MSE

WALCHAND COLLEGE OF ENGINEERING, SANGLI.

(An Autonomous Institute)

Final Year B.Tech. (Information Technology)

MID SEMESTER EXAMINATION SEMESTER- I SEPTEMBER-2018

CRYPTOGRAPHY AND NETWORK SECURITY (3IT401)

Day, Date and Time: Wednesday, 19/09/2018, 03.00pm to 04.30pm

Exam Seat Number: _____

Max Marks: **30**

IMP: Verify that you have received question paper with correct course, code, branch etc.

Instructions: i) All questions are compulsory. Writing question number is compulsory. The answers may not be assessed if question number is not written. Assume suitable data wherever necessary.
ii) Figures to the right of question text indicate full marks.
iii) Mobile phones are strictly prohibited.
iv) Except Exam Seat Number writing anything on question paper is not allowed.
Exchange/Sharing of stationery, calculator etc. not allowed.

Text on the right of marks indicates course outcomes (only for faculty use).

			Marks	
Q1	A)	Using suitable example, explain design principle of: (Any Two) i) Hill Cipher ii) Playfair Cipher iii) Row Transposition Cipher	6	CO1
Q1	B)	Differentiate active and passive attacks with necessary countermeasures.	3	CO1
Q2		Complete following table comparing Output Feedback and Counter modes of data operation w.r.t. given parameters.	9	CO3

Sr. No.	Parameter ↓ / Mode →	OFB	CTR
1	Input Mode (Stream/Block)		
2	Use of synchronized IV (Y/N)		
3	Encryption Parallelizable (Y/N)		
4	Decryption Parallelizable (Y/N)		
5	Random Read Access (Y/N)		
6	Error Propagation (Y/N)		
7	Supports Authentication than Confidentiality (Y/N)		
8	Working Design (In the form of En/Decryption component Figure)		

Q3	A)	For RSA algorithm, if primes p= 13, q=19 are used with encryption parameter e= 7; Calculate following: i) Decryption Parameter d (Forming minimum value valid pair with e) ii) Cipher C1 for plaintext M1=100 iii) Plaintext M2 back from Cipher C2= 120	9	CO2
Q3	B)	Fill in the blanks with appropriate integer values. Design criteria of DES algorithm uses:- i) Total _____ rounds of operation. ii) Individual round applies _____ bit key. iii) Block size = ___bits. iv) Total number of S boxes = _____ v) Input to each S box = _____ bits vi) In 3DES/2, the total key bits used are = _____	3	CO2

WALCHAND COLLEGE OF ENGINEERING, SANGLI.

MakeUp

(An Autonomous Institute)

Final Year B.Tech. (Information Technology)

MAKEUP EXAMINATION: SEMESTER I MAY-2019

CRYPTOGRAPHY AND NETWORK SECURITY (3IT401)

Exam Seat Number: _____

Day, Date and Time: Thursday, 09/05/2019, 02.00pm to 05.00pm

Max Marks: 100

IMP: Verify that you have received question paper with correct course, code, branch etc.

Instructions: i) All questions are compulsory. Writing question number is compulsory. The answers may not be assessed if question number is not written. Assume suitable data wherever necessary.
ii) Figures to the right of question text indicate full marks.
iii) Mobile phones and programmable calculators are strictly prohibited.
iv) Except Exam Seat Number writing anything on question paper is not allowed.
Exchange/Sharing of stationery, calculator etc. not allowed.

Text on the right of marks indicates course outcomes (only for faculty use).			Marks	
Q1	A)	Using suitable block diagram explain design principles of: i) Output Feedback Mode (OFB) of Data Transfer ii) DES algorithm round iii) Kerberos	18	CO2
Q1	B)	i) GCD(60, -12) = _____ ii) 7^5 mod 119 = _____	6	CO1
Q2	A)	Differentiate following: i) Private and Public key cryptography ii) Cryptography and Hash functions iii) Transport & Tunnel IP mode	18	CO3
Q2	B)	If plaintext is, 'helloworld', find ciphertext using: i) Caeser Cipher (key = 3) ii) Rail Fence Cipher (key = 2)	6	CO1
Q3	A)	In RSA public cryptosystem, if primes are p=5 and q=11, encryption parameter e=3 and plaintext M=9; Calculate lowest decryption parameter d and cipher C	9	CO2
Q3	B)	Draw structure of X.509 certificate showing various components.	8	CO3
Q4	A)	Using appropriate mathematical function, explain design and key exchange criteria of Diffie-Hellman algorithm. How a common key is calculated from both end users?	9	CO2
Q4	B)	How firewall is useful in system security? Enlist its various types.	8	CO3
Q5		Write Notes on: i) IP Security Architecture ii) Intrusion Detection Systems iii) Email Security	18	CO1

