UKA TARSADIA UNIVERSITY

B.Tech (CE)(Semester 6)
IT5003(2021-22)
Data Warehousing and Data Mining

Date :14/11/2022 **Time :**9:30AM- 12:30PM
Max. Marks:60

Instructions :

1. Attempt all questions.
2. Write each section in a separate answer book.
3. Make suitable assumptions wherever necessary.
4. Draw diagrams/figures whenever necessary.
5. Figures to the right indicate full marks allocated to that question.
6. Follow usual meaning of notations/abbreviations.

SECTION - 1

Q 1 A) Answer the following in brief. (Any 2) **[4]**

I) Explain roll-up operation in multidimensional data modeling.

II) Discuss how data warehouse is subject-oriented.

III) How are OLAP and OLTP different in terms of access patterns?

Q 1 B) Answer the following . (Any 1) **[5]**

I) Explain 3-tier structure of Data warehouse.

II) Explain the implementation of galaxy schema in data warehouse.

Q 2 A) Explain close frequent pattern. [2]

Q 2 B) Which are the limitations of the Apriori algorithm? Explain techniques to improve the efficiency of Apriori algorithm. [7]

OR

Q 2 B) Consider the following transactional dataset and find frequent itemsets using ECLATE algorithm.
Consider minimum support count=2 and minimum confidence=60%.
Find strong association rules.

Transactional ID	Itemset
T1	{A, B, D, E}
T2	{B, C, E}
T3	{A, B, D, E}
T4	{A, B, C, E}
T5	{A, B, C, D, E}
T6	{B, C, D}

Q 3 A) Answer the following in brief. (Any 1) **[2]**

I) Enlist various types of data used in data mining.

II) What is data integration and data transformation?

Q 3 B) Answer the following in detail. (Any 2) **[10]**

I) What is noise? Explain methods to handle noise in dataset.

II) Consider the following data and normalize it using given methods:
200, 300, 400, 600, 1000
i. Min-max normalization by setting min = 0 and max = 1
ii. z-score normalization, where standard deviation of data is 282.84

III) Explain concept hierarchy generation and dimensionality reduction.

SECTION - 2

Q 4 A) Answer the following . (Any 1) **[4]**

I) Explain grid-based methods and hierarchical methods.

II) Enlist the strength and weakness of K-Means algorithm.

Q 4 B) Answer the following. (Any 1) **[5]**

I) Explain DBSCAN algorithm with its advantages and disadvantages.

II) Find out clusters of given objects in below data set using K-Means clustering algorithm for single iteration. Take value of K=2 and O1 and O2 as initial centroids by using Euclidean distance measure.

Data Objects	Height(H)	Weight(W)
O1	185	72
O2	170	56
O3	168	60
O4	179	68
O5	182	72
O6	188	77
O7	180	71
O8	180	70
O9	183	84
O10	180	88

Q 5 A) Explain multimedia data in data mining applications. [2]

Q 5 B) Explain data mining application for customer relationship management (CRM). [7]

OR

Q 5 B) Explain in detail the role and use of data mining in healthcare sector.

Q 6 Answer the following in detail. (Any 2) **[12]**

I) Why naïve bayesian classification is called “naïve”? Describe naïve bayesian classification with example.

II) Explain the steps of the ID3 algorithm for generating decision trees.

III) Explain the methods to measure the accuracy of a classifier or predictor.

UKA TARSADIA UNIVERSITY

B.Tech (CyberSec)/B.Tech (IT) (Semester 5)

IT5003(2021-22)/IT5003(2022-23)

Data Warehousing and Data Mining

Date :14/11/2022 **Time :**9:30AM- 12:30PM

Max. Marks:60

Instructions :

1. Attempt all questions.
2. Write each section in a separate answer book.
3. Make suitable assumptions wherever necessary.
4. Draw diagrams/figures whenever necessary.
5. Figures to the right indicate full marks allocated to that question.
6. Follow usual meaning of notations/abbreviations.

SECTION - 1

Q 1 A) Answer the following in brief. (Any 2) **[4]**

I) Explain drill-down operation in multidimensional data model.

II) Give two points of difference between OLTP and OLAP server.

III) Define the following terms with respect to data warehousing:
a. Time variant
b. Non-volatile

Q 1 B) Explain data warehouse models from the architecture point of view. [5]

OR

Q 1 B) Draw the star schema for online library management system. What is the limitation of star schema?

Q 2 A) Explain Maximal frequent pattern. [2]

Q 2 B) Generate candidate itemsets, frequent itemsets and association rules using Apriori algorithm on the following data set with the minimum support count is 2 and minimum confidence is 70%. [7]

TID	List of items_IDs
T100	I1,I2,I5
T200	I2,I4
T300	I2,I3
T400	I1,I2,I4
T500	I1,I3
T600	I2,I3
T700	I1,I3
T800	I1,I2,I3,I5
T900	I1,I2,I3

OR

Q 2 B) State the Apriori Property. Generate large itemsets and association rules using Apriori algorithm on the following data set with minimum support value and minimum confidence value set as 50% and 75% respectively

TID	Items Purchased
T101	Cheese, Milk, Cookies
T102	Butter, Milk, Bread
T103	Cheese, Butter, Milk, Bread
T104	Butter, Bread

Q 3 A) Answer the following in brief. (Any 1) **[2]**

I) Explain noise in data with example.

II) Consider the following data and calculate mean and median for it.
13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

Q 3 B) Answer the following in detail. (Any 2) **[10]**

I) Enlist functionalities of data mining and explain any two of them.

II) What is the need of data reduction? Explain attribute subset selection method for data reduction.

III) Consider the following 2 x 2 contingency table summarizes sales of game and video for 10,000 customers, where game and non-game refers to the customer who purchase and not purchase games respectively, video and non-video refers to the customer who purchase and not purchase videos respectively.
The actual count for each joint event is given. Find expected frequency for each joint event and calculate chi-square coefficient for the given data.

	Game	Non-game	Total
Video	4000	3500	7500
Non-video	2000	500	2500
Total	6000	4000	10,000

SECTION - 2

Q 4 A) Answer the following in brief. (Any 2) **[4]**

I) Define the dendrogram with an example.

II) Define density-connectivity points.

III) Describe grid-based clustering.

Q 4 B) Answer the following . (Any 1) **[5]**

I) Enlist categories of major fundamental clustering methods and explain any two of them with its merits.

