CHRIST COLLEGE OF ENGINEERING
IRINJALAKUDA (AUTONOMOUS)
A CMI Institution

PG CURRICULUM – 2025

PG Programme Structure

FIRST SEMESTER: COMPUTER SCIENCE AND ENGINEERING

BRANCH: COMPUTER SCIENCE AND ENGINEERING

STREAM: CSE

Sl. No:	Slot	Course Code	Course Type	Course Category	Course Title (Course Name)	Credit			SS	Total		Credits	Hrs./Week
						L	T	P		CIE	ES		
THEORY													
1	A	R25CST103	BS	PC	Foundations of Computer Science	3	0	0	4.5	40	60	3	3
2	B	R25CST101	PC	PC	Advanced Machine Learning	3	0	0	4.5	40	60	3	3
3	C	R25CST102	PC	PC	Advanced Database Management	3	0	0	4.5	40	60	3	3
4	D	R25CST113	PE	PE	Program Elective I	3	0	0	4.5	40	60	3	3
5	E	R25CST123	PE	PE	Program Elective II	3	0	0	4.5	40	60	3	3
	F	G25RMT106	GC	GC	Research Methodology &IPR	2	0	0	3	40	60	2	2
PRACTICALS													
6	G	R25CSP107	PC	L	Computing Lab 1	0	0	2	1	100	0	1	3
Total									**26.5**			**18**	**20**

PROGRAM ELECTIVES														
Sl. No:	Slot	Course Code	Course Type	Course Category	Course Title (Course Name)	Credit				SS	Total		Credits	Hrs./Week
						L	T	P	R		CIA	ESE		
PROGRAM ELECTIVE-I (SEMESTER I)														
1	D	R25CST113	PE	PE1	Object Oriented Software Engineering	3	0	0	0	4.5	40	60	3	3
2	D	R25CST113	PE	PE1	Advanced Data Mining	3	0	0	0	4.5	40	60	3	3
3	D	R25CST113	PE	PE1	Cloud Computing	3	0	0	0	4.5	40	60	3	3
4	D	R25CST113	PE	PE1	Web Services	3	0	0	0	4.5	40	60	3	3
5	D	R25CST113	PE	PE1	Computational Intelligence	3	0	0	0	4.5	40	60	3	3
6	D	R25CST113	PE	PE1	Automated Verification	3	0	0	0	4.5	40	60	3	3
PROGRAM ELECTIVE-II (SEMESTER I)														
5	E	R25CST123	PE	PE2	Advanced Computer Networks	3	0	0	0	4.5	40	60	3	3
6	E	R25CST123	PE	PE2	Pattern Recognition	3	0	0	0	4.5	40	60	3	3
7	E	R25CST123	PE	PE2	Advanced Computer Architecture	3	0	0	0	4.5	40	60	3	3
8	E	R25CST123	PE	PE2	Natural Language Processing & Text Mining	3	0	0	0	4.5	40	60	3	3
9	E	R25CST123	PE	PE2	Advanced Compiler Design	3	0	0	0	4.5	40	60	3	3
10	E	R25CST123	PE	PE2	Bioinformatics	3	0	0	0	4.5	40	60	3	3

M.Tech. - COMPUTER SCIENCE AND ENGINEERING

SEMESTER I

Discipline: COMPUTER SCIENCE AND ENGINEERING
Stream: CSE
(COMPUTER SCIENCE AND ENGINEERING)

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
R25CST103	FOUNDATIONS OF COMPUTER SCIENCE	DISCIPLINE CORE	3	0	0	3

Course Objectives:

1. To expose the students to proof techniques and counting principles for developing mathematical reasoning in problem solving
2. To equip the students with the ability to apply recurrence relations, generating functions, and probability theory in computational contexts.
3. To equip the students with knowledge of algebraic structures and classical probability problems for applications in computer science.

Course Outcomes: After the completion of the course, the student will be able to

Course Outcomes		Bloom's Knowledge Level(KL)
CO 1	Apply Direct proof technique, Indirect proof technique and Mathematical Induction to prove various theorems and results	K3
CO 2	Solve counting problems using Pigeon hole principle, Principle of Inclusion exclusion, Permutations, Combinations, Cantor's Diagonalization argument and Derangements	K3
CO 3	Solve Recurrence relations and counting problems using Generating Functions	K3
CO 4	Solve problems on probability using the fundamentals of Probability, Bayes theorem, and Probability Distributions	K3
CO5	Solve problems using concepts in algebraic structures such as Groups, Cosets and Lagrange's Theorem.	K3
CO6	Design solutions for various computational problems using the mathematical concepts of computer science and prove the correctness of the solution developed.	K4

Note: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	3		2		2		
CO 2	2		2	3	2		
CO 3	2		2	3	2		
CO 4	3		2	3	2		
CO 5	2	2	2	3	2		
CO 6	3	3	3	3	3	2	

(1- Weak, 2-Medium, 3- strong)

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	**25**
Analyse	**25**
Evaluate	**5**
Create	**5**

Mark distribution

Total Marks	CIA	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation: 40 marks

Micro project/Course based project : 20 marks

Course based task/Seminar/ : 10 marks

Test paper, 1 no. : 10 marks

The project shall be done individually. Group projects not permitted.

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern

The end semester examination will be conducted by the University. There will be two parts; Part A and Part B. Part A contain 5 numerical questions with 1 question from each module, having 5 marks for each question. (such questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students). Students shall answer all questions.

Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

Total duration of the examination will be 150 minutes

SYLLABUS

MODULE	DESCRIPTION	CONTACT HOURS
1	**Theorem Proving Techniques** Theorem proving techniques: Direct Proof, Indirect proof - Proof by Contrapositive, Proof by contradiction and Proof by exhausting cases, Principle of mathematical induction, Complete induction and Well-ordering principle. The Pigeonhole principle	7
2	**Fundamentals of Counting** The Basics of counting, Addition and multiplication principles, Permutations and Combinations. Countable and uncountable sets, Principle of inclusion and exclusion – applications, derangements	8
3	**Generating Functions** Recurrence Relations, Modeling problems with recurrence relations. Generating functions, solving counting problems using Generating functions, solving recurrence relations using Generating functions.	8
4	**Probability Theory** Probability theory – Properties of Probability, Conditional Probability, Independent Events, Bayes Theorem, Mathematical Expectation and Variance of Random variables. Discrete Distributions and its mean and variance- Binomial Distribution, Bernoulli Distribution, Geometric Distribution, Poisson Distribution. Continuous Distributions and its mean and variance- Uniform and Exponential Distributions, Normal Distribution.	8
5	**Classic Problems in Probability and Algebraic Structures** Classic Problems in Probability- Birthday Paradox, The Hat Problem, Coupon Collector Problem.	8

	Groups and subgroups, generators for a group, Homomorphism theorems, cosets and normal subgroups, Lagrange's theorem.	

COURSE PLAN

No	Topic	No. of Lectures (40 hrs)
1	**Module 1: Theorem Proving Techniques (8)**	
1.1	Theorem proving techniques: Direct Proof	1
1.2	Indirect proof - Proof by Contrapositive	1
1.3	Proof by contradiction	1
1.4	Principle of mathematical induction, Complete induction	1
1.5	Principle of mathematical induction, Complete induction	1
1.6	Well-ordering principle	1
1.7	The Pigeonhole principle	1
1.8	The Pigeonhole principle	1
2	**Module 2: Fundamentals of Counting (7)**	
2.1	The Basics of counting, Addition and multiplication principles	1
2.2	Permutations and Combinations	1
2.3	Permutations and Combinations	1
2.4	Countable and uncountable sets	1
2.5	Countable and uncountable sets	1
2.6	Principle of inclusion and exclusion – applications	1
2.7	Derangements	1

3	**Module 3: Generating Functions (7)**	
3.1	Recurrence Relations, Modeling problems with recurrence relations	1
3.2	Generating functions	1
3.3	Generating functions	1
3.4	Solving counting problems using Generating functions	1
3.5	Solving recurrence relations using Generating functions	1
3.6	Solving recurrence relations using Generating functions	1
3.7	Solving recurrence relations using Generating functions	1
4	**Module 4: Probability Theory (10)**	
4.1	Probability theory – Properties of Probability	1
4.2	Conditional Probability	1
4.3	Independent Events	1
4.4	Bayes Theorem	1
4.5	Mathematical Expectation and Variance of Random variables	1
4.6	Discrete Distributions and its mean and variance- Binomial Distribution, Bernoulli Distribution	1
4.7	Geometric Distribution, Poisson Distribution	1
4.8	Continuous Distributions and its mean and variance	1
4.9	Uniform and Exponential Distributions	1
4.10	Normal Distribution	1

5	**Module 5: Classic Problems in Probability and Algebraic Structures(8)**	
5.1	Classic Problems in Probability- Birthday Paradox	1
5.2	The Hat Problem	1
5.3	Coupon Collector Problem	1
5.4	Coupon Collector Problem	1
5.5	Groups and subgroups, generators for a group	1
5.6	Homomorphism theorems	1
5.7	Cosets and normal subgroups	1
5.8	Lagrange's theorem	1

Reference Books

1. Kenneth H. Rosen, "Discrete Mathematics and its Applications" 7/e, McGraw Hill Inc, 2011.

2. J. P. Tremblay, R. Manohar, "Discrete Mathematical Structures with Application to Computer Science" Tata McGrawHill, 2000

3. Sheldon M. Ross, "Introduction to Probability Models"

4. Sheldon M. Ross, "A First Course in Probability"

5. William Feller, "An introduction to probability theory and its applications" Volume1.Wiley,1957.

6. Rajeev Motwani and Prabhakar Raghavan, "Randomized Algorithms" Cambridge University Press 1995

MODEL QUESTION PAPER

Course Code: R25CST103

FOUNDATIONS OF COMPUTER SCIENCE

Max. Marks: 60 Duration: 2.5 hours

PART A

(Answer all Questions: Each question carries 5 marks)

1. Prove that √3 is irrational using proof by contradiction.
2. Show that the set of real numbers is uncountable using Cantor's diagonalization principle.
3. In how many different ways can eight identical cookies be distributed among three distinct children if each child receives at least two cookies and no more than four cookies?
4. A woman has 11 close friends and she wants to invite 5 of them to dinner. In how many ways can she invite them if

   (i) there is no restriction on the choice.

   (ii) two particular persons will not attend separately.

   (iii) two particular persons will not attend together. Explain briefly the components of time series.
5. State and prove Birthday Paradox.

PART B

(Answer any five questions: Each carry 7 marks)

6. (a) Determine which amounts of postage can be formed using four- and seven-rupees stamps. Prove your answer using principle of mathematical induction. (4)
   (b) Prove your answer using strong induction. (3)
7. Every sequence of $(n^2 + 1)$ distinct real numbers contains a subsequence of length $(n + 1)$ that is either strictly increasing or strictly decreasing. Prove the statement using Pigeonhole principle. (7)
8. (a) (i) Find the number of permutations that can be formed from the letters of the string "ELEVEN". (4)

   (ii) How many of them begin and end with E?

   (iii) How many of them have three Es together?

   (iv) How many begin with E and end with N?