WALCHAND COLLEGE OF ENGINEERING

(Government Aided Autonomous Institute)
Vishrambag, Sangli - 416415

BE-IT.
MSE-ODD
2022-23

Final Year B. TECH. (Information Technology)
MSE , ODD SEMESTER, AY 2022-23
Cryptography and Network Security (5IT402)

MSE

Day, Date and Time: Tuesday, 11/10/2022, 03.00 pm to 04.30 pm PRN: _____

Max Marks: 30

IMP: Verify that you have received question paper with correct course, code, branch etc.

Instructions: a) All questions are compulsory.
b) Writing question number on answer book is compulsory otherwise answers may not be assessed.
c) Assume suitable data wherever necessary.
d) Figures to the right of question text indicate full marks.
e) Mobile phones and programmable calculators are strictly prohibited.
f) Except PRN anything else writing on question paper is not allowed.
g) Exchange/Sharing of stationery, calculator etc. not allowed.

Text on the right of marks indicates course outcomes (only for faculty use).

		Marks	
Q1 A)	Compare Following (Draw supporting diagrams): (Any Two) i) CBC and OFB ii) Symmetric and Asymmetric Encryption iii) Steganography and Cryptography	8	CO1
Q1 B)	Let sender A and receiver B agree upon for using RSA algorithm to send message M from A to receive ciphertext C at B. Let the two primes selected are 3 and 11. If public key component of e is chosen as 7; i) Find the lowest possible private key component d ii) Calculate the C for M= 2; by demonstrating the working of RSA.	4	CO2
Q2 A)	If Hill Cipher is calculated as C=KP mod 26; where C is cipher text and P is plaintext with given key K. Calculate C as output if given input text string P = 'meet' with positional convention as: a=1 … z=16 $K=\begin{bmatrix}9 & 4\\ 5 & 7\end{bmatrix}$	4	CO2
Q2 B)	Join the table From section A to B (Answer the correct pair e.g. A1-B2)	4	CO3

Section A	Section B
A1- Active Attack-	B1- Threat of frequency analysis
A2- Passive Attack-	B2- ECB
A3- Monoalphabetic-	B3- Repudiation
A4- Polyalphabetic-	B4- Counter Mode
A5- Diffusion -	B5- Spying
A6- Confusion-	B6- Playfair cipher
A7- Data transfer allowing parallelization of encryption while not requiring any decryption module -	B7- Complex relationship between ciphertext and key
A8- Secure block mode, however ciphertext exposes repetition of plaintext -	B8 - Complex relationship between plaintext and ciphertext

Q3 Fill in the blanks using most suitable word from the given table:

1) In LSB Steganography, to replace the least k-LSBs, the maximum error occurred is: _____

2) _____ maintains the same relationship with congruence equality relation given by $C \equiv (21p + 17) \bmod 26$ (Variables p and C represent plaintext and ciphertext respectively with cardinality of variable set is 26)

3) $(7^{(3+j)}) \bmod 19 =$ _____ (Where ^ denote the power function)

4) For 8 bit binary key if attacker is confirmed that MSB is 1 and LSB is 0. Then probability of guessing the remaining key <= _____

5) If plaintext M= "sangli" is to be encrypted using One Time Pad technique using key which is designed over a set of alphabets with cardinality = 26, then the achieved key space = _____

6) If plaintext M= "walchand" is to be encrypted using rail fence technique using key=3; then the ciphertext C= _____

7) $(36^{106} \bmod 107) \bmod 37 =$ _____

8) If Monoalphabetic cipher is to be encoded with key using 26 alphabets then the key space is equal to _____; for attempting Brute force attack.