II) Consider the given data and performs one iteration to divide the data into three groups using the k-means clustering method. Take points A, B, and C as initial centroid points.
Data: A (2, 4), B (6, 7), C (9, 10), D (1, 5), E (3, 3) and F(8, 9).
Use Euclidean distance measure to find similarity.

Q 5 A) Enlist types of data required for implementing data mining algorithms. [2]

Q 5 B) Explain the role of data mining in e-commerce websites. [7]

OR

Q 5 B) Explain the role and process of data mining in spam detection.

Q 6 Answer the following in detail. (Any 2) **[12]**

I) What is decision tree? Explain how classification is done using decision tree induction.

II) Explain genetic algorithm and fuzzy set approach for classification.

III) For the below data set, predict the class label for “Single” using naïve Bayesian classifier for the sample X =<year=4, Height=Tall, Pocket Money=Average, Grade= High>.

Year	Height	Pocket Money	Grade	Class: Single
1	Average	Low	High	Yes
2	Tall	Average	Low	No
3	Short	High	High	No
4	Average	Average	Low	No
2	Tall	High	Low	Yes
3	Tall	Low	High	No
3	Average	High	Average	Yes
1	Tall	Average	Average	Yes
4	Short	Average	High	Yes

UKA TARSADIA UNIVERSITY

B.Tech (CE)(Semester 6)
IT5003(2021-22)
Data Warehousing and Data Mining

Date :02/05/2023 **Time :**9:30AM-12:30PM
Max. Marks:60

Instructions :

1. Attempt all questions.
2. Write each section in a separate answer book.
3. Make suitable assumptions wherever necessary.
4. Draw diagrams/figures whenever necessary.
5. Figures to the right indicate full marks allocated to that question.
6. Follow usual meaning of notations/abbreviations.

SECTION - 1

Q 1 A) Answer the following in brief. (Any 2) **[4]**

I) Enlist measures of data cube.

II) What is data cleaning and data extraction?

III) Explain in brief 'Online Analytical Processing'.

Q 1 B) Explain snowflake and fact constellation schema. [5]

OR

Q 1 B) Write full form of OLAP.
Explain the following operations of OLAP:
a) Roll-up
b) Drill-down
c) Slice
d) Pivot

Q 2 A) Define support and confidence in association rule mining. [2]

Q 2 B) Consider the following transactional database consist of the sets. Consider minimum support count=3 and minimum confidence=75%. [7]
Apply FP-growth algorithm and find frequent itemsets.
Find strong association rules.

Transactional ID	Itemset
T1	{M, O, N, K, E, Y }
T2	{D, O, N, K, E, Y}
T3	{M, A, K, E}
T4	{M, U, C, K, Y}
T5	{C, O, O, K, I, E}

OR

Q 2 B) What are the disadvantages of Apriori algorithm? Explain hash-based and partitioning based techniques to improve efficiency of Apriori algorithm.

Q 3 Answer the following in detail. (Any 2) **[12]**

I) Define data mining. Discuss the KDD process with neat sketch.

II) Explain attribute selection methods.

III) Normalize the given data using z-score normalization and decimal scalling methods.
Data: 213, 457, 134, and 367

SECTION - 2

Q 4 A) Answer the following . (Any 1) **[4]**

I) Enlist general characteristics of density based clustering.

II) Differentiate between AGNES and DIANA algorithms.

Q 4 B) Answer the following. (Any 1) **[5]**

I) Assume that the number of clusters (k=3). Use k-medoids clustering to cluster the following set of data: P1(0, 2), P2(5, 0), P3(7, 3), P4(0, 5), P5(3, 1), P6(5, 2). Consider P1, P2 and P3 as intital medoids and calculate for two iterations. Use euclidean distance measure for the distance calculation.

II) Write k-means clustering algorithm with its limitations.

Q 5 A) Define the terms term frequency and document frequency in text mining. [3]

Q 5 B) Explain sequence data mining. [6]

OR

Q 5 B) Explain the use of data mining in the financial sector.

Q 6 Answer the following in detail. (Any 2) **[12]**

I) Explain k-nearest neighbor(KNN) classification.

II) Answer the following:
a) Define: Supervised learning and unsupervised learning.
b) What is the role of attribute selection in decision tree classifier?
c) Explain limitation of rule based classifier.

III) Define Linear regression.
For the following data:
a) Find the regression coefficient values.
b) Find the Y value for the given value of X=42 using the least square method for the given data.

X	Y
35	25
11	33
20	45
20	23
44	77

UKA TARSADIA UNIVERSITY

B.Tech (IT) (Semester 5)
IT5003(2021-22)
Data Warehousing and Data Mining

Date :11/05/2023 **Time :**1:30PM- 4:30PM
Max. Marks:60

Instructions :

1. Attempt all questions.
2. Write each section in a separate answer book.
3. Make suitable assumptions wherever necessary.
4. Draw diagrams/figures whenever necessary.
5. Figures to the right indicate full marks allocated to that question.
6. Follow usual meaning of notations/abbreviations.

SECTION - 1

Q 1 A) Define data warehouse. State the need of data warehouse. [3]

Q 1 B) Explain multi-tier data warehouse architecture. [6]

OR

Q 1 B) Define multidimentional model. State the difference between OLAP and OLTP.

Q 2 A) Answer the following . (Any 1) **[4]**

I) What is association rule mining?
Consider the following dataset. For which {O, K, E} is the frequent itemset. Find the strong rules from given frequent itemset.
Let minimum support= 60% and minimum confidence = 80%.

Transactional ID	Itemset
T1	{M, O, N, K, E, Y }
T2	{D, O, N, K, E, Y}
T3	{M, A, K, E}
T4	{M, U, C, K, Y}
T5	{C, O, O, K, I, E}

II) Answer the following:
a) What are close and maximal frequent patterns?
b) Enlist the advantage and disadvantage of FP-Growth method.

Q 2 B) Answer the following. (Any 1) **[5]**

I) Define: Frequent itemset.
Consider the following dataset of a grocery store transaction. Let minimum support count=3.
Find all frequent itemsets using ECLAT method.

Transaction ID	Items Bought
T1	Milk, Bread, Cheeze
T2	Bread, Jam
T3	Bread, Butter
T4	Milk, Bread, Jam
T5	Milk, Butter
T6	Bread, Butter
T7	Milk, Butter
T8	Milk, Bread, Butter, Cheeze
T9	Milk, Bread, Butter

II) Enlist techniques to improve the efficiency of Apriori algorithm and explain any two techniques from it.