(b) Determine the number of integers between 1 and 10000 that are not divisible by 6, 7, or 8. (3)

9. Solve the recurrence relation $a_n = 4a_{n-1} = 6 \times 4^n$ with initial condition $a_0 = 1$, using generating function. (7)

10. Entry to a certain University is determined by a national test. The scores on this test are normally distributed with a mean of 500 and a standard deviation of 100. Tom wants to be admitted to this university and he knows that he must score better than at least 70% of the students who took the test. Tom takes the test and scores 585. Will he be admitted to this university? Why? (7)

11. There are n distinct coupons placed in an urn. Coupons are randomly selected one at a time (with replacement) until at least one of each type of coupon has been selected. Find the expected number of selections made until all distinct coupons are collected. (7)

12. State and prove Lagrange's Theorem. (7)

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
R25CST101	ADVANCED MACHINE LEARNING	PROGRAM CORE 1	3	0	0	3

Course Objective

1. To develop a solid understanding of fundamental machine learning paradigms including supervised, unsupervised, semi-supervised, and reinforcement learning, along with parameter estimation and regression methods.
2. To equip learners with the ability to apply core machine learning algorithms such as regression, classification, clustering, neural networks, and support vector machines, while understanding their mathematical foundations and implementation techniques.
3. To provide knowledge of regularization, ensemble methods, and advanced techniques such as boosting, kernel tricks, and graphical models for building efficient and generalized predictive models.
4. To enable students to critically evaluate and validate machine learning models using appropriate performance metrics, resampling methods, and bias-variance analysis, with attention to practical aspects such as data preprocessing and model selection.

Course Outcomes: After the completion of the course the student will be able to:*

Course Outcome		Bloom's Knowledge Level(KL)
CO 1	Analyse the Machine Learning concepts, classifications of Machine Learning algorithms and basic parameter estimation methods.	K4
CO 2	Illustrate the concepts of regression and classification techniques	K3
CO 3	Describe unsupervised learning concepts and dimensionality reduction techniques.	K3
CO 4	Explain Support Vector Machine concepts and graphical models.	K3
CO 5	Choose suitable model parameters for different machine learning techniques and to evaluate a model performance.	K3
CO6	Design, implement and analyse machine learning solution for a real world problem.	K6

Note: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1		2	2				
CO 2		2	2				
CO 3					2		
CO 4					3		
CO 5					3		
CO 6			2	2	3	3	

(1-Weak, 2-Medium, 3- Strong)

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	48
Analyse	12
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation : 40 marks

Micro project/Course based project : 20 marks

Course based task/Seminar/Quiz : 10 marks

Test paper, 1 no. : 10 marks

The project shall be done individually. Group projects not permitted. Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the University. There will be two parts; Part A and Part B. Part A contain 5 numerical questions with 1 question from each module, having 5 marks for each question. (such questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students). Students shall answer all questions.

Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

Total duration of the examination will be 150 minutes.

SYLLABUS

Module	Description	Contact Hours
1	Parameter Estimation and Regression: Overview of machine learning: supervised, semi-supervised, unsupervised learning, reinforcement learning. Basics of parameter estimation: Maximum Likelihood Estimation(MLE), Maximum a Posteriori Estimation (MAP). Gradient Descent Algorithm, Batch Gradient Descent, Stochastic Gradient Descent. Regression algorithms: least squares linear regression, normal equations and closed form solution, Polynomial regression.	8
2	**Regularization techniques and Classification algorithms:** Overfitting, Regularization techniques - LASSO and RIDGE. Classification algorithms: linear and non-linear algorithms, Perceptrons, Logistic regression, Naive Bayes, Decision trees. Neural networks : Concept of Artificial neuron, Feed-Forward Neural Network, Back propagation algorithm.	9
3	**Unsupervised learning:** Unsupervised learning: clustering, k-means, Hierarchical clustering, Principal component analysis,	8

	Density-based spatial clustering of applications with noise (DBSCAN). Gaussian mixture models: Expectation Maximization (EM) algorithm for Gaussian mixture model.	
4	**Support Vector Machine and Graphical Models:** Support vector machines and kernels : Max margin classification, Nonlinear SVM and the kernel trick, nonlinear decision boundaries, Kernel functions. Basics of graphical models - Bayesian networks, Hidden Markov model - Inference and estimation.	**7**
5	**Evaluation Metrics and Sampling Methods:** Classification Performance Evaluation Metrics: Accuracy, Precision, Precision, Recall, Specificity, False Positive Rate (FPR), F1 Score, Receiver Operator Characteristic (ROC) Curve, AUC. Regression Performance Evaluation Metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R Squared/Coefficient of Determination. Clustering Performance Evaluation Metrics: Purity, Jaccard index, Normalized Mutual Information, Clustering Accuracy, Silhouette Coefficient, Dunn's Index. Boosting: AdaBoost, gradient boosting machines. Resampling methods: cross-validation, bootstrap. Ensemble methods: bagging, boosting, random forests Practical aspects in machine learning: data preprocessing, overfitting, accuracy estimation, parameter and model selection Bias-Variance tradeoff	**8**

COURSE PLAN

No	Topics	No. of Lectures (40)
1	**Module-1 (Parameter Estimation and Regression) 8 hours**	
1.1	Overview of machine learning: supervised, semi-supervised, unsupervised learning, reinforcement learning.	1
1.2	Basics of parameter estimation: Maximum Likelihood Estimation(MLE)	1
1.3	Basics of parameter estimation: Maximum Likelihood Estimation(MLE) - Examples	1
1.4	Basics of parameter estimation: Maximum a Posteriori Estimation (MAP)	1
1.5	Basics of parameter estimation: Maximum a Posteriori Estimation (MAP) - Example	1
1.6	Gradient Descent Algorithm, Batch Gradient Descent, Stochastic Gradient Descent	1

1.7	Regression algorithms: least squares linear regression, normal equations and closed form solution	1
1.8	Polynomial regression	1
2	**Module-2 (Regularization techniques and Classification algorithms) 9 hours**	
2.1	Overfitting, Regularization techniques - LASSO and RIDGE	1
2.2	Classification algorithms: linear and non-linear algorithms	1
2.3	Perceptrons	1
2.4	Logistic regression	1
2.5	Naive Bayes	1
2.6	Decision trees	1
2.7	Neural networks : Concept of Artificial neuron	1
2.8	Feed-Forward Neural Network	1
2.9	Back propagation algorithm	1
3	**Module-3 (Unsupervised learning) 8 hours**	
3.1	Unsupervised learning: clustering, k-means	1
3.2	Hierarchical clustering	1
3.3	Principal component analysis	1
3.4	Density-based spatial clustering of applications with noise (DBSCAN)	1
3.5	Gaussian mixture models: Expectation Maximization (EM) algorithm for Gaussian mixture model	1
3.6	Gaussian mixture models: Expectation Maximization (EM) algorithm for Gaussian mixture model	1
4	**Module-4 (Support Vector Machine and Graphical Models) 7 hours**	
4.1	Support vector machines and kernels : Max margin classification	1
4.2	Support vector machines: Max margin classification	1
4.3	Nonlinear SVM and the kernel trick, nonlinear decision boundaries	1
4.3	Kernel functions	1
4.5	Basics of graphical models - Bayesian networks	1
4.6	Hidden Markov model - Inference and estimation	1
4.7	Hidden Markov model - Inference and estimation	1

4.8	Hidden Markov model - Inference and estimation	1
5	**Module-5 (Evaluation Metrics and Sampling Methods) 8 hours**	
5.1	Classification Performance Evaluation Metrics: Accuracy, Precision, Precision, Recall, Specificity, False Positive Rate (FPR), F1 Score, Receiver Operator Characteristic (ROC) Curve, AUC	1
5.2	Regression Performance Evaluation Metrics: Mean Absolute Error	1
	(MAE), Root Mean Squared Error (RMSE), R Squared/Coefficient of Determination	1
5.3	Clustering Performance Evaluation Metrics: Purity, Jaccard index, Normalized Mutual Information, Clustering Accuracy, Silhouette Coefficient, Dunn's Index	1
5.4	Boosting: AdaBoost, gradient boosting machines.	1
5.5	Resampling methods: cross-validation, bootstrap.	1
5.6	Ensemble methods: bagging, boosting, random forests	1
5.7	Practical aspects in machine learning: data preprocessing, overfitting, accuracy estimation, parameter and model selection	1
5.8	Bias-Variance tradeoff	1

Reference Books

1. Christopher Bishop. Neural Networks for Pattern Recognition, Oxford University Press, 1995.
2. Kevin P. Murphy. Machine Learning: A Probabilistic Perspective, MIT Press 2012.
3. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements Of Statistical Learning, Second edition Springer 2007.
4. Ethem Alpaydin, Introduction to Machine Learning, 2nd edition, MIT Press 2010.
5. Tom Mitchell, Machine Learning, McGraw-Hill, 1997.

MODEL QUESTION PAPER

Course Code: R25CST101

ADVANCED MACHINE LEARNING

Max. Marks: 60 Duration: 2.5 hours

PART A

(Answer *ALL* questions; each question carries 5 marks)

1. Explain the principle of the gradient descent algorithm.
2. In a two-class logistic regression model, the weight vector **w = [4, 3, 2, 1, 0]**. We apply it to some object that we would like to classify; the vectorized feature representation of this object is $\mathbf{x} = [-2, 0, -3, 0.5, 3]$. What is the probability, according to the model, that this instance belongs to the positive class?
3. Expectation maximization (EM) is designed to find a maximum likelihood setting of the parameters of model when some of the data is missing. Does the algorithm converge? If so, do you obtain a locally or globally optimal set of parameters?
4. What is the basic idea of a Support Vector Machine?
5. What is the trade-off between bias and variance?

PART B

(Answer *any FIVE* questions; each question carries 7 marks)

6. Suppose that ***X*** is a discrete random variable with the following probability mass function: where $0 \leq \theta \leq 1$ is a parameter. The following 10 independent observations were taken from such a distribution: ***(3, 0, 2, 1, 3, 2, 1, 0, 2, 1)***. What is the maximum likelihood estimate of θ.
7. Derive the gradient descent training rule assuming for the target function ***od*** = ***w0 + w1x1 + ... + wnxn***. Define explicitly the squared cost/error function ***E***, assuming that a set of training examples ***D*** is provided, where each training example ***d ε D*** is associated with the target output ***td***.
8. Cluster the following eight points representing locations into three clusters:

A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9) .

Initial cluster centers are: ***A1(2, 10)***, ***A4(5, 8)*** and ***A7(1, 2)***.

The distance function between two points ***a = (x1, y1)*** and ***b = (x2, y2)*** is

defined as ***D(a, b) = |x2 – x1| + |y2 – y1|***

Use **k**-Means Algorithm to find the three cluster centers after the second

iteration.

9. Describe Principal Component Analysis. What criterion does the method minimize? What is the objective of the method? Give a way to compute the solution from a matrix ***X*** encoding the features. Distinguish between break-even analysis and benefit cost analysis.
10. How does random forest classifier work? Why is a random forest better than a decision tree?
11. Consider a two-class classification problem of predicting whether a

    photograph contains a man or a woman. Suppose we have a test dataset of 10

    records with expected outcomes and a set of predictions from our

    classification algorithm. Compute the confusion matrix, accuracy, precision, recall, sensitivity and specificity on the following data.

Sl.No.	Actual	Predicted
1	man	woman
2	man	man
3	woman	woman
4	man	man
5	man	woman
6	woman	woman
7	woman	man
8	man	man
9	man	woman
10	woman	woman

12. Classifier A attains 100% accuracy on the training set and 70% accuracy on the test set. Classifier B attains 70% accuracy on the training set and 75% accuracy on the test set. Which one is a better classifier. Justify your answer.

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
R25CST102	ADVANCED DATABASE MANAGEMENT	PROGRAM CORE 1	3	0	0	3

Course Objective:

1. To provide an in-depth understanding of advanced query processing techniques including indexing, hashing, cost analysis, and optimization strategies for efficient data retrieval.
2. To develop knowledge of database security challenges and solutions, including threats, access control mechanisms, and preventive measures against vulnerabilities such as SQL injection.
3. To impart conceptual and practical insights into modern database architectures, covering centralized, client-server, parallel, and distributed systems along with concurrency control, commit protocols, and query processing.
4. To explore advanced data models and indexing techniques by introducing semi-structured data (XML), object-relational systems, and specialized indexing methods such as Bloom filters, bitmap indices, spatial indices, and hash-based approaches.