9) The appropriate mechanism to avoid traffic analysis, DoS and Replay attacks; is to maintain _____

10) To avoid release of message contents, the countermeasure used is _____

$C \equiv (5p - 9) \bmod 26$	WHACADIN	$(7^{j}) \bmod 19$	2^{k-1}	1 mod 19
26*26*26*26*26*26	WALDNACH	$2^{k} - 1$	1	0.75
26*25*24*23*22*21	encipherment	23	0.25	19
$C \equiv (-5p + 9) \bmod 26$	access control	2^{26}	26!	Masquerade
$C \equiv (-5p - 9) \bmod 26$	Data integrity	26^{26}	26*6	2^{k}

WALCHAND COLLEGE OF ENGINEERING

(Government Aided Autonomous Institute)
Visharambag, Sangli – 416415

Final Year B.Tech. Information Technology
MSE, ODD SEMESTER, AY 2023-24
Cryptography and Network Security (5IT402)

MSE

PRN: _____

y & Date: Friday, 22/09/2023 Time : 3.00 pm to 4.30 pm

Max Marks: **30**

IMP: Verify that you have received question papers with correct course code, branch etc.

structions
a) All questions are compulsory.
b) Writing question number on answer book is compulsory otherwise answers may not be assessed.
c) Assume suitable data wherever necessary.
d) Figures to the right of question text indicate full marks.
e) Mobile phones, smart gadgets and programmable calculators are strictly prohibited.
f) Except PRN anything else writing on question paper is not allowed.
g) Exchange/Sharing of stationery, calculator etc. not allowed.

xt on the right of marks indicates course outcomes (Only for faculty use)

		Marks
A)	Fill in the blanks with appropriate integer values:	CO1
i)	The DES round sub_key size = _____ bits.	
ii)	By default, the key of Caesar cipher = ____	
iii)	The key size for 3DES/2 algorithm is = _____ bits	
iv)	In the given pair of prime numbers p=59 and q=61; the only safe prime is = _____	5
v)	(13+20) = X mod 26 ; The minimum value of X = ____	
B)	If RSA encrypts plaintext M to cipher_text C= 106 using prime factors p=11, q=13; and encryption parameter e=11 ; Find decryption parameter d and plaintext M.	CO2 5
A)	If Hill Cipher is calculated as C=KP mod 26; where C is cipher text and P is plaintext with given key K: $K=\begin{bmatrix}9 & 4\\ 5 & 7\end{bmatrix}$	CO2
i)	Calculate C as output if given input text string P = 'meet' with positional convention as a=1 … z=26	5
ii)	Select appropriate answer for following: a) Hill cipher is a monoalphabetic cipher (True/ False) b) Hill cipher applies *confusion* during its encryption cipher (True/ False)	

	B)	Draw working diagrams and compare CBC and OFB modes of data transfer	5	CO3
Q3	A)	Using suitable example explain following techniques: (Any Two)		CO1
	i)	Play Fair Classical Encryption - Comment on attack possibilities and crypt-complexity		
	ii)	Steganography - Comment on its properties, application and types	10	
	iii)	DES Round Function - Comment on number of rounds, expansion function and S boxes		

· · · · · End of question paper · · · · ·

WALCHAND COLLEGE OF ENGINEERING

(Government Aided Autonomous Institute)
Vishrambag, Sangli – 416415

Final Year B.Tech. Information Technology
Re-Exam, Odd and Even Semester AY 2023-24
Cryptography and Network Security (5IT402)

Re-Exam

PRN: _____

y & Date: Wednesday, 12/06/2024 Time : 10.30 am to 1.30 pm

Max Marks: **100**

IMP: Verify that you have received question papers with correct course code, branch etc.

structio ns

a) All questions are compulsory.
b) Writing question number on answer book is compulsory otherwise answers may not be assessed.
c) Assume suitable data wherever necessary.
d) Figures to the right of question text indicate full marks.
e) Mobile phones, smart gadgets and programmable calculators are strictly prohibited.
f) Except PRN anything else writing on question paper is not allowed.
g) Exchange/Sharing of stationery, calculator etc. not allowed.

ct on the right of marks indicates course outcomes (Only for faculty use)

		Marks	
A)	Using suitable example, explain following algorithms. (Any Two) i) DES ii) RSA iii) Diffie –Hellman	**18**	CO2
A)	Compare Following: i) Symmetric and Asymmetric Cryptography ii) Encryption and Hashing	**8**	CO2
B)	How IP security is achieved through Transport and Tunnel Mode?	**8**	CO1
A)	During Email Security process, What PGP and S/MIME are used for?	**8**	CO2
B)	What is SSL and TLS in Web Security?	**8**	CO1
A)	Using a neat diagram, demonstrate the typical format for X.509 certificate indicating certificate versions. What is the significant use of Certificate Revocation List?	**8**	CO3
B)	Explain the working of Kerberos as a private key authentication system.	**8**	CO3

Q5 A) What are Packet Filter Firewall and Application Level Gateways?

