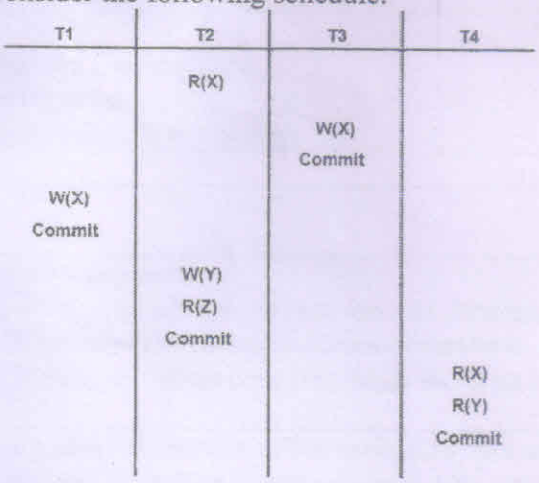
SOMAIYA VIDYAVIHAR UNIVERSITY

23/05/2022(E)

Semester: January 2022 – May 2022 **Maximum Marks:**100 **Examination: ESE Examination** **Duration:**3 hrs		
Programme code: 01 **Programme:** B.Tech Computer Engineering	**Class:** SY	**Semester:** IV (SVU 2020)
Name of the Constituent College: K. J. Somaiya College of Engineering	**Name of the department:** COMP	
Course Code: 116U01C403	**Name of the Course:** Relational Database Management System	
Instructions: **1) All Questions are Compulsory.** **2) Draw neat diagrams.** **3) Assume suitable data if necessary.**		

Question No.		Max. Marks
Q 1	**Attempt any two.** i) List and explain the various users of database and their roles. ii) Describe different applications of database. iii) State and explain concerns while using an enterprise database.	10 M
Q 2 (a)	A university registrar's office maintains data about the following entities: (a) courses including course_number, title, credits, syllabus, and prerequisites (b) course_offerings including year, semester, section_number, instructor(s), timings, and classroom_no (c) students including student_id, name, and program (d) instructors including identification_number, name, department, and title. Further, the enrollment of students in courses and grades awarded to students in each course they are enrolled for must be appropriately modeled. i. Construct an E-R diagram for the registrar's office. Document all assumptions that you make about the mapping constraints. ii. Show relational mapping for the ER diagram. **OR** Draw EER model and show relational mapping for Insurance Management System. Assume required data.	10 M
Q 2 (b)	Consider following schema; Branch (branch_name, branch_city, assets) customer (customer_name, customer_street, customer_city) loan (loan_number, branch name, amount) borrower (customer_name, loan_number) account (account_number, branch name, balance) depositor (customer_name, account_number) Write SQL for the following; i) Find all customers who have a loan at the bank but do not have an account. ii) Find all customers that have both an account and a loan. iii) Write PL/SQL function to add records in the Branch table. iv) Apply not null constraint to amount attribute by altering structure of table loan.	10 M

<table>
<tr><td>Q 3 (a)</td><td>Attempt any two.
i) Explain Domain Constraint and Referential Integrity constraint with On Delete Cascade with suitable example
ii) Describe security mechanism in SQL.
iii) Explain the working of REVOKE and TRUNCATE command with example.</td><td>10 M</td></tr>
<tr><td>Q 3 (b)</td><td>Consider the following relational database schema consisting of the four relation schemas:
passenger (pid, pname, pgender, pcity)
agency (aid, aname, acity)
flight (fid, fdate, time, src, dest)
booking (pid, aid, fid, fdate)
Answer the following questions using relational algebra queries;
i) Find only the flight numbers for passenger with pid 123 for flights to Chennai before 06/11/2020.
ii) Find the passenger names for those who do not have any bookings in any flights.
iii) Get the details of flights that are scheduled on both dates 01/12/2020 and 02/12/2020 at 16:00 hours.
iv) Find the details of all male passengers who are associated with agency.</td><td>10 M</td></tr>
<tr><td>Q 4 (a)</td><td>Draw Extendable Hashing structure to store following records. Consider maximum bucket size = 2.
<table><tr><th>Town</th><th>f(Town)</th></tr><tr><td>Brighton</td><td>0010</td></tr><tr><td>Clearview</td><td>1101</td></tr><tr><td>Downtown</td><td>1010</td></tr><tr><td>Mianus</td><td>1000</td></tr><tr><td>Perryridge</td><td>1111</td></tr><tr><td>Redwood</td><td>1011</td></tr><tr><td>Round Hill</td><td>0101</td></tr></table>OR
Let relation R (ABCDEFG) and FDs {AB → C, AC → B, AD → E B → D, BC → A, E → G}. Consider the decomposition for the given schema D = {ABC, ACDE, ADG}. Find it is lossy or lossless decomposition assuming same set of functional dependency holds.</td><td>10 M</td></tr>
<tr><td>Q 4 (b)</td><td>Show the equivalence between given set of functional dependencies with proper steps.
A relation R (A , C , D , E , H) is having two functional dependencies sets F and G as shown-
Set F- AC → D, E → AD, E → H
Set G- A → CD, E → AH</td><td>10 M</td></tr>
</table>