Course Outcomes: After the completion of the course the student will be able to

Course Outcome		Bloom's Knowledge Level
CO 1	Identify various measures of query processing and optimization.	K3
CO 2	Analyze and implement security mechanisms to secure a database system.	K2
CO 3	Apply knowledge and awareness of the different database architectures in different scenarios	K3
CO 4	Analyze implementation aspects of distributed system on database architecture	K2
CO 5	make use of semi structured data, XML and XML queries for data management	K3
CO 6	Design, Develop and Implement innovative ideas on advanced database concepts and techniques	K6

Note: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create

Mapping of Course Outcomes with Program Outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	3		1				1
CO 2	3		1	2			2
CO 3	3		1	2			2
CO 4	3		1	2			1

CO 5	3		1	2			1
CO 6	3	1	1	2			2

(1-Weak, 2-Medium, 3- Strong)

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	70-80%
Analyze	30-40%

Mark Distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 Hrs

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design-based questions (for both internal and end semester examinations).

Continuous Internal Evaluation : 40 marks

- Micro project/Course based project: 20 marks
- Course based task/Seminar/Quiz: 10 marks
- Test paper, 1 no. : 10 marks

The project shall be done individually. Group projects not permitted. Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the University. There will be two parts; Part A and Part B. Part A contain 5 numerical questions with 1 question from each module, having 5 marks for each question. (such questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students). Students shall answer all questions.

Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

Total duration of the examination will be 150 minutes.

SYLLABUS

Module	Description	Contact Hours
1	Review of indexing and Hashing - Overview- Measures of query cost- Algorithms for Selection and Join with cost analysis- Evaluation of expressions- Optimization of RA expressions.	9
2	Threats to databases, control measures, database security and DBA, Discretionary access control, Mandatory access control (role-based only), SQL injection.	9
3	Centralized and Client-Server Architectures – Centralized server systems - Server System Architectures - Parallel Systems- - Parallel storage - Data partitioning, replication and indexing in Parallel Databases- Parallel query processing.	9
4	Distributed System architecture- Distributed storage - Distributed file systems – Distributed RDB design- Transparency– Distributed Transactions - Commit Protocols – Concurrency Control - Distributed Query Processing Advanced indexing Techniques: Bloom filter - Bitmap indices - Indexing spatial data - Hash indices.	9
5	Semi-structured Data and XML Databases: XML Data Model – XSD – XPath and XQuery – Example Queries. Native XML databases, Object Relational Systems	9

COURSE PLAN

No	Topic	No. of Lectures (40 Hours)
1	**Module-1 Query Processing and Optimization(8)**	
1.1	Introduction to Query Processing and Optimization	1
1.2	Review of indexing	1
1.3	Hashing - Overview	1
1.4	Measures of query cost	1
1.5	Algorithms for Selection with cost analysis	1
1.6	Algorithms for Join with cost analysis	1
1.7	Evaluation of expressions	1
1.8	Optimization of RA expressions	1
2	**Module-2 Database Security (7)**	
2.1	Introduction to Database Security	1
2.2	Threats to databases	1
2.3	Control measures	1
2.4	Database security and DBA	1

2.5	Discretionary access control	1
2.6	Mandatory access control (role-based only)	1
2.7	SQL injection	1
3	**Module-3 Database System Architectures (9)**	
3.1	Introduction to Database System Architecture	1
3.2	Overview of Centralized and Client-Server Architectures	1
3.3	Centralized server systems	1
3.4	Server System Architectures	1
3.5	Parallel Systems	1
3.6	Parallel storage	1
3.7	Data partitioning, replication in Parallel Databases	1
3.8	Indexing in Parallel Databases	1
3.9	Parallel query processing	1
4	**Module-4 Distributed System Architecture(10)**	
4.1	Introduction to Distributed System architecture	1
4.2	Distributed storage & Distributed file systems	1
4.3	Distributed RDB design & its Transparency	1
4.4	Distributed Transactions	1
4.5	Commit Protocols & Concurrency Control	1
4.6	Distributed Query Processing	1
4.7	Advanced indexing Techniques: Bloom filter	1
4.8	Bitmap indices	1
4.9	Indexing spatial data	1
4.10	Hash indices	1
5	**Module-5 Semi-structured Data (6)**	
5.1	Introduction to Semi-structured Data and XML Databases	1
5.2	XML Data Model – XSD	1
5.3	XPath and XQuery	1
5.4	Example Queries	1

5.5	Native XML databases	1
5.6	Object Relational Systems	1

Reference Books

1. R. Elmasri, S.B. Navathe, “Fundamentals of Database Systems”, 7/e, Pearson Education/ Addison Wesley, 2016
2. Thomas Cannolly and Carolyn Begg, “Database Systems, A Practical Approach to Design, Implementation and Management”, 3/e, Pearson Education, 2010.
3. Henry F Korth, Abraham Silberschatz, S. Sudharshan, “Database System Concepts”, 7/e, Tata McGraw Hill, 2019.
4. Joe Fawcett, Danny Ayers, Liam R. E. Quin, Beginning XML, 5/e, John Wiley & Sons, 2012 Grigoris Antoniou. Frank van Harmelen, “A Semantic Web Primer”, The MIT Press,Cambridge, Massachusetts, 2003.

MODEL QUESTION PAPER

Course Code: R25CST102

ADVANCED DATABASE MANAGEMENT

Max. Marks: 60 Duration: 2.5 hours

PART A

Answer All Questions. Each Question Carries 5 Marks

1.	Why is it not desirable to force users to make an explicit choice of a query-processing strategy? Are there cases in which it is desirable for users to be aware of the costs of competing query-processing strategies? Explain your answer.	5
2.	What are the relative merits of using Discretionary Access Control or Mandatory Access Control? What is role-based access control? In what ways is it superior to DAC and MAC?	5
3	Suppose relation r is stored partitioned and indexed on A, and s is stored partitioned and indexed on B. Consider the query: $r.C\gamma_{count(s.D)}((\sigma_{A>5}(r)) \bowtie_{r.B=s.B} s)$ i. Give a parallel query plan using the exchange operator, for computing the subtree of the query involving only the select and join operators. Now extend the above to compute the aggregate. Make sure to use preaggregation to minimize the data transfer.	5
4	Insert and query on a Bloom filter of size m = 10 and number of hash functions k= 3. Let H(x) denote the result of the three hash functions which will write as a set of three values {h 1 (x), h 2 (x), h 3 (x)}. Has functions used: A= x mod 10, B= x mod 7, C= (sum of digits) mod 9.	5
5	Design an XML document for storing hostel mess food details (meals taken such as breakfast, lunch, dinner) with their charges for the month of June 2022. Charges may vary depending on the food taken. Students can opt not to take any meals on certain days. Write a sample XML for 2 students for 2 days. Write a XQuery to return the lunch details of all. i Create an XSD for the same.	5

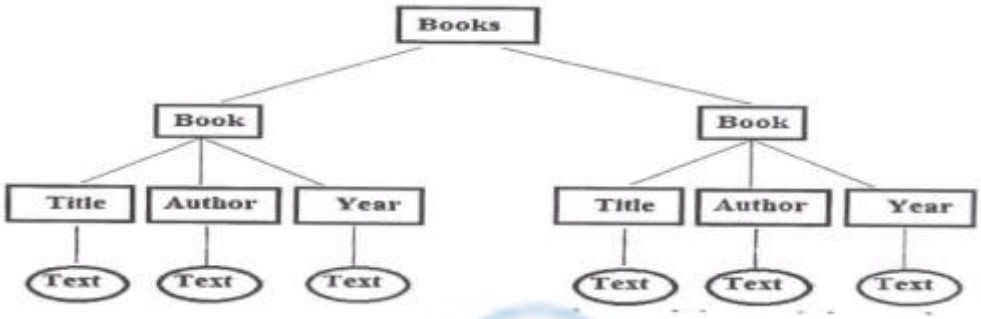
Part B

(Answer any five questions. Each question carries 7 marks)

5a	Consider the issue of interesting orders in optimization. Suppose you are given a query that computes the natural join of a set of relations S. Given a subset S1 of S, what are the interesting orders of S1?	4
b	Suppose you want to get answers to r ⋈ s sorted on an attribute of r, and want only the top K answers for some relatively small K. Give a good way of evaluating the query: i. When the join is on a foreign key of r referencing s, where the foreign key attribute is declared to be not null.	3

		When the join is not on a foreign key.	
6		A database relation may have the values of certain attributes encrypted for security. Why do database systems not support indexing on encrypted attributes? Using your answer to this question, explain why database systems do not allow encryption of primary-key attributes.	7
7		A database relation may have the values of certain attributes encrypted for security. Why do database systems not support indexing on encrypted attributes? Using your answer to this question, explain why database systems do not allow encryption of primary-key attributes.	7
8		If a parallel data-store is used to store two relations *r* and *s* and we need to join r and *s*, it may be useful to maintain the join as a materialized view. What are the benefits and overheads in terms of overall throughput, use of space, and response time to user queries? Explain in detail.	7
9		Consider the bitmap representation of the free-space map, where for each block in the file, two bits are maintained in the bitmap. If the block is between 0 and 30 percent full the bits are 00, between 30 and 60 percent the bits are 01, between 60 and 90 percent the bits are 10, and above 90 percent the bits are 11. Such bitmaps can be kept in memory even for quite large files.	
	i	Outline two benefits and one drawback to using two bits for a block, instead of one byte as described earlier in this chapter.	7
	ii	Describe how to keep the bitmap up to date on record insertions and deletions. Outline the benefit of the bitmap technique over free lists in searching for free space and in updating free space information	
10		Assume a relationship RAB at site 1 and relationship SCD at site 2 as follows: R= {(1,2), (3,4), (5, 6), (7, 8), (9, 10)} S= {(1, 0), (8,1), (9, 2), (10, 3), (11, 4)} Compute R⋈ S using bloom join with A=C and explain the intermediate steps. Show the tuples transferred with the hash function mod 4.	7
11a		Consider the country data. Write XPath for the following:	
	iii	Return the area of India. Return the names of all countries with population greater than 100 million. Return the names of all countries whose population is less than one thousandth that of some city (in any country). Return the names of all cities that have the same name as the country in which they are located.	4
b		Consider the country data. Write XQuery for the following: Return the name of the country with the highest population. Return the name of the country that has the city with the highest population. Return the average population of Russian-speaking countries.	3

Consider the following XML Tree



12 7

Write an XML schema for the above, and also provide an XQuery expression to get the books published in the year 1992.

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
G25RMT106	RESEARCH METHODOLOGY & IPR	GENERAL COURSE	2	0	0	2

Course Objective

1. To develop research aptitude by fostering creativity, problem identification through literature survey, and adoption of suitable methodologies for solving engineering problems.
2. To enhance professional skills by delivering effective technical presentations, preparing quality technical reports, and disseminating research outcomes through publications or patents.

Course Outcomes: After the completion of the course the student will be able to

Course Outcome		Bloom's Knowledge Level(KL)
CO 1	Approach research projects with enthusiasm and creativity	K3
CO 2	Conduct literature survey and define research problem	K4
CO 3	Adopt suitable methodologies for solution of the problem	K3
CO 4	Deliver well-structured technical presentations and write technical reports.	K3
CO 5	Publish/Patent research outcome.	K4

Note: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	2	2				2	
CO 2	2	2				2	
CO 3	2	2				2	
CO 4	2	2				2	
CO 5	2	2				2	

(1-Weak, 2-Medium, 3- Strong)

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	70
Analyse	30
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Course based task: 15 marks

Some sample course based tasks that can be performed by the student given below.

Conduct a group discussion based on the good practices in research.

Conduct literature survey on a suitable research topic and prepare a report based on this.

Seminar: 15 marks

Test paper: 10 marks

.