B) Using a neat diagram, describe the working of dual home and single homed bastion host firewall configuration.

Q6 A) Write short notes on: (Any Three)

i) Hill Cipher
ii) Steganography
iii) Malwares
iv) Password Management
v) Active and Passive attacks

..... End of question paper .....

WALCHAND COLLEGE OF ENGINEERING, SANGLI.

(An Autonomous Institute)

Final Year B.Tech. (Information Technology)

MID SEMESTER EXAMINATION SEMESTER- I SEPTEMBER-2018

DATA MINING (3IT402)

Exam Seat Number: _____

Day, Date and Time: Friday, 21/09/2018, 03.00pm to 04.30pm

Max Marks: **30**

IMP: Verify that you have received question paper with correct course, code, branch etc.

Instructions: i) All questions are compulsory. Writing question number is compulsory. The answers may not be assessed if question number is not written. Assume suitable data wherever necessary.

ii) Figures to the right of question text indicate full marks.

iii) **Mobile phones and programmable calculators are strictly prohibited.**

iv) **Except Exam Seat Number writing anything on question paper is not allowed.**

Exchange/Sharing of stationery, calculator etc. not allowed.

Text on the right of marks indicates course outcomes (only for faculty use).

		Marks	
Q1	Fill in the blanks 1. What is meant by discrete data? a) One that allows real numbers only, b) One that allows Only finite set of values c) Both a and b , d) One that allows float values only. 2. Which is the main key that is used by Decision Tree Algorithms to construct a Decision Tree. a) Entropy, b) Gain ratio, c) Information gain, d) All above a), b) and c). 3. Decision Trees algorithm will always tries Information gain _____. a) Equals to 1, b) Equals to 0 , c) Maximize, d) Minimize 4. _____stores data in a summarized version. a) Cube, b) Roll up, c) Both a and b, d) None 5. Which algorithm is used to find correlations among different attributes in a data set? a) Associative algorithm, b) Association algorithm, c) Time Series algorithm, d) Series algorithm 6. _____ is a process of taking operational data from one or more sources and mapping it, field by field, onto a new data structure in the data warehouse a) Transformation, b) Cleansing, c) Integration, d) Scrubbing 7. Summarization is a simple addition of values along one or more data dimensions a) True, b) False 8. Z-score normalization is also called as a) Zero-standard deviation normalization, b) Zero-mean deviation normalization c) Absolute mean normalization, d) Min-max at origin 9. FP-tree reduces a) Search cost, b) Memory consumption, c) CPU (Processing) cost, d) All above 10. A_____ plot is one of the most effective graphical method for determining if there appears to be a relationship, pattern, or trend between 2 quantitative variables a) Scatter, b) Loess curve, c) q-q, d) quantile	5	CO1

Q2	A)	Use min-max normalization techniques for transforming values 36,103 and 52 from following data into 1 to 10. Sample data {44,36,72,103,29,1,52,3,1,11}.	3	CO2
Q2	B)	Define data reduction and state strategies for it.	2	CO1
Q2	C)	Write 3-4-5 rule? Why it is used?	2	CO2

Q3	A)	State the possible ways of integrating a data mining system with data warehouse.	2	CO1
Q3	B)	State the major types of Concept hierarchies	2	CO2

Q3 C) For following data, find Entropy and Gain for (Play_Golf, Outlook)

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

Q4 A) How does ARCS (Association Rules Clustering System) work? Draw block diagram and give the limitations of ARCS. 3

Q4 B) Find correlation and comment on type of correlation. 2

Advertisement budget (Million Rs)	Annual Sale (Million Rs)
9	19
7	13
5	12
8	16
6	15
3	10
4	8

Q4 C) Draw FP tree for following Transaction data. State advantages of FP tree over candidate generation algorithms. 5

TID	Items purchased
1	b,a
2	b,d,c
3	d,e,a,c
4	a,d,e
5	c,b,a
6	a,c,b,d
7	a,f
8	b,a,c
9	b,d,a
10	c,e,b

WALCHAND COLLEGE OF ENGINEERING, SANGLI.