Q 5 (a)	Consider a file of 8192 records. Each record is 16 bytes long and its key field is of size 6 bytes. The file is ordered on a key field, and the file organization is unspanned. The file is stored in a file system with block size 512 bytes, and the size of a block pointer is 10 bytes. If the primary index is built on the key field of the file, and a multivalued index scheme is used to store the primary index, then find the number of first level and second level blocks in the multilevel index.	10 M
Q 5 (b)	Explain **any two** from the following with suitable diagram. i) State diagram of Transaction. ii) Shadow paging iii) Deadlock handling- wait for graph.	10 M
Q 6 (a)	Consider the following schedule.  Determine the given schedule is conflict serializable, recoverable or cascadeless. Provide stepwise solution.	5 M
Q 6 (b)	Describe the significance of Thomas write rule in concurrency control process. **OR** Explain the recovery process using log based recovery.	5 M

18/5/2022 (E)

SOMAIYA VIDYAVIHAR UNIVERSITY

Semester: January 2022 – May 2022 **Maximum Marks: 100** **Examination: ESE Examination** **Duration: 3hrs**		
Programme code: 01&04 **Programme: B Tech Comp/IT**	**Class: SY**	**Semester: IV (SVU 2020)**
Name of the Constituent College: **K. J. Somaiya College of Engineering**	**Name of the department:** COMP/IT	
Course Code: 116U01C401 /116U04C401	**Name of the Course: Probability, Statistics and Optimization Techniques**	
Instructions: 1)Draw neat diagrams 2)Assume suitable data if necessary		

Q No			MAX MARKS
Q1	a	If X_1 has mean 5 and variance 5, X_2 has mean -2 and variance 3, find $E(2X_1+3X_2-5)$, $V(2X_1+3X_2-5)$	5
	b	**Solve any THREE of the following**	21
	(i)	Determine the constant 'a' and find mean, $P(4 \le x \le 7)$ if the distribution function of a continuous random variable is defined as $f(x)=\frac{a}{x^5}, 2 \le x \le 10$	
	(ii)	If the height of 1000 students is normally distributed with mean 69 inches and standard deviation 4 inches. Find the expected number of students having heights: i) greater than 67 inches, ii) less than 68 inches, iii) between 65 & 71 inches	
	(iii)	The number of phone calls coming in to a telephone exchange between 2 & 4 P.M. say X is a random variable has Poisson distribution with parameter 2. Similarly the number of phone calls coming between 4 & 6 P.M. say Y is a random variable has Poisson distribution with parameter 6. If X& Y are independent Poisson random variables find the probability that during 2&6 P.M. there will be i) no phone calls at all ii) more than 3 calls. (iii) at most two calls	
	(iv)	A box to be constructed so that its height is 12 inches and its base is X inches by X inches. If X has a uniform distribution over the interval (2, 10), then what is the expected volume of the box in cubic inches?	
	(v)	The joint probability distribution function of (X,Y) is given by $f(x,y)=e^{-(x+y)} \quad 0 \le x, \ 0 \le y$ Compute $P(X>2), P(1<X+Y<3))$	