Q 3 A) Answer the following in brief. (Any 1) **[2]**

I) What is pattern evaluation and data data mining?

II) Enlist any four advanced datasets.

Q 3 B) Answer the following in detail. (Any 2) **[10]**

I) Explain Knowledge Discovery from Data (KDD) process.

II) Discuss one application of data mining in detail.

III) Suppose a group of age records has been sorted as follows:
8, 10, 12, 13, 20, 38, 52, 55, 72, 72, 80, 95.
Partition them into three bins by equal-frequency partitioning.
Apply following smoothing techniques to each bin:
i. Bin-mean
ii. Bin-boundary

SECTION - 2

Q 4 A) Describe grid based clustering. [3]

Q 4 B) Explain hirarchical clusterng methods. [6]

OR

Q 4 B) Enlist and explain different clustering approaches.

Q 5 A) Which are different types of data used in data mining? [2]

Q 5 B) Explain any two data mining applications for retail industry. [7]

OR

Q 5 B) Explain any two data mining applications for telecommunication sector.

Q 6 Answer the following in detail. (Any 2) **[12]**

I) Explain tree pruning process of classification.

II) Define confusion matrix. Explain it with an example.

III) For the given dataset, use Naïve baye's algorithm and find the class for the sample: X=<Outlook=Rain, Temperature =Cold, Humidity= High>

Outlook	Temperature	Humidity	Play_Tennis
Sunny	Hot	High	No
Sunny	Hot	High	No
Overcast	Hot	High	Yes
Rain	Mild	High	Yes
Rain	Cool	Normal	Yes
Rain	Cool	Normal	No
Overcast	Cool	Normal	Yes
Sunny	Mild	High	No
Sunny	Cool	Normal	Yes
Rain	Mild	Normal	Yes

UKA TARSADIA UNIVERSITY

B.Tech (CE) (Semester 6)
IT5003(2021-22)
Data Warehousing and Data Mining

Date :01/12/2023 **Time :**9:30AM- 12:30PM
Max. Marks:60

Instructions :

1. Attempt all questions.
2. Write each section in a separate answer book.
3. Make suitable assumptions wherever necessary.
4. Draw diagrams/figures whenever necessary.
5. Figures to the right indicate full marks allocated to that question.
6. Follow usual meaning of notations/abbreviations.

SECTION - 1

Q 1 A) Enlist measures of data cube. [3]

Q 1 B) Explain multi-tier data warehouse architecture. [6]

OR

Q 1 B) Explain ETL function and metadata repository.

Q 2 A) Define the following terms: [2]
a. Close itemset
b. Maximal frequent patterns.

Q 2 B) Consider the following transactional dataset and find out frequent itemsets using the Apriori algorithm where minimum support count=3 and minimum confidence=75%. Find strong association rules. [7]

Transactions_ID	Item_set
T1	{M, O, N, K, E, Y }
T2	{D, O, N, K, E, Y}
T3	{M, A, K, E}
T4	{M, U, C, K, Y}
T5	{C, O, O, K, I, E}

OR

Q 2 B) Consider the following transactional dataset and find frequent itemsets using ECLAT algorithm.
Consider minimum support count=2 and minimum confidence=60%. Find strong association rules.

Transactions_ID	Item_set
T1	{A, B, C, D}
T2	{B, C, E}
T3	{A, B, C, D}
T4	{A, B, D, E}
T5	{A, B, C, D, E}
T6	{B, C, D}

Q 3 Answer the following in detail. (Any 2) **[12]**

I) Enlist data mining functionalities and explain any two of them.

II) Explain the data integration methods for numerical data.

III) Consider the given data and partition them into 3 bins using equal-width method and then apply smoothing by bin mean and smoothing by bin boundary methods.
Data: 12, 4, 22, 3, 3, 12, 2, 5, 6, 3, 2, 3, 5, 25

SECTION - 2

Q 4 A) Answer the following . (Any 1) **[4]**

I) Explain how k-mean clustering method differs from k-medoid clustering method.

II) What is clustering? What are different types of clustering?

Q 4 B) Answer the following. (Any 1) **[5]**

I) Calculate two clusters using k-means cluster algorithm for two iteration. For finding the distance use euclidean distance. Assume centroid 1 as subject1 and centroid 2 as subject4.

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

II) Explain density based clustering with its merits and demerits.

Q 5 A) Give any two examples of time-series data. [2]

Q 5 B) Explain in detail web mining. [7]

OR

Q 5 B) Explain role and use of data mining in financial data analysis.

Q 6 Answer the following in detail. (Any 2) **[12]**

I) Answer the following:
a) Define: Supervised learning and unsupervised learning.
b) What is the role of attribute selection in decision tree classifier?
c) Explain limitation of rule based classifier.

II) What is decision tree? Explain how classification is done using decision tree induction.

III) Explain tree pruning process of classification.

UKA TARSADIA UNIVERSITY

B.Tech (CyberSec)/B.Tech (IT) (Semester 5)
IT5003(2021-22)/IT5003(2022-23)
Data Warehousing and Data Mining

Date :01/12/2023 **Time :**9:30AM- 12:30PM
Max. Marks:60

Instructions :

1. Attempt all questions.
2. Write each section in a separate answer book.
3. Make suitable assumptions wherever necessary.
4. Draw diagrams/figures whenever necessary.
5. Figures to the right indicate full marks allocated to that question.
6. Follow usual meaning of notations/abbreviations.

SECTION - 1

Q 1 A) Answer the following in brief. (Any 2) **[4]**

I) Explain pivot operation in multidimensional data model with example.

II) Give the difference between star schema and snowflake schema.

III) Explain the role of concept hierarchies.

Q 1 B) Explain the measures of data cube for categorization and computation purpose. [5]

OR

Q 1 B) Enlist and explain key features of data warehousing.

Q 2 A) Enlist the properties of Apriori algorithm. [2]

Q 2 B) Generate candidate itemsets, frequent itemsets and association rules using F-P growth algorithm on the following data set with the minimum support count is 2 and minimum confidence is 70%. [7]

Transactions_ID	Item_set
T1	I1, I2, I5
T2	I2, I4
T3	I2, I3
T4	I1, I2, I4
T5	I1, I3
T6	I2, I3
T7	I1, I3
T8	I1, I2, I3, I5
T9	I1, I2, I3

OR

Q 2 B) Consider the following dataset of a grocery store transaction. Let minimum support count=3. Find all frequent itemsets using ECLAT method.