End Semester Examination Pattern:

Total Marks: 60

The examination will be conducted by the respective college with the question provided by the University. The examination will be for 150 minutes and contain two parts; Part A and Part B. Part A will contain 6 short answer questions with 1 question each from modules 1 to 4, and 2 questions from module 5. Each question carries 5 marks. Part B will contain only 1 question based on a research article from the respective discipline and carries 30 marks. The students are to answer the questions based on that research article.

SYLLABUS

MODULE	DESCRIPTION	CONTACT HOURS
1	**Introduction:** Meaning and significance of research, Skills, habits and attitudes for research, Types of research, Characteristics of good research, Research process Motivation for research: Motivational talks on research: "You and Your Research"- Richard Hamming. Thinking skills: Levels and styles of thinking, common sense and scientific thinking, examples, logical thinking, division into sub-problems, verbalization and awareness of scale. Creativity: Some definitions, illustrations from day to day life, intelligence versus creativity, creative process, requirements for creativity	6
2	**Literature survey and Problem definition:** Information gathering – reading, searching and documentation, types of literature. Integration of research literature and identification of research gaps. Attributes and sources of research problems, problem formulation, Research question, multiple approaches to a problem. Problem solving strategies – reformulation or rephrasing, techniques of representation, Importance of graphical representation, examples. Analytical and analogical reasoning, examples, Creative problem solving using Triz, Prescriptions for developing creativity and problem solving.	6
3	**Experimental and modelling skills:** Scientific method, role of hypothesis in experiment, units and dimensions, dependent and independent variables, control in experiment, precision and accuracy, need for precision, definition, detection, estimation and reduction of random errors, statistical treatment of data, definition, detection and elimination of systematic errors, Design of experiments, experimental logic, documentation. Types of models, stages in modelling, curve fitting, the role of approximations, problem representation, logical reasoning, mathematical skills. Continuum/meso/micro scale	5

	approaches for numerical simulation, Two case studies illustrating experimental and modelling skills.	
4	**Effective communication - oral and written:** Examples illustrating the importance of effective communication, stages and dimensions of a communication process. Oral communication –verbal and non-verbal, casual, formal and informal communication, interactive communication, listening, form, content and delivery, various contexts for speaking- conference, seminar etc. Guidelines for preparation of good presentation slides. Written communication – Rules of scientific writing, form, content and language, layout, typography and illustrations, nomenclature, reference and citation styles, contexts for writing – paper, thesis, reports etc. Tools for document preparation-LaTeX. Common errors in typing and documentation	5
5	**Publication and Patents**: Relative importance of various forms of publication, Choice of journal and reviewing process, Stages in the realization of a paper. Research metrics- Journal level, Article level and Author level, Plagiarism and research ethics. Introduction to IPR, Concepts of IPR, Types of IPR. Common rules of IPR practices, Types and Features of IPR Agreement, Trademark. Patents- Concept, Objectives and benefits, features, Patent process – steps and procedures	6

COURSE PLAN

No	Topic	No. of Lectures
1	**Module-1 Introduction(5)**	
1.1	Meaning and significance of research, Skills, habits and attitudes for research, Types of research,	1
1.2	Characteristics of good research, Research process	1
1.3	Motivation for research: Motivational talks on research: "You and Your Research"- Richard Hamming	1
1.4	Thinking skills: Levels and styles of thinking, common sense and scientific thinking, examples, logical thinking, division into sub-problems, verbalization and awareness of scale.	1
1.5	Creativity: Some definitions, illustrations from day to day life, intelligence versus creativity, creative process, requirements for creativity	1
2	**Module-2 Literature survey and Problem definition(5)**	
2.1	Information gathering – reading, searching and documentation, types of literature.	1
2.2	Integration of research literature and identification of research gaps	1
2.3	Attributes and sources of research problems, problem formulation, Research question, multiple approaches to a problem	1
2.4	Problem solving strategies – reformulation or rephrasing, techniques of representation, Importance of graphical representation, examples.	1
2.5	Analytical and analogical reasoning, examples, Creative problem solving using Triz, Prescriptions for developing creativity and problem solving.	1
3	**Module-3 Experimental and modelling skills(5)**	
3.1	Scientific method, role of hypothesis in experiment, units and dimensions, dependent and independent variables, control in experiment	1
3.2	precision and accuracy, need for precision, definition, detection, estimation and reduction of random errors, statistical treatment of data, definition, detection and elimination of systematic errors	1
3.3	Design of experiments, experimental logic, documentation	1

3.4	Types of models, stages in modelling, curve fitting, the role of approximations, problem representation, logical reasoning, mathematical skills.	1
3.5	Continuum/meso/micro scale approaches for numerical simulation, Two case studies illustrating experimental and modelling skills.	1
4	**Module-4 Effective communication - oral and written(5)**	
4.1	Examples illustrating the importance of effective communication, stages and dimensions of a communication process	1
4.2	Oral communication –verbal and non-verbal, casual, formal and informal communication, interactive communication, listening, form, content and delivery, various contexts for speaking- conference, seminar etc	1
4.3	Guidelines for preparation of good presentation slides	1
4.4	Written communication – Rules of scientific writing, form, content and language, layout, typography and illustrations, nomenclature, reference and citation styles, contexts for writing – paper, thesis, reports etc. Tools for document preparation-LaTeX.	1
4.5	Common errors in typing and documentation	1
5	**Module-5 Publication and Patents(6)**	
5.1	Relative importance of various forms of publication, Choice of journal and reviewing process, Stages in the realization of a paper.	1
5.2	Research metrics-Journal level, Article level and Author level, Plagiarism and research ethics	1
5.3	Introduction to IPR, Concepts of IPR, Types of IPR	1
5.4	Common rules of IPR practices, Types and Features of IPR Agreement, Trademark	1
5.5	Patents- Concept, Objectives and benefits, features, Patent process – steps and procedures	2

Reference Books

1. E. M. Phillips and D. S. Pugh, "How to get a PhD - a handbook for PhD students and their supervisors", Viva books Pvt Ltd
2. G. L. Squires, "Practical physics", Cambridge University Press.

3. Antony Wilson, Jane Gregory, Steve Miller, Shirley Earl, Handbook of Science Communication, Overseas Press India Pvt Ltd, New Delhi, 1st edition 2005
4. C. R. Kothari, Research Methodology, New Age International, 2004
5. Panneerselvam, Research Methodology, Prentice Hall of India, New Delhi, 2012
6. Leedy P. D., Practical Research: Planning and Design, McMillan Publishing Co
7. Day R. A., How to Write and Publish a Scientific Paper, Cambridge University Press, 1989.
8. William Strunk Jr., Elements of Style, Fingerprint Publishing, 2020
9. Peter Medawar, 'Advice to Young Scientist', Alfred P. Sloan Foundation Series, 1979.
10. E. O. Wilson, Letters to a Young Scientist, Liveright, 2014.
11. R. Hamming, You and Your Research, 1986 Talk at Bell Labs.

MODEL QUESTION PAPER

Course Code: G25RMT106

RESEARCH METHODOLOGY & IPR

Max. Marks: 60 Duration: 2.5 hours

PART A

	Answer all questions. Each question carries 5 marks	Marks
1	Discuss the salient recommendations for great research Marks 30 recommended by Richard Hamming in his famous talk “You and Your Research”	5
2	What are the characteristics of a good research question? Discuss with an example.	5
3	Explain the difference between continuum, meso-scale and micro scale approaches for numerical simulation.	5
4	Discuss any four rules of scientific writing.	5
5	What are the requirements for patentability?	5
6	What are the differences between copyright and trademark protection?	5
	Read the given research paper and write a report that addresses the following issues (The paper given can be specific to the discipline concerned)	
7	What is the main research problem addressed?	3

8	Identify the type of research	3
9	Discuss the short comings in literature review if any?	6
10	Discuss appropriateness of the methodology used for the study	6
11	Discuss the significance of the study and summarize the important results and contributions by the authors	6
12	Identify limitations of the article if any.	6

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
R25CSP107	COMPUTING LAB 1	LABORATORY	0	0	3	2

Course Objectives:

1. To enable the learners to make use of the machine learning concepts and algorithms to derive data insights.

2. To provide exposure to the design and implementation aspects of machine learning algorithms such as decision trees, regression, naive bayes algorithm, clustering algorithms and artificial neural network.

Course Outcomes: After the completion of the course the student will be able to

Course Outcomes		Bloom's Knowledge Level(KL)
CO 1	Apply modern machine learning notions in predictive data analysis	K3
CO 2	Analyze the range of machine learning algorithms along with their strengths and weaknesses	K4
CO 3	Design and develop appropriate machine learning models to solve real world problems.	K4
CO 4	Build predictive models from data and analyze their performance	K6

Note: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	2	2	2	2	2	2	
CO 2	2	2	2	2	2	2	
CO 3	2	2	2	2	2	2	
CO 4	3	3	2	3	2	3	

(1- Weak, 2-Medium, 3- strong)

Mark Distribution

Total Marks	CIE	ESE	ESE Duration
100	100		

The laboratory courses will be having only Continuous Internal Evaluation and carries 100 marks. Final assessment shall be done by two examiners; one examiner will be a senior faculty from the same department.

LIST OF EXPERIMENTS

1. Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.

2. Write a program to implement the naïve bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.

3. Assuming a set of documents that need to be classified, use the naïve bayesian Classifier model to perform this task. Calculate the accuracy, precision, and recall for your data set.

4. Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. You can use Python ML library classes/API.

5. Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering. You can add Python ML library classes/API in the program.

6. Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions. Python ML library classes can be used for this problem.

7. Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select the appropriate data set for your experiment and draw graphs.

8. Write a program to implement 5-fold cross validation on a given dataset. Compare the accuracy,

precision, recall, and F-score for your data set for different folds.

9. Implement SVM/Softmax classifier for CIFAR-10 dataset: (i) using KNN, (ii) using 3 layer neural network.

10. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.

11. Image Captioning with Vanilla RNNs .

12. Image Captioning with LSTMs.

13. Familiarisation of cloud based computing like Google colab.

SEMESTER I

PROGRAM ELECTIVE I

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
R25CST113	OBJECT ORIENTED SOFTWARE ENGINEERING	PROGRAM ELECTIVE 1	3	0	0	3

Course Objectives:

1. Students will understand the principles of object-oriented software engineering and various software lifecycle models.
2. The course will enable learners to analyze, design, develop, and deploy software using appropriate tools and techniques.
3. Students will gain the skills to perform software testing, maintenance, and configuration management efficiently

Course Outcomes: After the completion of the course the student will be able to

Course Outcome		Bloom's Knowledge Level(KL)
CO 1	Make use of project organization and management concepts and analyse the various tasks carried out	K3
CO 2	Identify and select suitable process model for a given problem	K3
CO 3	Analyse the requirements of a given software project and produce requirement specification	K3
CO 4	Examine the various designing principles and patterns of a software product.	K3
CO 5	Build the mapping of product design to code, its testing and maintenance.	K3
CO 6	Design, analyse object models and dynamic models for a given problem statement	K6

Note: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO1	2		2	2	2	1	
CO2	2		2	3	2	2	
CO3	3		3	3	2	2	
CO4	2		2	3	3	2	
CO5	2		2	3	3	2	
CO6	3		3	3	3	2	2

(1-Weak, 2-Medium, 3- Strong)

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	70%-80%
Analyse	30%-40%
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Course based task: 15 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