Q2 **a** A data for selection of students regarding placement is given below. Find the probability that a boy is selected for the placement and log of odds of this probability — **5**

Students	Selection in placement		Total
	yes	no	
Girls	753	102	855
Boys	382	158	540
Total	1145	250	135

b **Solve any ONE of the following** — **7**

(i) Calculate the correlation coefficient from the following data.

x	23	27	28	29	30	31	33	35	36	39
y	18	22	23	24	25	26	28	29	30	32

(ii) Obtain two lines of regression and coefficient of correlation from the following data-

X	65	66	67	68	69	70	72	67
y	67	68	65	72	72	69	71	66

Q3 **a** Two samples are drawn from two different population gave the following results. Find 95% confidence limits for the difference between the population means. — **5**

	Size	Mean	S.D
Sample I	400	124	14
Sample II	250	120	12

b **Solve any TWO of the following** — **14**

(i) Intelligence tests of two groups of boys & girls obtained from two normal populations having the same standard deviations gave the following results. Test at 1% level of significance whether the boys perform better than the girls.

	Size	Mean	S.D
Girls	121	84	10
Boys	181	81	12

(ii) A certain injection administered to 12 patients resulted in the following change of blood pressure 5,2,8, -1,3,0,6, -2,1,5,0,4. Can be concluded that the injection will be in general accompanied by an increase in blood pressure at 5% LOS?

(iii) From the following table, showing the number of plants having certain character, test the hypothesis that the flower colour is independent of flatness of leaf.

				Flat leaves	Curved leaves	Total	
			White Flowers	99	36	135	
			Red Flowers	20	5	25	
			Total	119	41	160	
Q4	a	Construct the Dual of the following LPP Maximize $z = 5x_1 + 2x_2 - 3x_3$ Subject to $2x_1 - 2x_2 + x_3 \geq 4$ $2x_1 + x_3 \leq 8$ $x_1 + x_2 + 3x_3 = 20$ $x_1, x_3 \geq 0$, x_2 unrestricted					5
	b	**Solve any THREE of the following**					21
	(i)	Using Simplex method solve the following LPP Maximize $z = 3x_1 + 2x_2 + 5x_3$ Subject to $x_1 + x_2 + x_3 \leq 9$ $2x_1 + 3x_2 + 5x_3 \leq 30$ $2x_1 - x_2 - x_3 \leq 8$ $x_1, x_2, x_3 \geq 0$					
	(ii)	Using Big M method solve the following LPP Maximize $z = 6x_1 + 4x_2$ Subject to $2x_1 + 3x_2 \leq 30$, $3x_1 + 2x_2 \leq 24, x_1 + x_2 \geq 3$, $x_1, x_2 \geq 0$					
	(iii)	Using Duality Solve the following linear programming problem Minimize $z = 4x_1 + 3x_2 + 6x_3$ Subject to $x_1 + x_3 \geq 2$, $x_2 + x_3 \geq 5$, $x_1, x_2, x_3 \geq 0$					
	(iv)	Using Dual simplex method Solve the following linear programming problem Minimize $z = 2x_1 + 2x_2 + 4x_3$ Subject to $2x_1 + 3x_2 + 5x_3 \geq 2$, $3x_1 + x_2 + 7x_3 \leq 3$, $x_1 + 4x_2 + 6x_3 \leq 5$ $x_1, x_2, x_3 \geq 0$					
	(v)	Solve the following NLPP Maximize $z = 2x_1^2 - 7x_2^2 + 12x_1x_2$ Subject to $2x_1 + 5x_2 \leq 98$, $x_1, x_2 \geq 0$					
Q5	a	In a bank cheques are cashed at a single 'teller' counter. Customers arrive at the counter in a Poisson manner at an average rate of 30 customers per hour. The teller takes, on an average, a minute and a half to cash a cheque The service time has been shown to be exponentially distributed. Calculate the % of time the teller is busy.					3