(An Autonomous Institute)

Final Year B.Tech. (Information Technology)

MAKEUP EXAMINATION: SEMESTER I MAY-2019

DATA MINING (3IT402)

Exam Seat Number: _____

Day, Date and Time: Saturday, 11/05/2019, 02.00pm to 05.00pm

Max Marks: **100**

IMP: Verify that you have received question paper with correct course, code, branch etc.

Instructions: i) All questions are compulsory. Writing question number is compulsory. The answers may not be assessed if question number is not written. Assume suitable data wherever necessary.
ii) Figures to the right of question text indicate full marks.
iii) Mobile phones and programmable calculators are strictly prohibited.
iv) Except Exam Seat Number writing anything on question paper is not allowed.
Exchange/Sharing of stationery, calculator etc. not allowed.

Text on the right of marks indicates course outcomes (only for faculty use).

			Marks	
Q1	A)	State and define in short: major Tasks in Data Preprocessing.	5	CO1
Q1	B)	Differentiate in between OLAP and OLTP.	5	CO1
Q1	C)	Draw and explain a starnet query model with suitable data.	5	CO1
Q2	A)	What is 'concept hierarchy'? State the major types of it and explain rule-based hierarchy in detail with example	5	CO2
Q2	B)	In following data, calculate information gain for "department" attribute. How we can use the information gain for identifying weakly relevant attributes?	6	CO3

Gender	Department	Grade	Count
M	IT	B	16
F	IT	A	22
M	CSE	A	18
F	IT	A	25
M	IT	A	21
F	CSE	A	18

Gender	Department	Grade	Count
M	IT	B	16
F	ELN	C	22
M	ELN	C	18
F	IT	C	25
M	CSE	B	21
F	CSE	A	18

Q2	C)	What is a box and whisker plot? A sample of 10 boxes has these weights (in Kg): 25,28,29,29,30,34,35,35,37,38	6	CO3
Q3	A)	Explain Apriori algorithm and find association rules for following data set with minimum support=20%.	6	CO3

Tid	Items
1	A, C, D
2	B, C, E
3	A, B, C, E
4	B, E

Q3	B)	State the methods to Improve Apriori's Efficiency .	5	CO1
Q3	C)	In constraint based association rule mining, which constrains are used.	5	CO2
Q4	A)	State the criteria for comparing and evaluating classification and prediction methods.	4	CO1
Q4	B)	What is Bayesian belief network (BBN)? State the characteristics of BBN.	4	CO1

Q4 C) From given data ,identify the class of following test case by using Naïve Bayes Classifier. 6
Test case - {Home loan Refund = No, Marital status= Married, Income=120K}

Marital status	Income (K)	Home loan refund	Defaulter
Single	125	Yes	No
Married	100	No	No
Single	70	No	No
Married	120	Yes	No
Divorced	95	No	Yes
Married	60	No	No
Divorced	220	Yes	No
Single	85	No	Yes
Married	75	No	No
Single	90	No	Yes

4 D) From given data identify the class for test case using Naïve Bayes Classifier. 6
{Have Legs= No, Give Birth= Yes, Can Fly= No, Live in water=Yes}

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Q5	A)	Describe major clustering approaches.	5	CO1
Q5	B)	Apply k-means clustering method on following data with k=2 up to 2 iterations only. Assume initial 2 centroids as (1,1) and (2,1).	6	CO3

Sr.no	Attribute1	Attribute2
A	1	1
B	2	1
C	4	3
D	5	4

Q5	C)	Apply agglomerative hierarchical clustering with single-linkage to form clusters for following data of distances between cities given in matrix form.	6	CO3

Distance	BA	FI	MI	NA	RM	TO
BA	0	662	877	255	412	996
FI	662	0	295	468	268	400
MI	877	295	0	754	564	138
NA	255	468	754	0	219	869
RM	412	268	564	219	0	669
TO	996	400	138	869	669	0

Q6	A)	What are the tteps for performing a similarity search.	5	CO2
Q6	B)	Precision and recall for Text Retrieval	5	CO2
Q6	C)	Draw Web Mining Taxonomy	5	CO2

WALCHAND COLLEGE OF ENGINEERING

(Government Aided Autonomous Institute)
Visharambag, Sangli - 416415

Final Year B.Tech. (Information Technology)
ESE , ODD SEMESTER, AY 2022-23
Data Mining (5IT401)