SYLLABUS

MODULE	DESCRIPTION	CONTACT HOURS
1	**Classical Paradigm:** System Design Concepts – Project Organization Concepts : Project Organizations , Roles , Tasks and Work Products ,Schedule – Project Communication concepts : Planned Communication , Unplanned Communication ,Communication Mechanism – Project Management Concepts : Tasks and Activities ,Work Products , Work Packages and Roles , Work Breakdown Structure .	10
2	**Process Models:** Life cycle models: Sequential Activity Centered Models, Iterative Activity Centered models, Entity Centered models – Unified Process – Iterative and Incremental – Workflow – Agile Processes	8
3	**Analysis:** Requirements Elicitation Concepts – An Overview of Unified Modeling Language –Analysis Concepts : Analysis Object Model and Analysis Dynamic Models – Non-functional requirements – Analysis Patterns – Executable specification	7
4	**Design:** System Design, Architecture – Design Principles - Design Patterns – Dynamic Object Modeling Static Object Modeling – Model based approach vs Document based approach – Interface Specification – Object Constraint Language	8
5	**Implementation & Testing**: Mapping Design (Models) to Code – Testing - Usability – Deployment – Configuration Management – Maintenance	4

COURSE PLAN

No.	Topic	No. of Lectures
1	**Classical Paradigm(10)**	
1.1	System Design Concepts	1
1.2	Project Organization Concepts- Project Organizations	1
1.3	Roles, Tasks, Work Products and Schedule	1
1.4	Project Communication concepts	1
1.5	Planned Communication, Unplanned Communication	1
1.6	Communication Mechanism	1
1.7	Project Management Concepts	1
1.8	Tasks and Activities, Work Products	1
1.9	Work Packages and Roles	1
1.10	Work Breakdown Structure	1
2	**Process Models(8)**	
2.1	Life cycle models	1
2.2	Sequential Activity Centered Models	1
2.3	Iterative Activity Centered models	1
2.4	Entity Centered models	1
2.5	Unified Process	1
2.6	Iterative and Incremental	1
2.7	Workflow	1
2.8	Agile Processes	1

3	**Analysis(7)**	
3.1	Requirements Elicitation Concepts	1
3.2	An Overview of Unified Modeling Language	1
3.3	Analysis Concepts	1
3.4	Analysis Object Model and Analysis Dynamic Models	1
3.5	Non-functional requirements	1
3.6	Analysis Patterns	1
3.7	Executable specification	1
4	**Design(8)**	
4.1	System Design, Architecture	1
4.2	Design Principles	1
4.3	Design Patterns	1
4.4	Dynamic Object Modeling	1
4.5	Static Object Modeling	1
4.6	Model based approach vs Document based approach	1
4.7	Interface Specification	1
4.8	Object Constraint Language	1
5	**Implementation, Deployment And Maintenance(7)**	
5.1	Mapping Design (Models) to Code	1
5.2	Mapping Design (Models) to Code(Continued)	1
5.3	Testing	1
5.4	Usability	1
5.5	Deployment	1

5.6	Configuration Management	1
5.7	Maintenance	1

Reference Books

1. Bernd Bruegge, Alan H Dutoit, Object-Oriented Software Engineering, 2nd edition, Pearson Education, 2004.

2. Craig Larman, Applying UML and Patterns 3rd edition, Pearson Education, 2005

3. Stephen Schach, Software Engineering 7th ed, McGraw-Hill, 2007.

4. Ivar Jacobson, Grady Booch, James Rumbaugh, The Unified Software Development Process, Pearson Education, 1999.

5. Alistair Cockburn, Agile Software Development 2nd ed, Pearson Education.

Model Question Paper

Course Code: R25CST113

Course Name: OBJECT ORIENTED SOFTWARE ENGINEERING

Max. Marks : 60 Duration: 2.5 Hours

PART A

Answer All Questions. Each Question Carries 5 Marks

1. Being the member of the design team to develop an interface for an online registration portal, you are not sure about the mandatory fields. People in what role can help you out. Whether planned or unplanned communication will be more beneficial in this situation. Why? (5)
2. For what type of project, the spiral model suit's best. Why? (5)
3. Describe the different Requirement Elicitation techniques. (5)
4. Discuss the design Principles of System Design. (5)
5. Demonstrate the different steps of software deployment. (5)

Part B

(Answer any five questions. Each question carries 7 marks)

6. (a) Explain Work Breakdown Structure. (3)

   (b) What are the types of project organizations? (4)

7. (a) Explain Iterative Activity Centered Models. (3)

   (b) Explain Agile process? (4)

8. Problem statement: Student Attendance Monitoring System. Student Attendance Monitoring System is used to track the attendance of students in an Institute. Faculty

advisor can add students into the system, which is verified and approved by HOD. Once the students list is approved, the teachers can mark attendance on the system. Students can apply for duty leaves to the faculty advisor. Faculty advisor forwards the application to the HOD for approval. Students, teachers, faculty advisor and HOD can view the attendance reports of every student. Draw the sequence diagram for this problem statement. (7)

9. (a) Identify any four functional and non-functional requirements of KTU website. (4)

   (b) Compare the Dynamic Object Modelling with the Static Object Modelling (3)

10. (a) Design patterns speed up the development process quiet a lot. Illustrate with example. (3)

    (b) Discuss about the Object Constraint Language? (4)

11. (a) Discuss System Documentation (3)

    (b) What is skill matrix? Briefly describe the project management activities? (4)

12. (a) Consider a method that will return the fare of a transport bus, given the source, destination and number of passengers. The source and destination are specified as integers. 1 represents station A, 2 represents station B etc. The total number of seats is 30. Generate test cases for Unit testing the system. (4)

    (b) List out the benefits of model transformation. (3)

R25CST113	ADVANCED DATA MINING	CATEGORY	L	T	P	CREDIT
		PROGRAM ELECTIVE 1	3	0	0	3

Course Objective

1. Understand key concepts and techniques in data mining and knowledge discovery.
2. Apply data preprocessing, warehousing, and mining methods to real-world datasets.
3. Analyze, evaluate, and interpret mining results with focus on accuracy and impact.

Course Outcomes:

After the completion of the course the student will be able to:*

	Course Outcome	Bloom's Knowledge Level(KL)
CO 1	Summarise basic concepts of Data mining and Illustrate feature vector representation for a given data collection.	K1
CO 2	Design Data Warehouse for problems in various domains.	K3
CO 3	Implement Association Rules for analysing Transactional databases	K3
CO 4	Implement major Classification and Clustering Algorithms to a given problem.	K4
CO 5	To develop Data Mining system and analyze the performance	K5

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	2		3			2	
CO 2	2		3			2	
CO 3	2		3	2	2	2	
CO 4	2		3	2	2	2	
CO 5	2	2	3	2	2	2	2

(1-Weak, 2-Medium, 3- Strong)

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	48
Analyse	12
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions

relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

SYLLABUS

Module	Description	Contact Hours
1	**Data Mining and Knowledge Discovery** Desirable Properties of Discovered Knowledge – Knowledge representation, Data Mining Functionalities, Motivation and Importance of Data Mining, Classification of Data Mining Systems, Integration of a Data Mining System with a Database or Data Warehouse System, Classification, Clustering, Regression, Data Pre-processing: Data Cleaning, Data Integration and Transformation, normalization, standardization, Data Reduction, Feature vector representation. importance of feature engineering in machine learning; forward selection and backward selection for feature selection; curse of dimensionality; data imputation techniques; No Free Lunch theorem in the context of machine learning, Data Discretization and Concept Hierarchy Generation	8
2	**Data Warehouse and OLAP Technology for Data Mining** Data warehouses and its Characteristics - Data warehouse Architecture and its Components, Data Warehouse Design Process, Data Warehouse and DBMS, Data marts, Metadata, Data Cube and OLAP, Extraction - Transformation – Loading - Schemas for Multidimensional Database: Stars, Snowflakes and Fact constellations, OLAP Cube - OLAP Operations - OLAP Server Architecture - Data Warehouse Implementation - From Data Warehousing to Data Mining, Trends in data warehousing	9

3	**Association Pattern Mining** Mining Frequent Patterns, Associations and Correlations –Mining Methods – Mining Various Kinds of Association Rules – Correlation Analysis – Constraint Based Association Mining, Single Dimensional Boolean Association Rules From Transaction Databases, Multilevel Association Rules from transaction databases – Multi dimension Association Rules from Relational Database and Data Warehouses, Frequent Item Set Generation, Apriori Algorithm, Improved Apriori Algorithm for Association Rules Mining, Methods to improve Apriori, FP Growth Algorithm - Generating association rules from frequent itemset, Compact Representation of Frequent Item set - Maximal Frequent Item Set - Closed Frequent Item Sets. Pattern Evaluation Methods- Relationship Between FP-Growth and Enumeration-Tree Methods From Association Analysis to Correlation Analysis, Lift	**8**
4	**Classification and Prediction** Classification Techniques, Decision Tree - Decision tree Construction, Measures for Selecting the Best Split - Algorithm for Decision tree Induction - CART, Bayesian Belief Networks, Instance-Based Learning, K-Nearest neighbor classification, Accuracy and Error measures, Multiclass Classification, Semi-Supervised Classification, Multi class Learning, Rare class learning, Active Learning, Transfer Learning, Fuzzy Set Approaches for Classification, Rough Set Approaches, Techniques to improve classification accuracy-Ensemble methods, Bias-Variance Trade-off, Improving classification accuracy of class imbalanced data	**7**
5	**Cluster Analysis** Desired features of cluster Analysis, Types of data in cluster analysis, Categorization of Major Clustering Methods, Density-Based Methods, Clustering High Dimensional Data, Constraint Based Cluster Analysis, GA based clustering, Dealing with Large Databases, Probabilistic Model Based Clustering, Clustering with Constraints, Semi supervised clustering, Cluster Ensembles, Quality and validity of cluster analysis methods, Outlier Analysis-Statistical Approaches, Proximity Based Approaches Advanced Mining: Multimedia Data Mining - Text Mining, Graph Mining and Social Network Analytics - Geospatial Data Mining, Temporal Mining, Data Mining Applications - Social Impacts of Data Mining.	**8**

Course Plan

No	Topic	No. of Lectures (40 Hours)
1	**Module 1: Data Mining and Knowledge Discovery**	**6**
1.1	Data Mining Functionalities, Motivation and Importance of Data Mining	1
1.2	Integration of a Data Mining System with a Database or Data Warehouse System, Major Issues in Data Mining. Classification, Clustering, Regression	1
1.3	Data Pre-processing: Data Cleaning, Data Integration and Transformation, normalization	1
1.4	Data Reduction, Different techniques	1
1.5	Feature vector representation. importance of feature engineering in machine learning;	1
1.6	Forward selection and backward selection for feature selection;	1
2	**Module 2: Data Warehouse and OLAP Technology for Data Mining**	**7**
2.1	Data warehouses and its Characteristics - Data warehouse Architecture and its Components	1
2.2	Data Warehouse and DBMS, Data marts, Metadata Extraction - Transformation – Loading in DW,	1
2.3	Multidimensional model	1
2	**Module 2: Data Warehouse and OLAP Technology for Data Mining**	**7**
2.1	Data warehouses and its Characteristics - Data warehouse Architecture and its Components	1
2.2	Data Warehouse and DBMS, Data marts, Metadata Extraction - Transformation – Loading in DW,	1
2.3	Multidimensional model	1
2.4	Schemas for Multidimensional Database: Stars, Snowflakes Fact constellations	1
2.5	Design Data Warehouse for problems in different domains	1