Transactions_ID	Item_set
T1	Peas, Tuvar, Moongdal
T2	Tuvar, Chickpeas
T3	Tuvar, Greenpeas
T4	Peas, Tuvar, Chickpeas
T5	Peas, Greenpeas
T6	Tuvar, Greenpeas
T7	Peas, Greenpeas
T8	Peas, Tuvar, Greenpeas, Moongdal
T9	Peas, Tuvar, Greenpeas

Q 3 Answer the following in detail. (Any 2) **[12]**

I) Discuss all the steps of data cleaning.

II) Write a note on Data Integration.

III) Consider the given data and partition them into three bins using equal-width method and then apply smoothing by bin mean and smoothing by bin boundary methods.
Data: 10, 6, 12, 4, 4, 12, 2, 5, 6, 7, 2, 7, 5, 26, 30

SECTION - 2

Q 4 A) Describe grid based clustering. [3]

Q 4 B) Enlist and explain different clustering approaches. [6]

OR

Q 4 B) Write k-means clustering algorithm with its limitations.

Q 5 A) What is data mining? [2]

Q 5 B) Explain the procedure for spam detection using concepts of data mining. [7]

OR

Q 5 B) Explain any two applications of data mining in an e-commerce website.

Q 6 Answer the following in detail. (Any 2) **[12]**

I) Explain k-nearest neighbor(KNN) classification.

II) Explain genetic algorithm and fuzzy set approach for classification.

III) Explain the major steps of decision tree classification.

UKA TARSADIA UNIVERSITY

B.Tech (IT) (Semester 5)
IT5003(2021-22)
Data Warehousing and Data Mining

Date :11/05/2023 **Time :**1:30PM- 4:30PM
Max. Marks:60

Instructions :

1. Attempt all questions.
2. Write each section in a separate answer book.
3. Make suitable assumptions wherever necessary.
4. Draw diagrams/figures whenever necessary.
5. Figures to the right indicate full marks allocated to that question.
6. Follow usual meaning of notations/abbreviations.

SECTION - 1

Q 1 A) Define data warehouse. State the need of data warehouse. [3]

Q 1 B) Explain multi-tier data warehouse architecture. [6]

OR

Q 1 B) Define multidimentional model. State the difference between OLAP and OLTP.

Q 2 A) Answer the following . (Any 1) **[4]**

I) What is association rule mining?
Consider the following dataset. For which {O, K, E} is the frequent itemset. Find the strong rules from given frequent itemset.
Let minimum support= 60% and minimum confidence = 80%.

Transactional ID	Itemset
T1	{M, O, N, K, E, Y }
T2	{D, O, N, K, E, Y}
T3	{M, A, K, E}
T4	{M, U, C, K, Y}
T5	{C, O, O, K, I, E}

II) Answer the following:
a) What are close and maximal frequent patterns?
b) Enlist the advantage and disadvantage of FP-Growth method.

Q 2 B) Answer the following. (Any 1) **[5]**

I) Define: Frequent itemset.
Consider the following dataset of a grocery store transaction. Let minimum support count=3.
Find all frequent itemsets using ECLAT method.

Transaction ID	Items Bought
T1	Milk, Bread, Cheeze
T2	Bread, Jam
T3	Bread, Butter
T4	Milk, Bread, Jam
T5	Milk, Butter
T6	Bread, Butter
T7	Milk, Butter
T8	Milk, Bread, Butter, Cheeze
T9	Milk, Bread, Butter

II) Enlist techniques to improve the efficiency of Apriori algorithm and explain any two techniques from it.

Q 3 A) Answer the following in brief. (Any 1) **[2]**

I) What is pattern evaluation and data data mining?

II) Enlist any four advanced datasets.

Q 3 B) Answer the following in detail. (Any 2) **[10]**

I) Explain Knowledge Discovery from Data (KDD) process.

II) Discuss one application of data mining in detail.

III) Suppose a group of age records has been sorted as follows:
8, 10, 12, 13, 20, 38, 52, 55, 72, 72, 80, 95.
Partition them into three bins by equal-frequency partitioning.
Apply following smoothing techniques to each bin:
i. Bin-mean
ii. Bin-boundary

SECTION - 2

Q 4 A) Describe grid based clustering. [3]

Q 4 B) Explain hirarchical clusterng methods. [6]

OR

Q 4 B) Enlist and explain different clustering approaches.

Q 5 A) Which are different types of data used in data mining? [2]

Q 5 B) Explain any two data mining applications for retail industry. [7]

OR

Q 5 B) Explain any two data mining applications for telecommunication sector.

Q 6 Answer the following in detail. (Any 2) **[12]**

I) Explain tree pruning process of classification.

II) Define confusion matrix. Explain it with an example.

III) For the given dataset, use Naïve baye's algorithm and find the class for the sample: X=<Outlook=Rain, Temperature =Cold, Humidity= High>

Outlook	Temperature	Humidity	Play_Tennis
Sunny	Hot	High	No
Sunny	Hot	High	No
Overcast	Hot	High	Yes
Rain	Mild	High	Yes
Rain	Cool	Normal	Yes
Rain	Cool	Normal	No
Overcast	Cool	Normal	Yes
Sunny	Mild	High	No
Sunny	Cool	Normal	Yes
Rain	Mild	Normal	Yes

UKA TARSADIA UNIVERSITY

B.Tech (CE)(Semester 6)
IT5003(2021-22)
Data Warehousing and Data Mining

Date :02/05/2023 **Time :**9:30AM-12:30PM
Max. Marks:60

Instructions :

1. Attempt all questions.
2. Write each section in a separate answer book.
3. Make suitable assumptions wherever necessary.
4. Draw diagrams/figures whenever necessary.
5. Figures to the right indicate full marks allocated to that question.
6. Follow usual meaning of notations/abbreviations.

SECTION - 1

Q 1 A) Answer the following in brief. (Any 2) **[4]**

I) Enlist measures of data cube.

II) What is data cleaning and data extraction?

III) Explain in brief 'Online Analytical Processing'.