	b	**Solve any TWO of the following**	14
	(i)	Patients arrive at a clinic according to Poissondistribution at the rate of 30 patients per hour. The waiting room does not accommodate more than 14 patients. The examination time per patient is exponential with mean rate of 20 per hour. (a) Find number of patients in the clinic before the examination (b) What is the probability that an arriving patient will not wait. (c) What is the expected waiting time until a patient is discharged from the clinic ?	
	(ii)	Trucks arrival at a factory is for collecting finished goods that are supposed to be transported to distant markets. As and when they come they are required to join awaiting line and are served on first come, first served basis. Trucks arrive at the rate of 10 per hour where as the loading rate is 15 per hour. It is also given that arrivals are Poisson and loading is exponentially distributed. (a) Transporters have complained that their trucks have to wait for nearly 12 minutes at the plant. Examine whether the complaint is justified. (b) Determine the number of trucks waiting in the queue before getting loaded. (c) Find the probability that a truck cannot be loaded immediately.	
	(iii)	Customer arrives at a box office window, being manned by a single individual, according to a Poisson input process with a mean rate of 30 per hour. The time required to serve a customer has an exponential distribution with a mean of 90 seconds Find the average time spent by a customer. Also determine the average number of customers in the system and the average queue length	

30/5/2022 (E)

SOMAIYA VIDYAVIHAR UNIVERSITY

Semester: January 2022 – May 2022		
Maximum Marks: 100	**Examination: ESE Examination**	**Duration: 03 hrs**
Programme code: 54 **Programme: Honours in Data Science and Analytics**	**Class:** SY	**Semester:** IV (SVU 2020)
Name of the Constituent College: **K. J. Somaiya College of Engineering**	**Name of the department:** COMP	
Course Code: 116h54C401	**Name of the Course: Applied Data Science**	
Instructions: 1)Draw neat diagrams 2)Assume suitable data if necessary		

Question No.		Max. Marks
Q1 (a)	**Attempt any TWO (2) of the following** 1. Draw and explain the data science Process. 2. Explain Chi-squared distribution with suitable example. 3. Explain Significance of variance and bias with respect to machine learning.	10 Marks (05 Marks each)
Q1 (b)	1. Explain K-Fold Cross Validation with suitable diagram. 2. Explain Poisson distribution with example.	10 Marks
Q2 (a)	**Attempt any TWO (2) of the following** 1. Compare Python and R programming. List of packages of R useful for data science. 2. You're appointed as data scientist for second hand material selling company. Company having 1M customer base throughout Maharashtra and Gujrat state. You're asked to provide the data science based solution to increase the revenue. 3. What is X2 (chi-square) test? Perform test on given data and give inference over a result.	10 Marks (05 Marks each)
Q2 (b)	**Attempt any TWO (2) of the following** 1. Explain skewness and kurtosis with a diagram? 2. What is regression? Explain the usability of regression in data science. 3. Explain the use of Brownian motions in the finance model.	10 Marks (05 Marks each)
Q3 (a)	**Attempt any TWO (2) of the following** 1. Why data science is iterative process? Justify your answer with suitable example. 2. Explain how to treat missing values during data cleaning process. 3. What is outlier and error? Explain the treatment to remove it from dataset.	10 Marks (05 Marks each)

	Play chess	Not play chess	Sum (row)
Like science fiction	250(90)	200(360)	450
Not like science fiction	50(210)	1000(840)	1050
Sum(col.)	300	1200	1500

<table>
<tr><td>Q3 (b)</td><td>What is sampling? What is need of sampling? List types of sampling. Explain any of the sampling with suitable example.

Or

What is Data Normalization? Explain need of Data Normalization? Explain Min-Max normalization with suitable example.</td><td>10 Marks</td></tr>
<tr><td>Q4 (a)</td><td>What is K-means clustering? Give Algorithm. Calculate mean and final cluster for Given data: {2, 3, 4, 10, 8, 12, 3, 20, 30, 11, 27, 35, 28, 19}, K=2</td><td>10 Marks</td></tr>
<tr><td>Q4 (b)</td><td>What is house price prediction? Which algorithm is applicable for house price prediction? Explain algorithm in detail.