ESE

PRN: _____

& Date: Tuesday, 13/12/2022 Time : 3.00 pm to 5.00 pm

Max Marks: **50**

IMP: Verify that you have received question papers with correct course code, branch etc.

structions
a) All questions are compulsory.
b) Writing question number on answer book is compulsory otherwise answers may not be assessed.
c) Assume suitable data wherever necessary.
d) Figures to the right of question text indicate full marks.
e) Mobile phones, smart gadgets and programmable calculators are strictly prohibited.
f) Except PRN anything else writing on question paper is not allowed.
g) Exchange/Sharing of stationery, calculator etc. not allowed.

ext on the right of marks indicates course outcomes (Only for faculty use)

			Marks	
1	A)	State and explain major Tasks in Data Preprocessing.	3	CO1
2	A)	What is Attribute Relevance Analysis? How it is performed?	3	CO1
3	A)	Draw FP-tree for following transaction data (Table 1).	4	CO2

Table 1

Transaction	List of items
T1	I1,I2,I3
T2	I2,I3,I4
T3	I4,I5
T4	I1,I2,I4
T5	I1,I2,I3,I5
T6	I1,I2,I3,I4

Q4	A)	Which difficulties may arise when a decision tree is constructed?	4	CO2
	B)	Why Gini index is used? Consider the data given in Table 2, to find Gini Index of - {(Supercar, Heavy),Economy}	5	CO3

Table No-2

Car type	Class
Economy	High
Supercar	High
Supercar	High
Economy	Low
Heavy	Low
Economy	High

1B094BDD3D2FA4D47187EF21A1627B6C

-OR-

Why Info Gain is used? A training data (Table 3) is given as follows, find the Info Gain for Outlook.

Table 3

St. No.	Outlook	Temperature	Humidity	Windy	Class
1	overcast	hot	high	false	Play
2	overcast	mild	high	true	Play
3	overcast	hot	normal	false	Play
4	rain	mild	high	false	Play
5	rain	cool	normal	false	Play
6	rain	cool	normal	true	NoPlay
7	rain	mild	normal	false	Play
8	rain	mild	high	true	NoPlay
9	sunny	hot	high	false	NoPlay
10	sunny	hot	high	true	NoPlay
11	sunny	mild	high	false	NoPlay
12	sunny	cool	normal	false	Play
13	sunny	mild	normal	true	Play

C) Consider Table 3 as training data and Use Bayes Classifier to predict class of following condition. 5

{Outlook=Sunny, Temperature= cool, Humidity= high, Windy= false}

Q5 A) State the categories of constraints in constraints-based clustering. 4

B) Explain Grid-Based clustering Method. Give example. 5

C) Apply agglomerative hierarchical clustering algorithm to form clusters for following data (Table 4) using single linkage approach. Draw dendrogram for resulting clusters. 5

Table 4

Data/Distance	A	B	C	D	E	F
A	0	-	-	-	-	-
B	662	0	-	-	-	-
C	877	295	0	-	-	-
D	255	468	754	0	-	-
E	412	268	564	219	0	-
F	996	400	138	869	669	0

Q6 A) State and elaborate in short- Classification of Web Mining Techniques. 4

B) Elaborate in short with example and diagram- Spatial trend analysis 4

C) How data mining can be employed on Digital Images? State any difficulties in Image mining. 4

End of question paper

WALCHAND COLLEGE OF ENGINEERING

(Government Aided Autonomous Institute)

Vishrambag, Sangli - 416415

Final Year B. TECH. (Information Technology)

MSE , ODD SEMESTER, AY 2022-23

Data Mining (5IT401)

MSE

Day, Date and Time: Monday, 10/10/2022, 03.00 pm to 04.30 pm

PRN: _____

Max Marks: 30

IMP: Verify that you have received question paper with correct course, code, branch etc.

Instructions: a) All questions are compulsory.
b) Writing question number on answer book is compulsory otherwise answers may not be assessed.
c) Assume suitable data wherever necessary.
d) Figures to the right of question text indicate full marks.
e) Mobile phones and programmable calculators are strictly prohibited.
f) Except PRN anything else writing on question paper is not allowed.
g) Exchange/Sharing of stationery, calculator etc. not allowed.