Q 1 B) Explain snowflake and fact constellation schema. [5]

OR

Q 1 B) Write full form of OLAP.
Explain the following operations of OLAP:
a) Roll-up
b) Drill-down
c) Slice
d) Pivot

Q 2 A) Define support and confidence in association rule mining. [2]

Q 2 B) Consider the following transactional database consist of the sets. Consider minimum support count=3 and minimum confidence=75%. [7]
Apply FP-growth algorithm and find frequent itemsets.
Find strong association rules.

Transactional ID	Itemset
T1	{M, O, N, K, E, Y }
T2	{D, O, N, K, E, Y}
T3	{M, A, K, E}
T4	{M, U, C, K, Y}
T5	{C, O, O, K, I, E}

OR

Q 2 B) What are the disadvantages of Apriori algorithm? Explain hash-based and partitioning based techniques to improve efficiency of Apriori algorithm.

Q 3 Answer the following in detail. (Any 2) **[12]**

I) Define data mining. Discuss the KDD process with neat sketch.

II) Explain attribute selection methods.

III) Normalize the given data using z-score normalization and decimal scalling methods.
Data: 213, 457, 134, and 367

SECTION - 2

Q 4 A) Answer the following . (Any 1) **[4]**

I) Enlist general characteristics of density based clustering.

II) Differentiate between AGNES and DIANA algorithms.

Q 4 B) Answer the following. (Any 1) **[5]**

I) Assume that the number of clusters (k=3). Use k-medoids clustering to cluster the following set of data: P1(0, 2), P2(5, 0), P3(7, 3), P4(0, 5), P5(3, 1), P6(5, 2). Consider P1, P2 and P3 as intital medoids and calculate for two iterations. Use euclidean distance measure for the distance calculation.

II) Write k-means clustering algorithm with its limitations.

Q 5 A) Define the terms term frequency and document frequency in text mining. [3]

Q 5 B) Explain sequence data mining. [6]

OR

Q 5 B) Explain the use of data mining in the financial sector.

Q 6 Answer the following in detail. (Any 2) **[12]**

I) Explain k-nearest neighbor(KNN) classification.

II) Answer the following:
a) Define: Supervised learning and unsupervised learning.
b) What is the role of attribute selection in decision tree classifier?
c) Explain limitation of rule based classifier.

III) Define Linear regression.
For the following data:
a) Find the regression coefficient values.
b) Find the Y value for the given value of X=42 using the least square method for the given data.

X	Y
35	25
11	33
20	45
20	23
44	77

UKA TARSADIA UNIVERSITY

B.Tech (CyberSec)/B.Tech (IT) (Semester 5)
IT5003(2021-22)/IT5003(2022-23)
Data Warehousing and Data Mining

Date :01/12/2023 **Time :**9:30AM- 12:30PM
Max. Marks:60

Instructions :

1. Attempt all questions.
2. Write each section in a separate answer book.
3. Make suitable assumptions wherever necessary.
4. Draw diagrams/figures whenever necessary.
5. Figures to the right indicate full marks allocated to that question.
6. Follow usual meaning of notations/abbreviations.

SECTION - 1

Q 1 A) Answer the following in brief. (Any 2) **[4]**

I) Explain pivot operation in multidimensional data model with example.

II) Give the difference between star schema and snowflake schema.

III) Explain the role of concept hierarchies.

Q 1 B) Explain the measures of data cube for categorization and computation purpose. [5]

OR

Q 1 B) Enlist and explain key features of data warehousing.

Q 2 A) Enlist the properties of Apriori algorithm. [2]

Q 2 B) Generate candidate itemsets, frequent itemsets and association rules using F-P growth algorithm on the following data set with the minimum support count is 2 and minimum confidence is 70%. [7]

Transactions_ID	Item_set
T1	I1, I2, I5
T2	I2, I4
T3	I2, I3
T4	I1, I2, I4
T5	I1, I3
T6	I2, I3
T7	I1, I3
T8	I1, I2, I3, I5
T9	I1, I2, I3

OR

Q 2 B) Consider the following dataset of a grocery store transaction. Let minimum support count=3. Find all frequent itemsets using ECLAT method.

Transactions_ID	Item_set
T1	Peas, Tuvar, Moongdal
T2	Tuvar, Chickpeas
T3	Tuvar, Greenpeas
T4	Peas, Tuvar, Chickpeas
T5	Peas, Greenpeas
T6	Tuvar, Greenpeas
T7	Peas, Greenpeas
T8	Peas, Tuvar, Greenpeas, Moongdal
T9	Peas, Tuvar, Greenpeas

Q 3 Answer the following in detail. (Any 2) **[12]**

I) Discuss all the steps of data cleaning.

II) Write a note on Data Integration.

III) Consider the given data and partition them into three bins using equal-width method and then apply smoothing by bin mean and smoothing by bin boundary methods.
Data: 10, 6, 12, 4, 4, 12, 2, 5, 6, 7, 2, 7, 5, 26, 30

SECTION - 2

Q 4 A) Describe grid based clustering. [3]

Q 4 B) Enlist and explain different clustering approaches. [6]

OR

Q 4 B) Write k-means clustering algorithm with its limitations.

Q 5 A) What is data mining? [2]

Q 5 B) Explain the procedure for spam detection using concepts of data mining. [7]

OR

Q 5 B) Explain any two applications of data mining in an e-commerce website.

Q 6 Answer the following in detail. (Any 2) **[12]**

I) Explain k-nearest neighbor(KNN) classification.

II) Explain genetic algorithm and fuzzy set approach for classification.

III) Explain the major steps of decision tree classification.

UKA TARSADIA UNIVERSITY

B.Tech (CE) (Semester 6)
IT5003(2021-22)
Data Warehousing and Data Mining

Date :01/12/2023 **Time :**9:30AM- 12:30PM
Max. Marks:60

Instructions :

1. Attempt all questions.
2. Write each section in a separate answer book.
3. Make suitable assumptions wherever necessary.
4. Draw diagrams/figures whenever necessary.
5. Figures to the right indicate full marks allocated to that question.
6. Follow usual meaning of notations/abbreviations.

SECTION - 1

Q 1 A) Enlist measures of data cube. [3]

Q 1 B) Explain multi-tier data warehouse architecture. [6]

OR

Q 1 B) Explain ETL function and metadata repository.

