable example.	**20**	**CO4**
	OR		
	Compare the advantages and disadvantages of Decision tree versus K-nearest neighbor classification.		**CO4, CO5**
Q11	The following table consists of training data from customer database. Construct	**20**	**CO4**

decision tree from the given data.

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

Name: Enrolment No:	UPES

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2018

Programme Name: BT-CSE_LLB_CL **Semester : VII**
Course Name : Data warehouse and Data mining **Time : 03 hrs**
Course Code : CSEG411 **Max. Marks : 100**
Nos. of page(s) :2
Instructions:

SECTION A

S. No.		Marks	CO
Q 1	What does subject-oriented data warehouse signify?	**4**	**CO1**
Q 2	Difference between OLTP and OLAP.	**4**	**CO2**
Q 3	What are the advantages of data mining?	**4**	**CO3**
Q 4	Explain the difference between Data Mining and Data Warehousing.	**4**	**CO2**
Q 5	What is Apriori algorithm?	**4**	**CO4**
	SECTION B		
Q6.	Design a **Snowflake** schema to track the sale for a Retail shop, the following dimension tables are: 1.Product 2. Customer 3.Account 4.Promotion. List the possible attributes for each of the dimension table	**10**	**CO3**
Q7.	Consider 30 evenly spaced points (with integer features 1 to 30) in 1D and 10 clusters. The current cluster assignments are: points 1,2,3 belong to cluster 1, points 4,5,6 belong to cluster 2, and so on until points 28,29,30 belong to cluster 10. The current cluster centers are c1=1, c2=4, c3=7,…, c10=28. Note these centers are not optimal. Run k-means until convergence.	**10**	**CO4, CO5**
Q8.	What is Web mining? Explain the following terms: 1. Web content mining 2. Web structure mining 3. Web usage mining	**10**	**CO5**
Q9.	Difference between classification and clustering with example.	**10**	**CO4**
	OR		
	Explain some key issues to be considered while planning for a data warehouse.		
	SECTION-C		
Q10	What is classification? Explain at least 2 algorithms with suitable example.	**20**	**CO4, CO5**

	OR		
	Compare the advantages and disadvantages of Decision tree end K-means algorithm.		**CO4, CO5**
Q11	The following table consists of training data from customer database. Construct decision tree from the given data.	**20**	**CO4**

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no