2.6	OLAP Cube - OLAP Operations	1
2.7	OLAP Server Architecture - Data Warehouse Implementation	1
3	**Module 3: Association Rule Mining**	**7**
3.1	Mining Frequent Patterns, Associations and Correlations	1
3.2	Mining Various Kinds of Association Rules – Correlation Analysis – Constraint Based Association Mining	1
3.3	Multilevel Association Rules from transaction databases – Multi dimension Association Rules from Relational Database and Data Warehouses	1
3.4	Frequent Item Set Generation, Apriori Algorithm, Apriori Algorithm-illustration with example	1
3.5	Methods to improve Apriori, FP Growth Algorithm	1
3.6	FP Growth Algorithm- illustration with example, Compact Representation of Frequent Item set	1
3.7	Pattern Evaluation Methods, Association Analysis to Correlation Analysis, Lift	1
4	**Module 4: Classification and Prediction**	**10**
4.1	Classification Techniques, Decision Tree - Decision tree Construction Measures for Selecting the Best Split	1
4.2	Decision tree Induction - illustration with example Algorithm for Decision tree Induction - CART	1
4.3	Bayesian Belief Networks	1
4.4	Bayesian Belief Networks- Training	1
4.5	K-Nearest neighbor classification, Accuracy and Error measures	1
4.6	Multiclass Classification, Semi-Supervised Classification	1
4.7	Active Learning, Transfer Learning	1
4.8	Fuzzy Set Approaches for Classification	1
4.9	Rough Set Approaches	1
4.10	Ensemble methods. Improving classification accuracy of class imbalanced data	1
5	**Module 5: Cluster Analysis**	**10**
5.1	Desired features of cluster Analysis, Types of data in cluster analysis,	1
5.2	Categorization of Major Clustering Methods, Density-Based Methods,	1
5.3	Semi supervised clustering, Clustering High Dimensional Data,	1

	Constraint Based Cluster Analysis,	
5.4	GA based clustering	1
5.5	Probabilistic Model Based Clustering	1
5.6	Quality and validity of cluster analysis methods, Outlier Analysis-Statistical Approaches, Proximity Based Approaches	1
5.7	Multimedia Data Mining	1
5.8	Text Mining	1
5.9	Graph Mining and Social Network Analytics	1
5.10	Geospatial Data Mining, Temporal Mining	1

References

1. Kevin Murphy, Machine Learning: A Probabilistic Perspective (MLAPP), MIT Press, 2012

2. Christopher Bishop, Pattern Recognition and Machine Learning (PRML), Springer, 2007.

3. Jiawei Han, Micheline Kamber, Jian Pei, “Data Mining: Concepts and Techniques”, Morgan Kaufmann, 2nd Ed., 2005

4. Charu C. Aggarwal ,Data Mining, Springer, ISBN 978-3-319-14141-1,2015

5. Data Mining Techniques,Arun K Puari, Universities Press,2001

6. Margaret H. Dunham, “Data Mining: Introductory and Advanced Topics”, Prentice Hall, 1st Ed., 2002.

7. David G. Stork, Peter E. Hart, and Richard O. Duda. Pattern Classification (PC), Wiley-Blackwell, 2000

8. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning (ESL), Springer, 2009

9. G. K. Gupta “Introduction to Data Mining with Case Studies”, Eastern Economy Edition, Prentice Hall of India, 2006.

10. Soumen Chakrabarti, "Mining the Web: Discovering Knowledge from Hypertext Data", Morghan Kaufmann, 1st Ed., 2005.

11. Da Ruan, Guoqing Chen, Etienne E. Kerre, Geert Wets, “Intelligent Data Mining: Techniques and Applications (Studies in Computational Intelligence)”, Springer, 1st Ed., 2010.

12. Masoud Mohammadian, “Intelligent Agents for Data Mining and Information Retrieval”, Idea Group Publishing, 2004.

13. I. H. Witten and E. Frank. Data Mining: Practical Machine Learning Tools and Techniques Morgan Kaufmann, 2000.

14. D. Hand, H. Mannila and P. Smyth. Principles of Data Mining. Prentice-Hall. 2001

15. Valliappa Lakshmanan, Jordan Tigani, Google BigQuery: The Definitive Guide**:** Data Warehousing, Analytics, and Machine Learning, O'Reilly Media, Inc.", 2019

MODEL QUESTION PAPER

Course Code: R25CST113

ADVANCED DATA MINING

Max. Marks: 60 Duration: 2.5 hours

PART A

(Answer *ALL* questions; each question carries 5 marks)

1. Differentiate between classification and regression with example
2. Explain concept hierarchy generation. With a suitable example show how is it done for categorical data.
3. How can you generate association rules from frequent item sets?
4. Why are nearest neighbor algorithms called lazy learners? What are the disadvantages of a lazy learner?
5. How do we relate text mining and web mining? Differentiate between spatial and non spatial data with example

PART B

(Answer *any FIVE* questions; each question carries 7 marks)

6(a). Why feature engineering is important? What is the output of feature engineering in machine learning?

(b) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (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. i) Use smoothing by bin means to smooth the above data, using a bin depth of 3. Illustrate your steps. ii) How might you determine outliers in the data.

7(a)How do data warehousing relate to data mining? Discuss

(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.

a) List three classes of schemas that are popularly used for modeling data warehouses.

b) Draw a schema diagram for the above data warehouse using one of the schema classes listed in(a).

Starting with the base cuboid [day,doctor,patient],what a specific OLAP operations should be performed in order to list the total fee collected by each doctor in 2022?

8(a) Why is the FP growth algorithm so efficient?

(b) Discuss FP growth algorithm. Using Apriori and FP growth algorithm find the frequent itemsets from the following transactional database? (min_sup= 2, confidence 70%). Compare the two processes

TID	List of item-IDs
T100	I1,I2,I3
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

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
R25CST113	CLOUD COMPUTING	PROGRAM ELECTIVE 1	3	0	0	3

Course Objectives:

1. Students will understand the core concepts of virtualization and its application in cloud computing, including server, network, and storage virtualization.
2. The course will equip students with the practical skills to deploy and manage both virtualization solutions and private cloud environments.
3. Students will be able to identify and address security challenges inherent in grid and cloud computing to ensure the protection of private cloud infrastructure.

Course Outcomes: After the completion of the course the student will be able to

Course Outcome		Bloom's Knowledge Level(KL)
CO 1	Employ the concepts of storage virtualization, network virtualization and its management	K3
CO 2	Apply the concept of virtualization in the cloud computing.	K3
CO 3	Apply domain knowledge in architecture, infrastructure and delivery models of cloud computing in designing and developing cloud applications.	K3
CO 4	Develop services using Cloud computing.	K3
CO 5	Analyse and choose security models appropriate to the cloud environment	K3
CO 6	Design, develop and implement cloud based applications	K6

Note: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO1	2		2	2	3	1	
CO2	2		2	3	2	1	
CO3	3		3	3	3	2	2
CO4	2		3	3	2	2	2
CO5	2		2	2	3	2	
CO6	3	2	3	3	3	2	2

(1-Weak, 2-Medium, 3- Strong)

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	70%-80%
Analyse	30%-40%
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Course based task: 15 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 pu blications shall be referred) : 15 marks

ii.Course based

task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

SYLLABUS

MODULE	DESCRIPTION	CONTACT HOURS
1	Virtualization - Basics of Virtual Machines, Process & System VMs, Emulation, Interpretation, Binary Translation, Taxonomy of Virtual Machines, Virtualization Management, Hardware Maximization, Architectures, Storage Virtualization, Network Virtualization	8
2	Virtualization Infrastructure - Comprehensive Analysis, Resource Pool, Testing Environment, Server Virtualization, Virtual Workloads, Provision of VMs, Desktop & Application Virtualization, CPU/Memory/I/O Virtualization, Virtual Clusters, Resource Management, Data Centre Automation	8
3	Cloud Platform Architecture - Cloud computing basics, History, Advantages & Disadvantages, Deployment Models (Public/Private/Hybrid), Everything as a Service (IaaS/PaaS/SaaS), Generic Cloud Architecture, Virtualization Support, Disaster Recovery, Public Cloud Platforms (GAE, AWS), Inter-cloud Resource Management	9
4	Programming Mode - Hadoop Framework, MapReduce, Input Splitting, Functions, Configuring & Running Jobs, Developing MapReduce Apps, Hadoop File System, Cluster Setup, Cloud Software Environments (Eucalyptus, Open Nebula, OpenStack, Nimbus)	8
5	Cloud Security - Infrastructure Security, Network/Host/Application Security, Data Security, Provider Data Security, Identity & Access Management Architecture, IAM Practices, SaaS/PaaS/IaaS Availability, Privacy Issues, Cloud Security & Trust Management	7

COURSE PLAN

No	Description	No. of Lectures
1	**Virtualization**	
1.1	Basics of Virtual Machines	1
1.2	Process Virtual Machines, System Virtual Machines	1
1.3	Emulation, Interpretation	1
1.4	Binary Translation	1
1.5	Taxonomy of Virtual Machines	1
1.6	Virtualization Management, Hardware Maximization	1
1.7	Architectures, Virtualization Management	1
1.8	Storage Virtualization, Network Virtualization	1
2	**Virtualization Infrastructure**	
2.1	Comprehensive Analysis, Resource Pool	1
2.2	Testing Environment, Server Virtualization	1
2.3	Virtual Workloads	1
2.4	Provision, Virtual Machines	1

2.5	Desktop Virtualization, Application Virtualization	1
2.6	Implementation levels of virtualization, virtualization structure, virtualization of CPU	1
2.7	Memory and I/O devices	1
2.8	Virtual clusters and Resource Management, Data centre automation	1
3	**Cloud Platform Architecture**	
3.1	Understanding Cloud computing, History, Advantages and Disadvantages	1
3.2	Cloud deployment models (Public/Private/Hybrid), Categories of cloud computing	1
3.3	Everything as a Service, Infrastructure	1
3.4	Platform, Software	1
3.5	Generic Cloud Architecture Design, Layered cloud Architectural Development	1
3.6	Virtualization Support and Disaster Recovery, Architectural Design Challenges	1
3.7	Public Cloud Platforms	1
3.8	GAE, AWS	1
3.9	Inter-cloud Resource Management	1
4	**Programming Mode**	

4.1	Introduction to Hadoop Framework, MapReduce	1
4.2	Input splitting	1
4.3	Map and reduce functions, Specifying input/output parameters	1
4.4	Configuring and running a job, Developing MapReduce Applications	1
4.5	Design of Hadoop file system, Setting up Hadoop Cluster	1
4.6	Cloud Software Environments: Eucalyptus	1
4.7	Open Nebula, Open Stack	1
4.8	Nimbus	1
5	**Cloud Security**	
5.1	Cloud Infrastructure security	1
5.2	Network, host and application level	1
5.3	Aspects of data security, Provider data security	1
5.4	Identity and access management architecture	1
5.5	IAM practices in the cloud	1
5.6	SaaS, PaaS, IaaS availability in the cloud	1
5.7	Key privacy issues, Cloud Security and Trust Management	1

Reference Books:

1. Greg Schulz, “Cloud and Virtual Data Storage Networking”, Auerbach Publications [ISBN: 978-1439851739], 2011.

2. Michael Miller, Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online, Que Publishing, August 2008.

3. GauthamShroff, “Enterprise Cloud Computing: Technology, Architecture, Applications”, Cambridge press, 2010.

4. EMC, “Information Storage and Management” Wiley; 2 edition [ISBN: 978-0470294215],2012.

5. Kai Hwang , Geoffrey C Fox, Jack J Dongarra : “Distributed and Cloud Computing – From Parallel Processing to the Internet of Things” , Morgan Kaufmann Publishers – 2012.

Model Question Paper

Course Code: R25CST113

Course Name: CLOUD COMPUTING

Max. Marks : 60 Duration: 2.5 Hours

PART A

Answer All Questions. Each Question Carries 5 Marks

1. Sketch the core the differences between a traditional computer and a virtual machine. (5)
2. Explain your understanding about virtualization. What is the role of VMM in virtualization? (5)
3. Illustrate PaaS model for cloud computing. (5)
4. Summarize the concept of Map Reduce? Explain the logical data flow of Map Reduce function using suitable example. (5)
5. Illustrate the major security challenges in clouds? (5)

Part B

(Answer any five questions. Each question carries 7 marks)

6. How Memory virtualization is implemented? Provide necessary examples and diagrams wherever necessary
   (7)
7. Investigate the functional modules of Google App Engine?
   (7)
8. Sketch the core idea about virtualization. What is the role of VMM in virtualization?(7)
9. Is it harder to establish security in the cloud? Justify
   (7)
10. With a neat diagram explain the Generic Cloud architecture and components.
    (7)
11. With neat diagram explain your understanding Security Architecture Design in cloud.
    (7)
12. Demonstrate Private Cloud Design using Open Nebula.
    (7)

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
R25CST113	WEB SERVICES	PROGRAM ELECTIVE 1	3	0	0	3

Course Objective

1. To understand the details of web services technologies: SOAP, WSDL, UDDI.
2. To learn how to implement and deploy web services clients and servers.
3. To explore interoperability between different frameworks.
4. To learn basic concepts of SOAP.