Q 2 A) Define the following terms: [2]
a. Close itemset
b. Maximal frequent patterns.

Q 2 B) Consider the following transactional dataset and find out frequent itemsets using the Apriori algorithm where minimum support count=3 and minimum confidence=75%. Find strong association rules. [7]

Transactions_ID	Item_set
T1	{M, O, N, K, E, Y }
T2	{D, O, N, K, E, Y}
T3	{M, A, K, E}
T4	{M, U, C, K, Y}
T5	{C, O, O, K, I, E}

OR

Q 2 B) Consider the following transactional dataset and find frequent itemsets using ECLAT algorithm.
Consider minimum support count=2 and minimum confidence=60%. Find strong association rules.

Transactions_ID	Item_set
T1	{A, B, C, D}
T2	{B, C, E}
T3	{A, B, C, D}
T4	{A, B, D, E}
T5	{A, B, C, D, E}
T6	{B, C, D}

Q 3 Answer the following in detail. (Any 2) **[12]**

I) Enlist data mining functionalities and explain any two of them.

II) Explain the data integration methods for numerical data.

III) Consider the given data and partition them into 3 bins using equal-width method and then apply smoothing by bin mean and smoothing by bin boundary methods.
Data: 12, 4, 22, 3, 3, 12, 2, 5, 6, 3, 2, 3, 5, 25

SECTION - 2

Q 4 A) Answer the following . (Any 1) **[4]**

I) Explain how k-mean clustering method differs from k-medoid clustering method.

II) What is clustering? What are different types of clustering?

Q 4 B) Answer the following. (Any 1) **[5]**

I) Calculate two clusters using k-means cluster algorithm for two iteration. For finding the distance use euclidean distance. Assume centroid 1 as subject1 and centroid 2 as subject4.

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

II) Explain density based clustering with its merits and demerits.

Q 5 A) Give any two examples of time-series data. [2]

Q 5 B) Explain in detail web mining. [7]

OR

Q 5 B) Explain role and use of data mining in financial data analysis.

Q 6 Answer the following in detail. (Any 2) **[12]**

I) Answer the following:
a) Define: Supervised learning and unsupervised learning.
b) What is the role of attribute selection in decision tree classifier?
c) Explain limitation of rule based classifier.

II) What is decision tree? Explain how classification is done using decision tree induction.

III) Explain tree pruning process of classification.

UKA TARSADIA UNIVERSITY

B.Tech (CE)(Semester 6)
IT5003(2021-22)
Data Warehousing and Data Mining

Date :24/04/2024 **Time :**9:30AM-12:30PM
Max. Marks:60

Instructions :

1. Attempt all questions.
2. Write each section in a separate answer book.
3. Make suitable assumptions wherever necessary.
4. Draw diagrams/figures whenever necessary.
5. Figures to the right indicate full marks allocated to that question.
6. Follow usual meaning of notations/abbreviations.

SECTION - 1

Q 1 A) Describe ETL process of data warehouse. [3]

Q 1 B) Explain OLAP operations with an example. [6]

OR

Q 1 B) Enlist data warehouse schemas and explain anyone with an example.

Q 2 A) Define the following terms: [3]
- Frequent itemset
- Closed itemset
- Maximal itemset

Q 2 B) Discuss vertical data format and explain ECLAT algorithm with an example. [6]

OR

Q 2 B) Consider the following dataset and find frequent itemset and association rules satisfying the minimum threshold as 2 and minimum confidence as 70% using any association rule mining method.

Transaction	List of items
T1	M, T, C
T2	E, T, D
T3	M, E, T, D
T4	E, D
T5	J

Q 3 A) Answer the following in brief. (Any 1) **[2]**

I) Enlist any two challenges in data mining.

II) Describe the final stage of data mining life cycle.

Q 3 B) Answer the following in detail. (Any 2) **[10]**

I) Explain data cube aggregation.

II) What is the need of data pre-processing? Explain each step of data pre-processing.

III) Explain following functionalities of data mining.

SECTION - 2

Q 4 A) Describe requirements of cluster analysis. [3]

Q 4 B) Explain density based clustering method with an example. [6]

OR

Q 4 B) Partition the given data into 2 clusters using k-means clustering algorithm. Consider points A and C as initial cluster centroids.
Data: A(1,5), B(2,6), C(5,3), D(4,9) and E(3,7)

Q 5 A) Explain multimedia data in data mining applications. [3]

Q 5 B) Explain sequence data mining. [6]

OR

Q 5 B) Explain use and concept of data mining with the help of any banking application.

Q 6 Answer the following in detail. (Any 2) **[12]**

I) What are the differences between classification and prediction? Explain the idea of prediction in detail.

II) Provide an explanation for each of the following:
a) Supervised learning and unsupervised learning
b) The significance of attribute selection in decision tree classifiers
c) Limitations associated with rule-based classifiers

III) What is rule-based classification? How does it work in the context of data mining?

UKA TARSADIA UNIVERSITY

B.Tech (CyberSec)/B.Tech (IT) (Semester 5)
IT5003(2021-22)/IT5003(2022-23)
Data Warehousing and Data Mining

Date :15/05/2024 **Time :**1:30PM- 4:30PM
Max. Marks:60

Instructions :

1. Attempt all questions.
2. Write each section in a separate answer book.
3. Make suitable assumptions wherever necessary.
4. Draw diagrams/figures whenever necessary.
5. Figures to the right indicate full marks allocated to that question.
6. Follow usual meaning of notations/abbreviations.

SECTION - 1

Q 1 A) Define data warehouse. State the need of data warehouse. [3]

Q 1 B) Explain the difference between OLAP and OLTP server. [6]

OR

Q 1 B) Explain three tier data warehouse architecture with a diagram.

Q 2 A) Define the following terms with respect to association rule mining: [2]
a. Support
b. Confidence

Q 2 B) State the Apriori Property. Generate large itemsets and association rules using Apriori algorithm on the following data set with minimum support value and minimum confidence value set as 50% and 75% respectively. [7]

TID	Items Purchased
T101	Laptop, mouse, pendrive
T102	Harddisk, mouse, USBcable
T103	Laptop, mouse, harddisk, USBcable
T104	Harddisk, USBcable
T105	Laptop, pendrive

OR

Q 2 B) Explain the concept of market basket analysis using an example.

Q 3 Answer the following in detail. (Any 2) **[12]**

I) Define data mining. Explain the KDD process with neat sketch.

