

**Eighth Semester B. E. (Computer Science and Engineering)
Examination**

DATA WAREHOUSING AND MINING

Time : 3 Hours]

[Max. Marks : 60

Instructions to Candidates :—

- (1) All questions carry marks as indicated against them.
- (2) **Number your answers properly.**
- (3) Assume suitable data and illustrate answers with neat sketches wherever necessary.

1. (a) Consider the star schema of an automobile data warehouse :

Autos(ModelId, modelname, serialNo, color)

Dealers(DealerId, name, city, state, phone)

Time(TimeId, day, week, month, year)

Sales(ModelId, DealerId, TimeId, QtySold, CountSold)

where the attribute QtySold is intended to be the total price of all automobiles for the given model, color, date and dealer, while CountSold is the total number of automobiles in that category.

Answer the following OLAP queries :—

- (i) Find total sales generated for model name (Maruti, Honda) and dealer state (Maharashtra, Gujarat) in September 2017 and October 2017 using ROLL–UP across three dimensions–ModelId, DealerId and TimeId.
- (ii) Find total sales generated for model name (Maruti, Honda) and dealer state (Maharashtra, Gujarat) in September 2017 and October 2017 using CUBE across the dimensions–ModelId, DealerId and TimeId.
- (iii) Comment on difference in output using ROLL–UP and CUBE aggregation clause.

- (iv) Find total sales generated for model name (Maruti, Honda) and dealer state (Maharashtra, Gujarat) in September 2017 and October 2017 using Partial ROLL – UP across dimensions DealerId and TimeId and group by ModelId.
- (v) Perform aggregation on amount generated. It should get aggregated by day first, then by all the weeks in each month and then across all months in the year.
- (vi) Why Groupid() clause is used in OLAP queries ?

6 (CO 2)

OR

- (b) Suppose that a data warehouse consists of the three dimensions time, doctor and patient and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit :
 - (i) Draw the lattice of cuboids (from apex to base cuboid) for the above data warehouse.
 - (ii) Draw a star schema diagram for the above data warehouse. For each dimension, include the appropriate attributes (conceptual hierarchies).
 - (iii) Starting with the base cuboid [day; doctor; patient], what specific OLAP operations (roll – up, drill down, dice, slice) should be performed in order to list the total fee collected by each doctor in 2004 ?
 - (iv) Assume that each dimension has four levels. How many cuboids will this cube contain ?
- (c) Design a data mart for analyzing the attendance of students in course. Show all the dimensions involved and arrange them in STAR schema. Write a SQL query to retrieve a particular student attendance from the factless fact table, Assume finest granularity.

6 (CO 2)

4 (CO 2)

2. (a) Differentiate between the following :—

- (i) ROLAP and MOLAP.
- (ii) Snowflake and fact constellation schema.

4 (CO 1)

(b) Explain the following types of Dimension with suitable examples :—

- Rapidly changing dimension.
- Slowly changing dimension.
- Degenerate dimension.

6 (CO 1)

OR

(c) Give suitable examples for Additive, Semi-Additive and Non-Additive Facts in Fact table of a data warehouse.

Compare and contrast between data warehouse and data mart. Also specify reasons for creating data mart.

6 (CO 1)

3. (a) State the advantages of partitioning in data-warehouse.

Write a SQL query to create composite List-Range partitioning for the following scenario :

- Customer table having attributes cust_id, cust_name, cust_state and time_id.
 - Perform list partitioning on state attributes and range partitioning on time-id.
 - Partition definitions for list are as below :
 - Partition East should accept values ('WB', 'JK')
 - Partition South should accept values ('TN', 'AP')
 - Partition North should accept values ('UP', 'HP')
 - Partition Temp should accept any other state.
 - Partition definitions for range are as below :
 - Partition P1 should accept values less than 01 - Jan - 2017
 - Partition P2 should accept values less than 01 - April - 2017
 - Partition P3 should accept values less than 01 - July - 2017
 - Partition P4 should accept values greater than 01 - July - 2017

Write SQL commands for the following :—

- (i) Insert 2 rows in the above table.
- (ii) Select data from any one partition.

5 (CO 2)

OR

- (b) Write SQL command to create Index Organized Table Employee with the attributes empno, empname and salary in tablespace tsa as directed :
- Empno is primary key for the table.
 - PCTTHRESHOLD is 20.
 - Specify Overflow and Including clause. Assume empname to be included in Including clause.
 - Give meaning of PCTTHRESHOLD, including and overflow clause.