Course Outcomes : After the completion of the course the student will be able to:*

Course Outcome		Bloom's Knowledge Level(KL
CO 1	Illustrate the need of web services in developing applications.	K2
CO 2	Make use of Server-side and Client-side RESTful web services.	K4
CO 3	Analyze how web services can be published in standalone web servers.	K4
CO 4	Employ techniques on creating dynamic web pages	K3
CO 5	Utilize emerging technologies in web services.	K3
CO6	Design, Develop, Implement and Present innovative ideas on modern web services concepts and techniques.	K5

Note: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	2		1	1		2	
CO 2	1		1	2		2	
CO 3	2		2	2		3	1
CO 4	1		2	2	2	1	1
CO 5	2		1	2	3	1	1
CO6	2	1	2	2	2	1	2

(1-Weak,2-Medium, 3-Strong)

Assessment Pattern

Bloom's Category	End Semester Examination
Understand	15
Apply	20
Analyse	15
Create	10

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

ContinuousInternalEvaluation: 40 marks

Preparing a review article based on peer reviewed original publications (minimum 10publications shall be referred): 15 marks

Coursebasedtask/Seminar/Datacollectionandinterpretation:15marks Test paper, 1 no.: 10 marks

Test paper shall include minimum 80% of the syllabus.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College. There will be two parts; Part A and Part B. Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question (such questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students).

Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to the theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

Total duration of the examination will be 150 minutes

SYLLABUS

Module	Description	Contact Hours
1	Web Services Standards Organizations, Service oriented architecture, Advantages of web services over distributed object architecture, SOAP-based web services, RESTful web services. Review of HTTP requests and responses- HTTP as an API, A RESTful example, Use of Servlets for RESTful Web Services.	**8**

2	A RESTful service as an HttpServlet, Implementation details. A RESTful Web Service as a JAX-RS Resource- JAX-RS Web Service Using Jersey, Publishing JAX-RS Resources with a Java Application, Publishing JAX-RS Resources with Tomcat, JAX-RS Generation of XML and JSON Responses, Porting the Predictions Web Service to JAX-RS. A RESTful Web Service as Restlet Resources. GraphQL- Introduction to GraphQL, GraphQL Architecture, Basic Queries	**8**
3	A Perl Client Against a Java RESTful Web Service. RESTful Clients and WADL Documents- The JAX-RS Client API, JSON for JavaScript Clients- JSONP and Web Services.	**8**
4	Introduction and Evolution of SOAP, Architecture of a typical SOAP-based service, Publishing a SOAP-Based Service with a Standalone Web Server. RandService- JavaClient Against the RandService, C# Client Against the RandService, A Perl Client Against the RandService. WSDL – WSDL document structure.	**8**
5	Introduction to React and Node-Basic Concepts and Applications, Rendering Elements and Components, Comparison and Purpose of Node.js and React.js, Angular JS Basics-Modules, Creating Components, Directives, Filters, Angular Forms-Services, Single page application and Multipage application, Use case of a real-time single page chat application.	**8**

COURSE PLAN

No	Topic	No. of Lectures (40 Hours)
	Module-1(Introduction to Web Services) 8 hours	
.1	Introduction to web services-Web Services Standards Organizations, Service oriented architecture	1
.2	Advantages of web services over distributed object architecture	1
.3	SOAP-based web services	1
.4	RESTful webservices	1
.5	Review of HTTP requests and responses	1
.6	HTTP as an API	1
.7	A RESTful example	1
.8	Use of Servlets for RESTful Web Services	1
	Module-2(RESTful Web Services:TheService Side) 8 hours	
2.1	A RESTful service as an HttpServlet, Implementation details	1
2.2	A RESTfulWeb Serviceas aJAX-RS Resource	1
2.3	JAX-RS Web Service Using Jersey	1
2.4	Publishing JAX-RS Resources with a Java Application	1

2.5	Publishing JAX-RS Resources with Tomcat	1
2.6	JAX-RS Generation of XMLand JSON Responses	1
2.7	Porting the Predictions Web Service to JAX-RS	1
2.8	GraphQL- Introduction to GraphQL,GraphQL Architecture,Basic Queries	1
3	**Module-3(RESTfulWebServices:TheClientSide)8 hours**	
3.1	A Perl Client Against a Java RESTful WebService(Lecture1)	1
3.2	A perl Client Against a JavaRESTful WebService(Lecture2)	1
3.3	RESTful Clients and WADL Documents(Lecture1)	1
3.4	RESTful Clients and WADL Documents(Lecture2)	1
3.5	The JAX-RS Client API	1
3.6	JSON for JavaScript Clients	1
3.7	JSONP and WebServices(Lecture1)	1
3.8	JSONP and WebServices (Lecture2)	1
4	**Module-4(SOAP-BasedWebServices) 8 hours**	
4.1	Introduction and Evolution of SOAP	1
4.2	Architecture of a typical SOAP-based service	1
4.3	Publishing a SOAP-Based Service with a Standalone Web Server	1

	(Lecture1)	
4.4	Publishing a SOAP-Based Service with a StandaloneWeb Server (Lecture2)	1
4.5	RandService-Java Client Against the RandService	1
4.6	C# Client Against the RandService	1
4.7	A perl Client Against the RandService	1
4.8	WSDL–WSDLdocument structure	1
5	**Module-5(IntroductiontoReact.js,Node.jsand Angular.js) 8 hours**	
5.1	Introduction to React and Node, Basic Concepts and Applications	1
5.2	Rendering Elements and Components	1
5.3	Comparison and Purpose of Node.js and React.js	1
5.4	Angular JS Basics-Modules	1
5.5	CreatingComponents,Directives,Filters	1
5.6	AngularForms-Services	1
5.7	Singlepage application and Multipage application	1
5.8	Usecase of a real-time single page chat application.	1

ReferenceBooks

1. Martin Kalin, Java Web Services: Up and Running, Second Edition, O'Reilly, 2013.

2. Robin Wieruch, The Road to Learn React, https://www.roadtoreact.com, 2022 Edition.

3. Andrew Grantt, Beginning Angular JS, Apress, 2014.

4. Vipul A M and Prathamesh Sonpatki, ReactJS by Example - Building Modern Web Applications with React, PACKT Publishing, 2016.

5. Node.js Web Development: Server-side web development made easy with Node 14 using practical examples, 5th Edition, PACKT Publishing, 2020.

6. Alex Banks, Eve Porcello, Learning GraphQL, O'Reilly Media, 2018.

MODEL QUESTION PAPER

CourseCode: R25CST113

WEB SERVICES

Max.Marks:60 Duration:2.5hours

PARTA

(Answer *ALL* questions; each question carries 5marks)

1. Examine the possible reasons for the choice of Web services over Web applications.

Why are HTTP servlets a convenient way for implementing RESTful web services?

2. Why are HTTP servlets a convenient way for implementing RESTful web.
3. Explain how a basic Client request can be created using the Client API.
4. Describe the architecture of a SOAP based Web Service.
5. Discuss the fact how react is different from angular.

PARTB

(Answer *any FIVE* questions; each question carries7 marks)

6. How are SOAP and RESTful Web services architecturally different from each other.
7. Illustrate the process of publishing JAX-RS resources with Tomcat.
8. JSONP brings an event-driven API to client-side processing- Justify the statement using examples.
9. Explain how a documents client service REST Client can be created from DOCS DWADL.
10. Describe the structure of a WSDL document with example
11. Analyze the purpose of render in React. How would you prevent unnecessary component re-render in ReactJS?
12. Demonstrate with examples the two approaches that AngularJS takes to build forms.

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
R25CST113	COMPUTATIONAL INTELLIGENCE	PROGRAM ELECTIVE 1	3	0	0	3

Course Objectives:

1. Students will understand the core concepts and principles of computational intelligence.
2. The course will enable learners to apply fuzzy logic, genetic algorithms, and swarm optimization to solve complex problems.
3. Students will be able to design and implement effective computational intelligence solutions for real-world applications.

Course Outcomes: After the completion of the course the student will be able to

Course Outcome		Bloom's Knowledge Level(KL)
CO 1	Apply fuzzy logic to handle uncertainty and solve engineering problems	K3
CO 2	Apply Fuzzy Logic Inference methods in building intelligent machines.	K3
CO 3	Design genetic algorithms for optimized solutions in engineering problems	K4
CO 4	Analyze the problem scenarios and apply Ant colony system to solve real optimization problems.	K4
CO 5	Apply PSO algorithm to solve real world problems.	K3
CO 6	Design, develop and implement solutions based on computational intelligence concepts and techniques.	K6

Note: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create

Mapping of course outcomes with program outcomes

CO / PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7
CO1	2		2	3	3	1	
CO2	2		2	3	2	1	
CO3	3		3	3	3	2	2
CO4	2		2	3	3	2	2
CO5	2		2	3	3	2	
CO6	3	2	3	3	3	2	2

(1-Weak, 2-Medium, 3- Strong)

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	70%-80%
Analyse	30%-40%
Evaluate	
Create	

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Course based task: 15 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

SYLLABUS

MODULE	DESCRIPTION	CONTACT HOURS
1	Crisp sets vs fuzzy sets- Operations and properties of Fuzzy sets. Membership functions - Linguistic variables. Operations on fuzzy sets- Fuzzy laws- Operations on fuzzy relations, Fuzzy composition- Max- min, Max – product. Alpha-cut representation.	9
2	Fuzzy Reasoning – GMP and GMT. Fuzzy Inference System: Defuzzification methods - Fuzzy Controllers - Mamdani FIS, Larsen Model	7
3	Introduction to Genetic Algorithms – Theoretical foundation - GA encoding, decoding - GA operations – Elitism – GA parameters – Convergence. Multi-objective Genetic Algorithm – Pareto Ranking.	7
4	Swarm intelligent systems - Background Ant colony systems – Biological systems- Development of the ant colony system- - Working - Pheromone updating- Types of ant systems- ACO algorithms for TSP	8
5	Basic Model - Global Best PSO- Local Best PSO- Comparison of 'gbest' to 'lbest'- PSO Algorithm Parameters- Problem Formulation of PSO algorithm- Working. Rate of convergence improvements -Velocity clamping- Inertia weight- Constriction Coefficient- Boundary Conditions- Guaranteed Convergence PSO- Initialization, Stopping Criteria, Iteration Terms and Function Evaluation.	9

COURSE PLAN

No.	Description	No. of Lecture Hours
1	Fuzzy Logic(9)	
1.1	Crisp sets vs fuzzy sets, Operations and properties of Fuzzy sets	1
1.2	Membership functions	1
1.3	Linguistic Variables	1
1.4	Operations on fuzzy sets	1
1.5	Fuzzy laws	1
1.6	Operations on fuzzy relations	1
1.7	Fuzzy Composition – Max-min	1
1.8	Fuzzy Composition – Max-Product	1
1.9	Alpha-cut representation	1
2	Fuzzy Systems(7)	
2.1	Fuzzy Reasoning – GMP	1
2.2	Fuzzy Reasoning – GMT	1
2.3	Fuzzy Inference System	1
2.4	Defuzzification methods	1