II) Consider the following data and find the Pearson's correlation coefficient between X and Y.

X	Y
3.1	8
3.5	12
4.2	11
4.6	7
5.1	10

III) Consider the following data and normalize it using given methods: 20, 30, 40, 60, 100
a. Min-max normalization by setting min = 0 and max = 1
b. Z-score normalization, where standard deviation of data is 28.284

SECTION - 2

Q 4 A) Answer the following . (Any 1) **[4]**

I) Write an algorithm of K-Means clustering.

II) Explain the following terms with an example:
1. Cluster Analysis
2. Data Segmentation

Q 4 B) Answer the following. (Any 1) **[5]**

I) Explain K-Medoids algorithm with its advantages and disadvantages.

II) Write the short note on the following:
1. Agglomerative hierarchical clustering
2. Divisive hierarchical clustering

Q 5 A) Enlist any two applications of data mining where world wide web data can be used. [2]

Q 5 B) Write a short note on customer relationship management. [7]

OR

Q 5 B) Explain the use of data mining in the financial sector.

Q 6 Answer the following in detail. (Any 2) **[12]**

I) Enlist and explain techniques used to estimate accuracy of classification model.

II) What is classification and prediction?
For the following dataset, use Naïve Bayesian classifier and predict the class label “Stolen” for the given sample X = <Color = Red, Type = SUV, Origin = Domestic>

Color	Type	Origin	Class: Stolen
Red	Sports	Domestic	Yes
Red	Sports	Domestic	No
Red	Sports	Domestic	Yes
White	Sports	Domestic	No
Yellow	Sports	Import	Yes
Yellow	SUV	Import	No
Yellow	SUV	Import	Yes
White	SUV	Domestic	No
Red	SUV	Import	No
Red	Sports	Import	Yes

III) Write a short note on the following with respect to the classification:
a) Case based reasoning
b) Fuzzy logic

UKA TARSADIA UNIVERSITY

B.Tech (IT) (Semester 5)
IT5003(2021-22)
Data Warehousing and Data Mining

Date :27/11/2024 **Time :**9:30AM- 12:30PM
Max. Marks:60

Instructions :

1. Attempt all questions.
2. Write each section in a separate answer book.
3. Make suitable assumptions wherever necessary.
4. Draw diagrams/figures whenever necessary.
5. Figures to the right indicate full marks allocated to that question.
6. Follow usual meaning of notations/abbreviations.

SECTION - 1

Q 1 A) What is fact constellations schema? Explain it with example. [3]

Q 1 B) Describe the multidimensional data model used in data warehousing. Explain the concepts of data cubes, star schema including examples for each. [6]

OR

Q 1 B) Define the role of the metadata repository in a data warehouse. How does metadata support the management and usage of data within a data warehouse environment?

Q 2 A) Answer the following . (Any 1) **[4]**

I) Consider the following dataset. For which {I1, I2, I3} and {I1, I2, I5} are the frequent itemsets. Find the strong rules for given frequent itemsets. Let minimum support count=2 and minimum confidence = 70%.

Transaction ID	Items
T1	{I1, I2, I5}
T2	{I2, I4}
T3	{I2, I3}
T4	{I1, I2, I4}
T5	{I1, I3}
T6	{I2, I3}
T7	{I1, I3}
T8	{I1, I2, I3, I5}
T9	{I1, I2, I3}

II) Explain hash-based and sampling techniques to improve the efficiency of Apriori algorithm.

Q 2 B) Answer the following. (Any 1) **[5]**

I) Explain market basket analysis with an example.

II) What is the limitation of Apriori algorithm?
Consider the following transactional database consisting of the sets. Let minimum support=60%. Find frequent itemsets using Apriori algorithm.

Transactional ID	Itemset
T1	{M, O, N, K, E, Y }
T2	{D, O, N, K, E, Y}
T3	{M, A, K, E}
T4	{M, U, C, K, Y}
T5	{C, O, O, K, I, E}

Q 3 Answer the following in detail. (Any 2) **[12]**

I) What are the steps involved in the Knowledge Discovery from Data (KDD) process, and provide an explanation of each step.

II) Suppose for a group of products, its price has been recorded as follows:
4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34
Partition them into three bins by equal-frequency method. Apply Bin-mean and Bin-boundary to remove noise.

III) What does data cleaning involve, and describe three techniques used to replace missing values in a dataset?

SECTION - 2

Q 4 A) Answer the following . (Any 1) **[4]**

I) Explain the general characteristics of density based clustering.

II) Give the difference between agglomerative and divisive hierarchical clustering.

Q 4 B) Answer the following. (Any 1) **[5]**

I) Explain DBSCAN algorithm with its advantages and disadvantages.

II) Consider the given data and performs one iteration to divide the data into three groups using the k-means clustering method. Take points A, B, and C as initial centroid points. Data: A (2, 10), B (6, 12), C (5, 10), D (2, 5), E (3, 7) and F (9, 8). Use Euclidean distance measure to find similarity.

Q 5 A) Identify and describe two applications of data mining that utilize data from the World Wide Web. [3]

Q 5 B) Discuss how data mining techniques are used in fraud detection within the banking and financial sectors. Provide real-world examples to support your answer. [6]

OR

Q 5 B) Discuss the applications of data mining in the telecommunications industry. Explain how data mining techniques are utilized to improve customer service.

Q 6 Answer the following in detail. (Any 2) **[12]**

I) Discuss the Genetic Algorithm and Fuzzy Set Approach as classification methods in data mining. Explain how each method works.

II) Describe the steps involved in the ID3 algorithm for constructing decision trees. Explain the concept of information gain.

III) What is Bayesian classification? Explain the Naive Bayes classifier and how it applies Bayes' theorem for classification tasks. Provide an example to illustrate its application.

UKA TARSADIA UNIVERSITY

B.Tech (CyberSec) (Semester 5)
IT5003(2022-23)
Data Warehousing and Data Mining

Date :25/11/2024 **Time :**9:30AM- 12:30PM
Max. Marks:60

Instructions :

1. Attempt all questions.
2. Write each section in a separate answer book.
3. Make suitable assumptions wherever necessary.
4. Draw diagrams/figures whenever necessary.
5. Figures to the right indicate full marks allocated to that question.
6. Follow usual meaning of notations/abbreviations.