Text on the right of marks indicates course outcomes (only for faculty use)

		Marks	
Q1 A)	State the methods to fill in the missing values for attributes in data mining process.	3	CO1
Q1 B)	State the possible ways of integrating a data mining system with data warehouse.	3	CO2
Q1 C)	State use of normalization of data. Perform min-max normalization of following data in the range of 0 to 1.	4	CO2

Name	Years of Experience
A	8
B	20
C	10
D	15

Q2 A)	State the basic principles of Attribute-Oriented Induction.	3	CO1
Q2 B)	What is concept hierarchy? State the major types of concept hierarchy with example of each type.	3	CO2
Q2 C)	Find the 5 number summary for the given data set and draw box-plot for it Data : 23, 42, 12, 10, 15, 14, 9.	4	CO3
Q3 A)	Give the classification of association rule mining based on different criteria.	3	CO1
Q3 B)	Draw flowchart for ARCS (Association Rules Clustering System)? Give the limitations of ARCS.	3	CO2
Q3 C)	Find maximal frequent itemset from following transaction data. Assume minimum support of 50%.	4	CO2

TID	A	B	C	D	E	F
T_1	1	0	1	1	0	0
T_2	0	1	0	1	0	0
T_3	1	1	1	0	1	0
T_4	0	1	0	1	0	1

WALCHAND COLLEGE OF ENGINEERING

(Government Aided Autonomous Institute)
Visharambag, Sangli – 416415

Final Year B.Tech. Information Technology
Re-Exam, Odd and Even Semester AY 2022-23
Data Mining (5IT401)

Re-Exam

PRN: _____

Date: Monday, 10/07/2023 Time : 02.00 pm to 05.00 pm

Max Marks: **100**

MP: Verify that you have received question papers with correct course code, branch etc.

uctions
a) All questions are compulsory.
b) Writing question number on answer book is compulsory otherwise answers may not be assessed.
c) Assume suitable data wherever necessary.
d) Figures to the right of question text indicate full marks.
e) Mobile phones, smart gadgets and programmable calculators are strictly prohibited.
f) Except PRN anything else writing on question paper is not allowed.
g) Exchange/Sharing of stationery, calculator etc. not allowed.

on the right of marks indicates course outcomes (Only for faculty use)

		Marks	
A)	State and explain in short - the stages in data mining	4	CO1
B)	What is Dimensionality Reduction? State basic Heuristic methods of attribute subset selection.	4	CO1
C)	Apply min- max and Z-Score Normalization for following data. Data = {1000, 2000, 3000, 5000, 9000}	4	CO3
D)	What you mean by binning? State different binning methods and smooth following data by these method Data- 4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34	5	CO3
A)	Define and give example of- Schema Hierarchies and Set-grouping hierarchies.	6	CO1
B)	When Analytical Characterization is performed? How it is performed?	4	CO2
C)	Classify data mining Primitives with the help of real word examples.	4	CO1
D)	What is Cube? Elaborate Roll-up and Drill-down operation.	4	CO3
A)	Write a short note on Brute-force approach for mining association rule.	4	CO1
B)	Give the advantages of FP-tree for association rule mining.	4	CO3
C)	Find out the type of correlation analysis for following data.	4	CO2

A	B
20	8
12	34
9	4

D)	For the following given transaction data-set generate rules using Apriori algorithm.	5	CO2

0143143A582C1DBAADC2A5C568A5605D

Q4 **A) Construct decision tree for given dataset.**

Sore Throat	Fever	Swollen Glands	Congestion	Headache	Diagnosis.
Yes	Yes	Yes	Yes	Yes	Streep Throat
No	No	No	Yes	Yes	Allergy
Yes	Yes	No	Yes	No	Cold
Yes	No	Yes	No	No	Streep Throat
No	Yes	No	Yes	No	Cold
No	No	No	Yes	No	Allergy
No	No	Yes	No	No	Streep Throat
Yes	No	No	Yes	Yes	Allergy
No	Yes	No	Yes	Yes	Cold
Yes	Yes	No	Yes	Yes	Cold

6

B) From given data ,identify the class of following test case by using Naïve Bayes Classifier. Test case - $X = (\text{Refund} = \text{No}, \text{Married}, \text{Income} = 120\text{K})$