Mention advantages of IOT over B-tree indexes. 5 (CO 2)

- (c) Describe the concept of compress index and with the help of suitable example illustrate how it helps in saving storage space. 3 (CO 2)
- (d) Bring out the difference between Cost based and Rule based optimizer. 2 (CO 2)

4. (a) Give is the frequency of stop words in documents (The values are given in increasing order) :

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.

Apply the following methods and show the results :—

- (i) Use smoothing by bin means with a depth of 3.
- (ii) Use Min – Max normalization to transform the value 35 into the range 0.0 to 1.0.
- (iii) Use z – score normalization to transform the value 35 where the standard deviation of the above frequency is 12.94.
- (iv) Use normalization by decimal scaling to transform the value 35.
- (v) Plot an equi – width histogram of width 10. 7 (CO 3)

OR

- (b) Suppose a student collected the price and weight of 20 products in shop with the following result :—

price	5.89	149	59.98	129	15.89	56.99	35.75	42.19	31	125.5
weight	1.4	1.5	2.2	2.7	3.2	3.9	4.1	4.1	4.6	4.8
price	4.5	22	52.9	61	33	328	122	142.19	229	89.4
weight	4.9	5.1	5.5	5.8	5.8	8.9	9.6	18.0	36.9	38.2

- (i) Give 5 number summary for price.
 - (ii) Draw boxplot for price. Identify outliers, if any.
 - (iii) Draw scatter plot based on these two variables.
 - (iv) Calculate the Pearson correlation coefficient. Are these two variables positively or negatively correlated ? 7 (CO 3)
- (c) What are the major challenges of mining a huge amount of data (such as billion of tuples) in comparison with mining a small amount of data (few hundred tuple data set). 3 (CO 3)
5. (a) State the Apriori Property. Generate candidate itemsets, frequent itemsets and association rules using Apriori algorithm on the following data set with minimum support count 2, minimum confidence 60%. List strongest association rules.

TID	List of items_IDs	
1. T100	I1, I2, I5	
2. T200	I2, I4	
3. T300	I2, I3	
4. T400	I1, I2, I4	
5. T500	I1, I3	
6. T600	I2, I3	
7. T700	I1, I3	
8. T800	I1, I2, I3, I5	
9. T900	I1, I2, I3	6 (CO 4)

OR

- (b) Apply Naive Bayesian classifier for the following dataset and find class(x) by executing it in the given training set.

X = (age < 30, Income = Medium, Student = Yes, Credit_rating = Fair)

Age	Income	Student	Credit_rating	Buys_laptop
≤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

6 (CO 4)

- (c) Two classifiers designed to predict patient's susceptibility to allergy are being designed and tested, independently of one another. Each of the classifiers predict that a patient is either positive (allergic) or negative (normal) based on a combination of observable factors. The tests result in the following two confusion matrices, one for each classifier :

Matrix : A

		predicted		
		allergic	normal	total
actual	Allegric	20	80	100
	Normal	30	490	520
total		50	570	620

Matrix : B

		predicted		
		allergic	normal	total
actual	Allegric	80	20	100
	Normal	190	330	520
	total	270	350	620

Calculate the accuracy, recall and F-measure for each of the classifiers based on these tables. Based on their values, can you recommend one classifier over the other, considering the type of application they are used for ? Justify your answer. 4 (CO 4)

6. (a) Use the k-means algorithm and Euclidean distance to cluster the following 10 examples into 3 clusters X1(2, 10); X2(2, 5); X3(8, 4); X4(9, 4); Y1(5, 8); Y2(7, 5); Y3(6, 4); Z1(1, 2); Z2(4, 9); Z3(6, 10).

Suppose that the initial seeds (centers of each cluster) are X1, X4 and Z2.

Run the k-means algorithm for 3 iteration. At the end of each iteration, show :

- The new clusters. (i.e. the examples belonging to each cluster).
- The centers of the new clusters.
- Draw a 10 by 10 space with all the 10 points and show the clusters after each iteration.