SECTION - 1

Q 1 A) What is a data warehouse, and what are its main characteristics and functionalities? [3]

Q 1 B) How do OLAP (Online Analytical Processing) and OLTP (Online Transaction Processing) differ from each other in their functions and purposes? [6]

OR

Q 1 B) Provide a brief overview of the Star Schema and Snowflake Schema data modeling techniques used in data warehousing?

Q 2 A) Enlist the advantages of Apriori algorithm. [2]

Q 2 B) Consider the following fruits dataset and find frequent itemsets using ECLATE algorithm. Consider minimum support count=2 and minimum confidence=50%. Find strong association rules. [7]

Transactional_ID	Itemset
T1	Mango, apple, banana, orange
T2	Apple, chickoo, orange
T3	Mango, apple, banana, orange
T4	Mango, apple, chickoo, orange
T5	Mango, apple, chickoo, banana, orange
T6	Apple, chickoo, banana
T7	Chickoo, banana, orange

OR

Q 2 B) Explain hash-based and partitioning based techniques to improve efficiency of Apriori algorithm.

Q 3 Answer the following in detail. (Any 2) **[12]**

I) Explain the concepts of discretization and concept hierarchy generation in data preprocessing. How do these techniques help in simplifying the data for analysis?Provide examples to illustrate your points.

II) Consider two stocks, X and Y, with the following weekly closing prices:
Stock X: 4, 7, 9, 6, 8 and Stock Y: 10, 12, 15, 14, 17
a) Calculate the covariance between the two stocks X and Y.
b) Based on the calculated covariance, explain whether the prices of stocks X and Y are likely to rise or fall together.

III) What is the significance of data pre-processing in data mining? Discuss each step involved in the data pre-processing process and explain how each step contributes to improving data quality for analysis.

SECTION - 2

Q 4 A) Define the dendrogram.Explain its usage with an application. [3]

Q 4 B) Partition the given data set int ot 3 clusters using K-Means clustering algorithm. [6]
Consider data points are: P1(1,3) , P2(2,2) , P3(5,8) , P4(8,5) ,P5(3,9) , P6(10,7) , P7(3,3) , P8(9,4) , P9(3,7).
Assume Initial cluster centers are P7(3,3), P9(3,7), P8(9,4) as C1, C2, C3.

OR

Q 4 B) Define desity based clustering.Explain the terms used in DBSCAN algorithm -noise,boundry and core point with example.

Q 5 A) Give any three examples of time-series data. [3]

Q 5 B) Describe the role and use of data mining in fraud detection. [6]

OR

Q 5 B) Explain any two applications of data mining in an e-commerce website.

Q 6 Answer the following in detail. (Any 2) **[12]**

I) Explain “case based reasoning” classifier.

II) Explain genetic algorithm for classification.

III) For the following dataset of “Stolen”, Find the attribute, which will be selected as a root node for the decision tree by using information gain measure.

Color	Type	Origin	Class: Stolen
Red	Sports	Domestic	Yes
Red	Sports	Domestic	No
Red	Sports	Domestic	Yes
White	Sports	Domestic	No
Yellow	Sports	Import	Yes
Yellow	SUV	Import	No
Yellow	SUV	Import	Yes
White	SUV	Domestic	No
Red	SUV	Import	No
Red	Sports	Import	Yes

UKA TARSADIA UNIVERSITY

B.Tech (CE)/(Semester 6)
IT5003(2021-22)
Data Warehousing and Data Mining

Date :23/11/2024 **Time :**1:30PM- 4:30PM
Max. Marks:60

Instructions :

1. Attempt all questions.
2. Write each section in a separate answer book.
3. Make suitable assumptions wherever necessary.
4. Draw diagrams/figures whenever necessary.
5. Figures to the right indicate full marks allocated to that question.
6. Follow usual meaning of notations/abbreviations.

SECTION - 1

Q 1 A) Discuss the need for a data warehouse in organizations. [3]

Q 1 B) Explain the basic concepts of a data warehouse. How does it differ from a traditional operational database system? Provide examples to illustrate your explanation. [6]

OR

Q 1 B) What is the ETL process in data warehousing? Describe the steps involved in ETL and explain the significance of each step in preparing data for analysis.

Q 2 A) Explain the two measures of rule interestingness: support and confidence. [2]

Q 2 B) Discuss the pattern-growth approach for mining frequent item sets. Compare it with the Apriori algorithm and explain how it addresses the limitations of Apriori using an example. [7]

OR

Q 2 B) Let the database of transactions consist of the sets {1,2,5}, {2,3,4}, {3,4}, {1,2,3,4}, {1,3}, {1,2,3}. Apply algorithm for vertical format and find frequent patterns for support count = 2 and confidence = 60%.

Q 3 Answer the following in detail. (Any 2) **[12]**

I) Normalize the following group of data:
1000,2000,3000,9000
i)using min-max normalization by setting min:0 and max:1
ii)decimal scaling

II) What do you mean by the term "Data Mining"? And provide an explanation of its various functionalities.

III) Suppose two stocks A and B have the following values in one week:
A=2,3,5,4,6 and B= 5,8,10,11,14
Question: If the stocks are affected by the same industry trends, will their prices rise or fall together? Find it using Co-Variance method.

SECTION - 2

Q 4 A) Differentiate between AGNES and DIANA algorithms. [3]

Q 4 B) Explain DBSCAN algorithm with its advantages and disadvantages. [6]

OR

Q 4 B) What is cluster analysis? Discuss the requirements of cluster analysis.

Q 5 A) What is the role of sequencial data mining? [2]

Q 5 B) Discuss data mining in sales management. [7]

OR

Q 5 B) How can data mining be implied in world wide web applications?

Q 6 Answer the following in detail. (Any 2) **[12]**

I) Explain linear and non-linear regression.

II) Explain rule based classification.

III) For the below data set predict the class label for “Flu” using Naïve Bayesian Classifier for the sample X=<Cold=Light, Weakness= Yes, Headache= Mild, Fever= Normal>

Cold	Weakness	Headache	Fever	Flu
Severe	No	Mild	High	No
Severe	Yes	No	Normal	Yes
Severe	No	Strong	High	Yes
Light	Yes	Mild	High	Yes
Light	No	No	Normal	No
Light	Yes	Strong	High	Yes
Light	Yes	Strong	Normal	No
Severe	Yes	Mild	High	Yes