Or

"Linear regression is not suitable for classification problem", Justify with suitable example. Explain logistic regression</td><td>10 Marks</td></tr>
<tr><td>Q5 (a)</td><td>Consider the problem of comet detection. Suppose we have a dataset of detected and not detected comet and we train a naïve Bayes classifier on the dataset. For ten instances, the figure below shows the predictions of the trained classifier of the probability of a comet being detected. The classifier classifies an instance as comet if and only if the predicted probability is greater than 0.700. Draw the confusion matrix and calculate the accuracy, recall, F-measure (F1), Precision, Error rate, Sensitivity, Specificity.
<table>
<tr><th>Predicted Probability</th><th>Actual label</th></tr>
<tr><td>0.012</td><td>Not Detected</td></tr>
<tr><td>0.201</td><td>Not Detected</td></tr>
<tr><td>0.321</td><td>Not Detected</td></tr>
<tr><td>0.432</td><td>Not Detected</td></tr>
<tr><td>0.699</td><td>Not Detected</td></tr>
<tr><td>0.721</td><td>Detected</td></tr>
<tr><td>0.734</td><td>Detected</td></tr>
<tr><td>0.801</td><td>Detected</td></tr>
<tr><td>0.907</td><td>Detected</td></tr>
<tr><td>0.701</td><td>Detected</td></tr>
</table></td><td>10 Marks</td></tr>
<tr><td>Q5 (b)</td><td>Attempt any TWO (2) of the following
1. Draw and Explain different phases of NLP.
2. What is Word Embedding? What are their types? Explain.
3. Give importance of Feature selection with example.</td><td>10 Marks (05 Marks each)</td></tr>
</table>

30|5|2022 (E)

SOMAIYA VIDYAVIHAR UNIVERSITY

Semester: January 2022 – May 2022 **Maximum Marks: 100** **Examination: ESE Examination** **Duration: 3 Hrs.**		
Programme code: 55 **Programme: Cyber Security & Forensics (Honors).**	**Class:** SY	**Semester:** IV **(SVU 2020)**
Name of the Constituent College: **K. J. Somaiya College of Engineering**	**Name of the Department:** Computer Engineering.	
Course Code: 116h55C401	**Name of the Course: Cyber Security, Forensics & Cyber Law**	
Instructions: **1) All questions are compulsory.** **2) Draw neat diagrams, as required.** 3) **Assume suitable data if necessary.**		

Question No.		Max. Marks
Q1 (a)	Elaborate on any five social engineering attacks.	10
Q1 (b)	Explain different goals of security with supporting examples of each. **OR** Classify different categories of cyber-criminals.	10
Q2 (a)	What do you understand by identity theft? List down the reasons leading towards identity theft. **OR** What is software piracy? Explain its different types.	10
Q2 (b)	What are ransomwares? Explain the different strategies to combat ransomwares.	10
Q3 (a)	Discuss different sources of data theft. Mention the strategies to prevent data theft. **OR** What do you understand by data privacy? List down the data privacy mechanisms.	10
Q3 (b)	What is meant by chain of custody? What are the challenges an organization would have to face in case it could not establish the chain of custody for electronic evidence? Explain with suitable examples. **OR** Explain the role of the following in forensic investigations: i. Volatile and Non-volatile evidences. ii. Forensic Duplicates.	10
Q4 (a)	What do you understand by Email Forensics? Outline the tasks for investigating email crimes and violations.	05