6 (CO 4)

(b)

	A	B	C	D	E	F
A	0.00	0.71	5.66	3.61	4.24	3.20
B		0.00	4.95	2.92	3.54	2.50
C			0.00	2.24	1.41	2.50
D				0.00	1.00	0.50
E					0.00	1.12
F						0.00

Use single link agglomerative clustering to group the data described by the following distance matrix. Produce the dendograms. 4 (CO 4)

**Eighth Semester B. E. (Computer Science and Engineering)
Examination**

INFORMATION SECURITY

Time : 3 Hours]

[Max. Marks : 60

Instructions to Candidates :—

- (1) All questions carry marks as indicated against them.
- (2) Solve Question **One** , Question **Two** , Question **Six** which are compulsory.
- (3) Solve Question **Three** b or c.
- (4) Solve Question **Four** or Question **Five** and Question **Seven** or Question **Eight**.
- (5) Assume Suitable Data and illustrate answer with neat sketch wherever necessary.

1. (a) Illustrate Network Access and Network security model for Encryption and decryption. Also list the characteristics of AES. 5 (CO 1)
- (b) Answer the following :—
 - State the number of rounds in 56 bit DES and 128 bit DES.
 - If the biases of three independent binary random variables X_1 , X_2 and X_3 are $-1/4$, $1/4$, $1/8$ then bias of $x_1 \oplus x_2 \oplus x_3$ will be ——— ?
 - Identify the attack that are based on linear crypt – analysis.
 - Which property provides a suggested substitution box.
 - Which mode will be an appropriate choice to implement a MAC using DES algorithm. 5 (CO 1)
2. (a) A row major plaintext matrix is arranged (on left) to generate a cipher block by applying a sequence of row transposition followed by column and again by row, the encrypted block is shown on the right. Now analyze

and identify what is the sequence of transposition that were performed ?

Plaintext						Encrypted
$\begin{bmatrix} s & e & c & u & r \\ e & y & o & u & r \\ n & e & t & w & o \\ r & k & n & o & u \end{bmatrix}$	\rightarrow					$\begin{bmatrix} w & n & r & k & o \\ r & c & s & e & u \\ r & o & e & y & u \\ o & t & n & e & u \end{bmatrix}$

5 (CO 1)

- (b) Apply a Hill cipher technique with $m = 3$ (block size = 3) with

the key $\begin{bmatrix} 23 & 3 & 7 \\ 5 & 9 & 21 \\ 11 & 8 & 13 \end{bmatrix}$

What is the ciphertext corresponding to plaintext "WAR" ?

Show the proof why Hill cipher decryption work.

5 (CO 1)

3. (a) Briefly explain the idea behind the RSA cryptosystem.

- What is the one way function in this system ?
- What is the trapdoor in this system ?
- Define the public and private keys in this system.
- Describe the security of this system.
- Identify various application based on RSA.

3 (CO 2)

- (b) In RSA, given $e = 13$ and $n = 100$.

Encrypt the message "MISSION" using 00 to 25 for letter A to Z and 26 for the spaces. Use different blocks to make $P < n$.

7 (CO 2)

OR

- (c) How will you discover that the integer 'a' is Z_n has a multiplicative inverse ? Identify such an algorithm and its proof to justify your answer and solve the following.

- Find multiplicative inverse of 8 in Z_{10} .
- Find all multiplicative inverse in Z_{10}, Z_{11} .

7 (CO 2)

4. Derive the equation of compression logic of MD5 design structure. Also design a three column table specifying all feature point differentiating between MD5 and SHA-I message digest algorithm. 10 (CO 3)

OR

5. (a) How does message digest is created with HMAC ? What are the problems that makes this algorithm not a choice for any application where non-repudiation is a need ? 5 (CO 3 , 4)
- (b) Identify two approaches to perform Digital signature. Also derive the equations for signing and verification process. 5 (CO 3 , 4)
6. (a) Why there is a need for Kerberos, in the same context ? What is realm ? Apply simple authentication dialogue to obtain ticket from AS, TGS server. 5 (CO 3)
- (b) Write the necessity of PGP certificates in to-days software application deployment for its privacy and integrity related issues over internet. 5 (CO 3)
7. (a) Identify the protocol that provides security to the conversation of web browser and web server ? Perceptualize and illustrate the steps that get involved in Record Protocol. 5 (CO 5)
- (b) Who all participants gets involved in any e-commerce transaction ? Give their roles of definition. 5 (CO 5)

OR

8. (a) Mention the roles of two categories of audit record. Explore with the help of example. 5 (CO 5)
- (b) Write various schemes that are designed for intrusion detection system. 5 (CO 5)

Course Code : CST 408-1

AZBY/MS – 18 / 4620

**Eighth Semester B. E. (Computer Science and Engineering)
Examination**

Elective – III

NATURAL LANGUAGE PROCESSING

Time : 3 Hours]

[Max. Marks : 60

Instructions to Candidates :—

(1) Assume suitable data and illustrate answers with neat sketches wherever necessary.