Marital status	Income (K)	Home loan refund	Defaulter
Single	125	Yes	No
Married	100	No	No
Single	70	No	No
Married	120	Yes	No
Divorced	95	No	Yes
Married	60	No	No
Divorced	220	Yes	No
Single	85	No	Yes
Married	75	No	No
Single	90	No	Yes

6

C) State the criteria for comparing and evaluating classification and prediction methods. 5

Q5 **A)** State the typical requirements of clustering in data mining. 3

B) How Density-Based Clustering Method works? What is *Eps* and *MinPts?* 3

C) Apply k-means clustering method on following data with k=2; assume *Initial value of centroids* are C1=(1,1) and C2=(2,1). **Find first two cluster assignments iterations only.**

Sr.no	Attribute1	Attribute2
A	1	1
B	2	1
C	4	3
D	5	4

6

D) When we prefer to use hierarchical clustering? Perform single linkage hierarchical clustering on following data. CO2 6

Distance	A	B	C	D	E
A	0	-	-	-	-
B	0.71	0	-	-	-
C	5.66	4.95	0	-	-
D	3.61	2.92	2.24	0	-
E	4.24	3.54	1.41	1.00	0
F	3.20	2.50	2.50	0.50	1.12

Q6 A) What is Hotspot analysis in spatial data mining? What are its applications? 4 CO2

B) Short note on - Similarity Search in Multimedia Data. 4 CO2

C) Elaborate First Order Markov Model. 5 CO1

. . . . . End of question paper . . . . .

...CHAND COLLEGE OF ENGINEERING
(Government Aided Autonomous Institute)
Visharambag, Sangli – 416415

Final Year B.Tech. Information Technology
MSE, ODD SEMESTER, AY 2023-24
Data Mining (5IT401)

MSE

...y & Date: Thursday, 21/09/2023 **PRN:** _____

Time : 3.00 pm to 4.30 pm

Max Marks: **30**

...structions **IMP: Verify that you have received question papers with correct course code, branch etc.**
a) All questions are compulsory.
b) Writing question number on answer book is compulsory otherwise answers may not be assessed.
c) Assume suitable data wherever necessary.
d) Figures to the right of question text indicate full marks.
e) Mobile phones, smart gadgets and programmable calculators are strictly prohibited.
f) Except PRN anything else writing on question paper is not allowed.
g) Exchange/Sharing of stationery, calculator etc. not allowed.

...n the right of marks indicates course outcomes (Only for faculty use)

		Marks	
A)	Answer in short (two/three) sentences.		
	i. What is role of IT engineer/expert in Data mining field?		CO2
	ii. How data mining helps in business?	3	
	iii. Give scope of data mining?		
B)	What do you mean by preprocessing of data in data mining? Give data preprocessing stages.	3	CO1
C)	Write 3-4-5 rule? Why it is used?	3	CO3
D)	Why is Normalization in Data Mining? Find normalized values for following data in the range [0, 1] using min-max normalization method.		CO2

Employee Name	Years of Experience
ABC	8
JKL	20
MNO	10
PQR	15

3

A)	Define concept hierarchy. State the major types with example of each.	3	CO1
B)	State the possible ways of integrating a data mining system with data warehouse.	3	CO1

C) The crosstab of t-weight and d-weight is given below. Fill the missing data in it. Also write quantitative description rule for target class 'TV'. CO2

Class\Region	State 1			State 2			Both States		
	Count	t-weight (%)	d-weight (%)	Count	t-weight (%)	d-weight (%)	Count	t-weight (%)	d-weight (%)
TV	80	25	40	240		30	320	100	32
PC	120		60	560	82.35	70	680	100	
Both class	200	20	100		80	100	1000	100	100

3

Q4 A) Compare Apriori and FP-growth algorithms for mining frequent patterns in large datasets. 3 CO1

B) State methods to improve Apriori efficiency. 3 CO3

C) Consider given transaction data and - CO3

i. How many association rules generated using Brute-Force approach?
ii. Find support of Itemset : {Milk, Bread, Egg}
iii. Find confidence of following association rules

R1: Bread → {Milk, Egg}

R2: {Milk, Egg} → Bread

Trans_ID	Items purchased
1	Milk, Bread, Egg
2	Milk, Juice
3	Juice, Butter
4	Milk, Bread, Egg
5	Coffee, Egg
6	Coffee
7	Coffee, Juice
8	Milk, Bread, Cookies, Egg
9	Cookies, Butter
10	Milk, Bread

3

····· End of question paper ·····