Q4 (b)	Mumbai police have arrested a hacker by name ABC for hacking into a financial website. Although the hacker couldn't break into the main server of the financial institution, which was well secured by the financial institution. The accused person could make some addition to the home page of the financial website and has added a string of text to the news module of the home page of the website. Police were able to crack the case by following the trace left by the hacker on the web server of the financial institution. The financial institution has maintained a separate server for financial online transactions, for which utmost security has been taken by the financial institution. The website was hosted on a different server which comparatively had lesser security. The hacker is a 10th Pass youngster of 23 years old. He has done computer courses like CCNA, MCSE etc. But he is a computer addict. He sits before the computer for almost 16 to 20 times each day. He has mostly used the readymade hacking tools, to hack into any website. He goes to a particular website on the web, which facilitates him to see the entire directory structure of that website. Then using various techniques, such as obtaining a password file, he gets into the administrator's shoes and hacks the website. You are appointed as a cyber expert for handling of above mentioned case. Explain the investigation procedure based on above mentioned scenario: i. List contents of pre-incident kit required for forensic analysis of given case. (02) ii. Explain the concrete steps taken for immediate and overall incident response.(05) iii. List the data collected for analysis. (02) iv. List the respective tools needed to investigation and collection of above data items and write detailed process of retrieving the data with those tools. (04) v. Write your conclusions/findings. (02)	15
Q5 (a)	Illustrate on the following terms associated with cyber-crime & mention the appropriate section of IT Act 2000 & applicable penalties: (1) Infringement of IPR. (2) Hacking. (3) Forgery. (4) Defamation.	10
Q5 (b)	Write short notes on the following ***(Any two)***: (1) Digital Millennium Copyright Act (DMCA). (2) OWASP attacks on networks. (3) GDPR.	10

25/05/2022 (E)

SOMAIYA VIDYAVIHAR UNIVERSITY

Semester: January 2022 – May 2022 **Maximum Marks: 100** **Examination: ESE Examination** **Duration: 3 Hrs**		
Programme code: 01 **Programme: B.Tech**	**Class:** SY	**Semester:** IV **(SVU 2020)**
Name of the Constituent College: **K. J. Somaiya College of Engineering**	**Name of the department:** COMP	
Course Code: **116U01C404**	**Name of the Course: Theory of Automata with Compiler Design**	
Instructions: 1)Draw neat diagrams 2)Assume suitable data if necessary		

Question No.		Max. Marks
Q1 (a)	Given the alphabet set Σ={0,1}. Write Regular Expression for the following languages : i. Language L1 where all words must start with 0 and end with 1. ii. Language L2 where all words must contain 11. iii. Language L3 where all words either start with 1 or end with 01 or both. iv. Language L4 where all words are not having three or more consecutive 1s.	 2 2 3 3
Q1 (b)	Consider L1 = {aaab*} and L2= {a*bbb}, find regular expression for L1 U L2, L1 ∩ L2 and –L1 (complement of L1) and draw their respective Automata. OR i. Design a DFA for the Language L={w | w is of even length and begins with 01} ii. Design Mealy Machine to convert each occurrence of substring 1000 by 1001	10 5 5
Q2 (a)	i. Write a CFG for the Language L over the alphabet ∑ = { (,) } where the words are balanced parenthesis. e.g (( )) ( ) ( ) (( )). ii. Construct a Parse tree using LMD showing that the string ( ) ( ) ( ) is derived by the CFG. iii. What is the necessary condition when a CFG is called an Ambiguous grammar?	5 5 4
Q2 (b)	Eliminate Null Production from the given Grammar: S->ACB / CbB / Ba A-> da / BC B-> bC / ε C-> ab / ε	6

Q3 (a)	Construct PDA for the given CFG: S -> AB A -> BB B -> AB A -> a B -> a B -> b	10
Q3 (b)	Design PDA for recognizing $L = \{a^n b^{2n+1} \mid n>=1\}$. OR Convert the following grammar to Greibach Normal Form G = ({A,B,C}, {a,b},P,S) Where P consists of the following A-> BC B->CA / b C->AB / a	10
Q4 (a)	Design Turing Machine to increment the value of any binary number by one. The output should also be a binary number with value one more the number given. Show the simulation of input string "101101" on your Turing Machine. OR Design a TM for even length palindrome $L= ww^R \mid w \in (a+b)^*$. Show the simulation of your Turing Machine with the help of an example.	15
Q4 (b)	Write a short note on Multitape Turing Machine.	5
Q5 (a)	Explain Pumping Lemma. Prove that the following language on alphabet = {a,b} is not CFL: $L=\{a^n b^{2n} a^n \mid n>0\}$	5 5
Q5 (b)	Write short note on any two: i. Post correspondence Problem ii. Rice's Theorem iii. Recursively Enumerable Language iv. Halting problem of Turing Machine	10