1. (a) Compute maximum likelihood estimation for following sentence :—

<s> I am a student <e>

<s> Student of good institute <e>

<s> Good institute good student <e>

Compute any three bigram probabilities. 5 (CO 1)

(b) What are different level of ambiguities, which make Natural Language Processing complex ? Discuss any three levels with example. 3 (CO 1)

(c) Write a python code to read a file and find number of stop words in the file. Use standard dataset of stop words. 2 (CO 1)

2. (a) Construct language model based on bigram probability method for following data set :

	Switch	Off	Power	Save	Nation
Switch	5	727	59	0	12
Off	0	10	410	0	22
Power	15	250	50	550	120
Save	0	27	671	9	320
Nation	30	0	15	0	0

The uni – gram details :

Switch	Off	Power	Save	Nation
1250	950	1900	1230	1122

Sentence : Save Nation Power Nation.

AZBY/MS-18 / 4620

Contd.

Apply Witten Bell smoothing technique and find out the probability of unseen event from Row – 2 of data set.

Using the smooth data set compute Entropy of the two bigram with maximum probability in dataset. 7 (CO 2)

- (b) Define Precision, Recall and F–measure. Explain the method to compute the values for Language Model. 3 (CO 2)

3. (a) Design CYK algorithm for the following Grammar and String. Find the final probability of the parse tree and demonstrate the result by redrawing the parse tree.

Grammar :

S → NP VP	0.9
S → VP	0.1
VP → V NP	0.5
VP → V	0.1
VP → V IN	0.3
VP → V PP	0.1
NP → NP NP	0.1
NP → NP PP	0.2
NP → N	0.7
PP → P NP	1
IN → NP PP	1

N → People	0.5
N → Fish	0.2
N → Tanks	0.2
N → rods	0.1
V → People	0.1
V → Fish	0.6
V → Tanks	0.3
P → with	1

Sentence : FISH TANKS FISH PEOPLE 10 (CO 3)

4. (a) Design Forward algorithm for suitable formulation. Use four word sentence for example. 5 (CO 3)

OR

Illustrate the drawback of PCFG using suitable example. 5 (CO 4)

- (b) Write generic semantic role labeling algorithm. Apply algorithm on sentence S1 and construct parse tree. Find Arguments, Target and Temporary variables. S1=The University Examiner declared a special incentive during last semester. 5 (CO 4)
5. (a) Explain minimal supervised WSD [Bootstrapping method] algorithm with suitable example. 5 (CO 4)
- (b) Define primitive decomposition.
Implement decomposition on :
Rahul opened the gate
The gate is open
The guard bought Mohan the vehicle. 5 (CO 4)

OR

Using NLTK toolkit, write python code to find out name–entity in a given file. Find out most common and uncommon name – entity. 5 (CO 5)

6. Solve any **Two** :—

- (a) Consider two languages [English and Hindi]
Green House The House
Harra Ghar Yeh Ghar
Compute MLE probability estimation. 5 (CO 5)
- (b) Describe symmetric alignment for phrase based machine translation.
Consider any two sentences for [English–Hindi] language and generate intersection matrix. 5 (CO 5)
- (c) Using suitable example : define term weight, vector length, term–by–document matrix, inverse document frequency, interpolated precision. 5 (CO 5)

**Eighth Semester B. E. (Computer Science and Engineering)
Examination**

Elective – III

DISTRIBUTED AND PARALLEL DATABASES

Time : 3 Hours]

[Max. Marks : 60

Instructions to Candidates :—

- (1) All questions carry marks as indicated.
 - (2) Assume suitable data and illustrate answers with neat sketches wherever necessary.
-
1.
 - (a) Draw and discuss the components of a commercial DDBMS. 5 (CO 1)
 - (b) Compare the following two parallel database architectures with their pros and cons :
 - (i) Hierarchical Architecture.
 - (ii) NUMA Architecture. 5 (CO 1)
 2.
 - (a) What are the various kinds of transparencies in distributed database design ? Explain each types with an example. 5 (CO 2)
 - (b) Consider the following relations :—
Books(BookNo, Author, Topic, Stock, Price)
BookStore(StoreNo, City, State, Zip, Inventory_Value)
Stock(StoreNo, BookNo, Qty)
 - (i) For the above global relations, give two simple predicates that would be meaningful for horizontal fragmentation of BookStore relation.
 - (ii) How would a derived horizontal fragmentation of Book be based on horizontal fragmentation of BookStore ?
 - (iii) Show the predicates by which Book may be horizontally fragmented. 5 (CO 2)
 3.
 - (a) Describe the behavior of two phase commit protocol in case of lost messages and network partitions. 6 (CO 2)

- (b) Draw and discuss the reference model for distributed transaction recovery.
4 (CO 2)

4. Solve any **Two** :—

- (a) How does distributed grouping and aggregate function evaluation takes place ?
Is there any need to introduce extension of relational algebra ? Explain the
properties of Group – by operation with proper example. 5 (CO 3)

- (b) Consider the database profile for following Global relations and their fragments :

Emp(name, age, sal, dno)

Dept(dno, dname, budget, mgr)

Card(Emp) = 20,000

	name	age	sal	dno
size	8	3	7	2
val	3000	60	2000	20

Card(Dept) = 20

	dno	dname	budget	mgr
size	2	15	8	2
val	20	20	1700	15

Card(Emp1) = 8000

	name	age	sal	dno
size	8	3	7	2
val	1200	40	900	10

Card(Dept1) = 10

	dno	dname	budget	mgr
size	2	15	8	2
val	10	10	1000	11

Consider the query to display name of employees working in department
having budget more than 50000. For simplicity disregard C0 and D0 in

the expression of cost and delays. Assume database profiles of Emp2 and Emp1 are similar with $\text{card}(\text{Emp2})=30000$ and database profile of Dept2 is similar to dept1 with $\text{card}(\text{Emp2})=10$. Evaluate any two execution strategies for implementing the query and calculate the delay. 5 (CO 3)

- (c) With example show the reduction of relations using semi joins. 5 (CO 3)
- 5. (a) Explain transaction recovery using checkpoints and UNDO/REDO method with a suitable example. 5 (CO 2)
- (b) Discuss the reason for failure in distributed systems and the basic fault tolerance approaches and techniques to cope with them. 5 (CO 2)
- 6. (a) Many Distributed data mining systems adopt Multi Agent System Architecture (MAS). Describe MAS architecture. 5 (CO 4)
- (b) How does optimized distributed mining algorithm works ? 5 (CO 4)

**Eighth Semester B. E. (Computer Science and Engineering)
Examination**

Elective – IV

WEB INTELLIGENCE AND BIG DATA

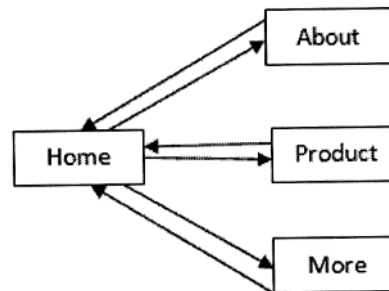
Time : 3 Hours]

[Max. Marks : 60

Instructions to Candidates :—

- (1) All questions carry marks as indicated against them.
- (2) Assume suitable data and illustrate answers with neat sketches wherever necessary.

1. (a) There is a database (DB) of 1 million fingerprints of the criminals in Nagpur. We have received a fingerprint (FP) from crime branch asking to verify the FP sample from the DB. How can we be able to do this in the best possible time ? Justify your answer. 3 (CO 1)
- (b) Find the Page rank for each web page in the given web graph considering $d = 0.5$



What can you conclude about the hierarchy in the web site design ?
4 (CO 1)

- (c) Suppose that we use 1000 functions chosen randomly from set of functions F . Next, we shall construct a new family of functions F_1 by performing a 1000-way OR on F . Then find :
 - (1) The probability that F_1 will put fingerprints from the same finger together in at least one bucket.

- (2) What is the probability that two fingerprints from different fingers will be placed in the same bucket ?
- (3) Calculate % false negative and % false positives.
(Use $k = 4$ and $p = 0.25$, $q = 0.9$) 3 (CO 1)

OR

- (d) Create an index on a set of documents having following texts. Comment on complexity of your algorithm. Use the method to search a query "Sachin, the Sachin". Documents :
- D1 : "The best batsman in the world is Sachin Tendulkar."
D2 : "The cricketers prefer to be a batsman."
D3 : "Sachin Tendulkar is an Indian cricketer." 3 (CO 1)

2. (a) Let's assume we have observed the reviews of a product during a span of one month from the date of its release. We have total 10,000 reviews. After summarizing the reviews
- 25% reviews say '**The product is very good and useful**';
 - 10% reviews say '**It is not worth the price, the other products in the market are better**'
 - 10% reviews say '**why waste money on buying new one when similar products with all functionalities are available**';
 - 20% reviews say '**not bad**';
 - 15% say '**one more product out, a good competitor though**';
 - 20% say '**nice to have one**'.

Create an environment based on the data which can automatically detect the opinion of the new review "**Not worth buying**". Assume the sentiments of the training reviews. 5 (CO 3)

- (b) What is mutual information ? How can it be used to determine the feature to be used in classification ? 3 (CO 4)

OR

- (c) How can you measure the efficiency of your classifier ? Explain with example. 3 (CO 4)

- (d) Can machine learning be used to find whether a person intends to shop or surf ? How can conditional probability be used to predict an event of buy if the following search keywords are given - Red, Flower, Gift and Cheap, derive the equation of probability of a buy = yes / no. 2 (CO 4)
3. (a) Illustrate how map-reduce uses parallel programming model to cater needs of web domain. Also illustrate the working of Mappers and reducers with example. 4 (CO 2)
- (b) "Efficiency of the Parallel computing system decreases with the increase in number of processors." Prove the efficiency of Map-Reduce in context of large problems. 3 (CO 2)
- OR**
- (c) "HDFS is fault tolerant" - Justify. 3 (CO 2)
- (d) What is the need of NO-SQL database ? Elaborate any NO-SQL database and its working. 3 (CO 2)
4. (a) Considering sample billing data listed below, find the frequent rules using association rule mining algorithm

Tid	Items
1	A, B, E
2	B, D
3	B, C
4	A, B, D
5	A, C
6	B, C
7	A, C
8	A, B, C, E
9	A, B, C

Given frequent set {A, B}, what association rules have minimum support = 4 and minimum confidence = 50% ? 5 (CO 3)

- (b) Apply Collaborative filtering. Find two items that are similar and hence suggest users that may have similar interests.

ID	User	Item	Rating
ID1	U1	M1	2
ID2	U1	M3	3
ID3	U2	M1	5
ID4	U2	M2	3
ID5	U3	M1	3
ID6	U3	M2	3
ID7	U3	M3	1
ID8	U4	M2	2
ID9	U4	M3	2

5 (CO 3)

5. (a) What are fundamental limits of logic ? Also discuss the differences between predicate and propositional logic. 4 (CO 4)

- (b) Consider the probability table as below :

Table T1-P(W S,R)				Table T2-P(R)		Table T3-P(S)	
W	S	R	P	R	P	S	P
y	y	y	0.9	y	0.2	y	0.3
y	y	n	0.7	n	0.8	n	0.7
y	n	y	0.8				
y	n	n	0.1				
n	n	n	0.9				
n	n	y	0.2				
n	y	n	0.3				
n	y	y	0.1				

Given the SQL query find the corresponding probability. What is this probability

and what is the evidence in each case :

- (1) Select R, Sum (P*P*P) from T1, T2 and T3 where W = y,
R = R, S = S group by R.
- (2) Select R, Sum (P*P*P) from T1, T2 and T3 where W = y
S = y, R1 = R2, R2 = R3 group by R. 4 (CO 4)
- (c) What is causality ? Is it similar to classification ? Justify. 2 (CO 4)

OR

- (d) Explain Predicate logic and its application on web data. 2 (CO 4)
6. (a) What is linear regression ? Explain with an example how we can predict a value of an output variable using linear regression. 4 (CO 1)

OR

- (b) Elaborate the concept of Hierarchical Temporal Memory. 4 (CO 1)
- (c) Explain how fuzzification can be used in decision making assuming any general example. 4 (CO 1)
- (d) Can we learn facts from free flowing texts ? Justify. 2 (CO 1)