2.5	Fuzzy Controllers	1
2.6	Mamdani Model	1
2.7	Larsen Model	1
3	Genetic Algorithms(7)	
3.1	Introduction to Genetic algorithm	1
3.2	Theoretical foundation	1
3.3	GA encoding – decoding	1
3.4	GA operations	1
3.5	Elitism, GA parameters, Convergence of GA	1
3.6	Multi–objective Genetic Algorithm	1
3.7	Pareto Ranking	1
4	Ant Colony Systems(8)	
4.1	Swarm intelligent systems	1
4.2	Background	1
4.3	Ant colony systems – biological systems	1
4.4	Development of the ant colony system	1
4.5	Working	1

4.6	Pheromone updating	1
4.7	Types of ant systems	1
4.8	ACO algorithms for TSP	1
5	Particle Swarm Optimization(9)	
5.1	Basic Model	1
5.2	Global Best PSO	1
5.3	Local Best PSO, Comparison of 'gbest' to 'lbest'	1
5.4	PSO Algorithm Parameters	1
5.5	Problem Formulation	1
5.6	Working	1
5.7	Rate of convergence improvements – velocity clamping	1
5.8	Inertia-weight - Constriction Coefficient - Boundary Conditions	1
5.9	Initialization, Stopping Criteria, Iteration Terms and Function Evaluation	1

Reference Books:

1. Samir Roy, Udit Chakraborty, Introduction to Soft Computing Neuro- Fuzzy Genetic Algorithms, Pearson, 2013
2. N.P. Padhy, Artificial Intelligence and Intelligent systems, Oxford Press, New Delhi, 2005.
3. Xin-She Yang School of Science and Technology, Middlesex University London, Nature-Inspired Optimization Algorithms, Elsevier, First edition, 2014
4. Satyobroto Talukder, Blekinge Institute of Technology, Mathematical Modelling and Applications of Particle Swarm Optimization, February 2011
5. Mitchell Melanie, An Introduction to Genetic Algorithm, Prentice Hall,1998

6. Andries Engelbrecht, Computational Intelligence: An Introduction, Wiley, 2007
7. Marco Dorigo and Thomas Stutzle, “Ant Colony optimization”, Prentice Hall of India, New Delhi 2005

Model Question Paper

FIRST SEMESTER M.TECH DEGREE EXAMINATION, MONTH & YEAR

Course Code: R25CST113

Course Name: COMPUTATIONAL INTELLIGENCE

Max. Marks : 60 Duration: 2.5 Hours

PART A

Answer All Questions. Each Question Carries 5 Marks

1. Consider the set of Colours A= {Blue, Red, Orange, Yellow, Green}, Attributes B = {Bright, Warmth, Dullness}, Feelings C= {Unpleasant, happiness, Angry}. Given R and S where R is the relationship between colours and their attributes and S is the relationship between colour attributes and feelings created. Find the relationship Q between colours and feelings created (5).

	Bright	Warmth	Dullness
Blue	0.8	0.6	0.4
Red	0.8	0.8	0.2
Orange	0.5	0.7	0.2
Yellow	0.3	0.6	0.4
Green	0.8	0.6	0.4

	Unpleasant	Happiness	Angry
Bright	0.2	0.8	0.6
Warmth	0.4	0.7	0.8

Dullness	0.8	0.3	0.6

2. Develop a membership function for "Tall". Based on that devise membership function for "Very Tall". Explain how it is done (5).

3. Mention the importance of objective (fitness) function in genetic algorithm (5).

4. Describe how pheromone is updated. What is elitist / elastic ants? Are they useful in this scenario? (5).

5. What is the significance of pbest and gbest particles in solving problems with particle swarm optimization? (5).

PART B

(Answer any five questions. Each question carries 7 marks)

6. (a) Consider the set of fruits F = {Apple, Orange, Lemon, Strawberry, Pineapple}.

$$\text{Let sweet fruits } B = \{\frac{0.8}{Apple} + \frac{0.6}{Orange} + \frac{0.2}{Lemon} + \frac{0.4}{Strawberry} + \frac{0.7}{Pineapple}\}$$
$$\text{Sour Fruits } F = \{\frac{0.6}{Apple} + \frac{0.8}{Orange} + \frac{0.9}{Lemon} + \frac{0.7}{Strawberry} + \frac{0.5}{Pineapple}\}$$

Find Fruits that are Sweet or Sour, Sweet but not Sour, Sweet and Sour (3).

6. (b) Consider two fuzzy Sets given by

$$P = \{\frac{0.9}{short} + \frac{0.3}{medium} + \frac{0.5}{tall}\}$$

$$Q = \{\frac{0.7}{positive} + \frac{0.4}{zero} + \frac{0.8}{negative}\}$$

Find the fuzzy relation for the Cartesian product of P and Q i.e. R=P×Q. Introduce a fuzzy set T given by

$$T = \{\frac{0.9}{short} + \frac{0.3}{medium} + \frac{0.6}{tall}\}$$

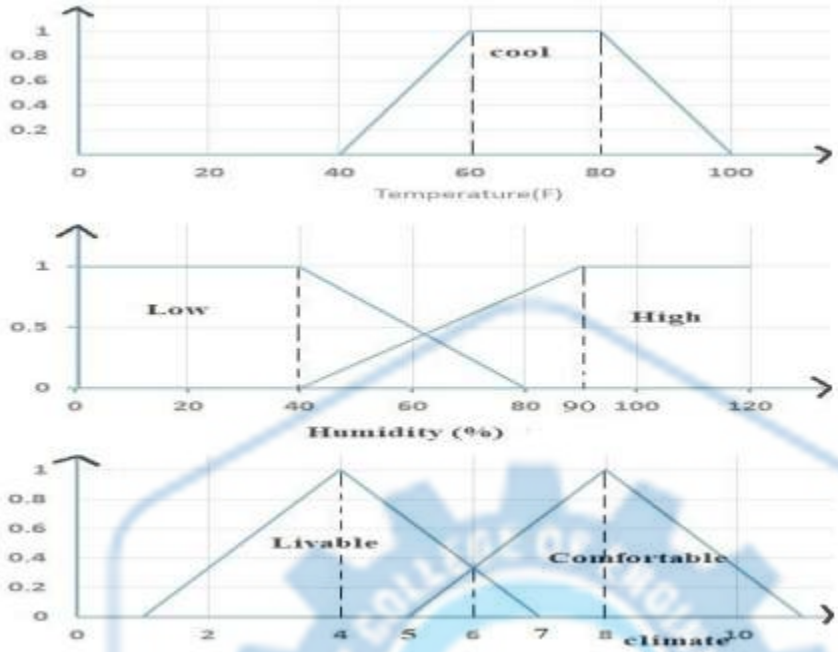
and Find T o R using max-min composition (4).

7. Consider a Fuzzy Inference System for checking climate comfortability of human beings for long time living. The system accepts two inputs temperature and humidity. The rules and membership functions of FIS is given below. Using Mamdani inference and center of sum, calculate output when the temperature is 50°Fahrenheit and humidity is 50% (7).

Rules:

- **Rule 1:** IF temperature is cool and humidity is low, THEN climate is comfortable.

- **Rule 2:** IF temperature is cool and humidity is high, THEN climate is livable.



The fuzzy sets "Easy Question Paper" and their corresponding "Student Performance" are given below:

$$\text{Easy\_QP} = \left\{\frac{0.8}{1} + \frac{0.2}{2} + \frac{0.6}{3} + \frac{0.7}{4}\right\}$$

$$\text{Stud\_Perf} = \left\{\frac{0.3}{a} + \frac{0.4}{b} + \frac{0.8}{c} + \frac{0.9}{d} \frac{0.8}{1} + \frac{0.2}{2} + \frac{0.6}{3} + \frac{0.8}{4} + \frac{0.7}{5}\right\}$$

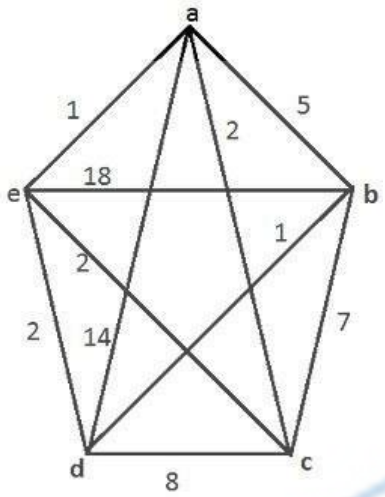
Find the performance of students c and d for the question paper "Somewhat Easy".

Somewhat Easy

$$\left\{\frac{0.7}{1} + \frac{0.3}{2} + \frac{0.5}{3} + \frac{0.6}{4}\right\}$$

1. Explain any procedure to map a solution to the corresponding chromosome and vice versa in genetic algorithms. Also illustrate it with an example (7).
2. Describe two methods used to select individuals from a population for the mating pool in Genetic Algorithms (7).

10. (a) Consider the TSP with the following edge costs. Given the evaporation factor ρ=0.02 and initial pheromone at all edges Tij=100. What is the cost of best tour? (1).



10. (b) Using the equation Tij(t+1)=(1−ρ)Tij(t)+△Tij(t,t+1), compute the Tij the edge <a,c> when 10 ants uses the edges <a, c>, using the following models:

- i. Ant Density Model (Constant Q=10)
- ii. Ant Quantity Model(Constant Q=100)
- where Q is the constant related to the pheromone updation (6).

11. Describe Ant Colony System. What are the different types of Ant systems? (7).

12. Consider a particle swarm optimization system composed of three particles and maximum velocity 10. Assume that both the random numbers r1 and r2 used for computing the movement of the particle towards the individual best position and social best position are 0.5. Also assume that the space of solutions is the two-dimensional real valued space and the current state of swarm is as follows (7):

- Position of particles: x1=(4,4), x2=(8,3), x3=(6,7)
- Individual best positions: x1∗=(4,4), x2∗=(7,3), x3∗=(5,6)
- Velocities: v1=(2,2); v2=(3,3); v3=(4,4)

What would be the next position of each particle after one iteration of the PSO algorithm if the inertia parameter that is used along with current velocity update formula is 0.8?

R25CST113	AUTOMATED VERIFICATION	CATEGORY	L	T	P	CREDIT
		PROGRAM ELECTIVE 1	3	0	0	3

Course Objective

1. To develop a clear understanding of formal reasoning methods, propositional logic, and their role in decision problems, with a focus on normal forms, SAT solvers, and Boolean structures.
2. To analyze equality logic and uninterpreted functions, including decision procedures such as congruence closure, graph-based reductions, and comparisons of different reduction methods.
3. To apply linear arithmetic techniques—such as the Simplex algorithm, Fourier-Motzkin elimination, and the Omega Test—for solving real-world decision problems with constraints.

Course Outcomes:

After the completion of the course the student will be able to:*

Course Outcome		Bloom's Knowledge Level(KL)
CO 1	Use the model-theoretic and proof-theoretic approaches towards formal reasoning.	K1
CO 2	Demonstrate how decision procedures can be developed for propositional logic using SAT solvers and Binary Decision Diagrams.	K3
CO 3	Develop methods to prove the validity and satisfiability of formulas using Equality Logic and Uninterpreted Functions.	K3
CO 4	Illustrate decision procedures using linear arithmetic	K4
CO 5	Design, develop and implement solutions based on the concepts of automated verification.	K5

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	2		3	3		2	

CO 2	2		3	2		2	
CO 3	2		3	2		2	
CO 4	2		3	2		2	
CO 5	2	2	3	2	2	2	2

(1-Weak, 2-Medium, 3- Strong)

Assessment Pattern

Bloom's Category	**End Semester Examination**
Apply	48
Analyse	12
Evaluate	
Create	

Mark distribution

Total Marks	**CIE**	**ESE**	**ESE Duration**
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall

achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Syllabus

Module	Description	Contact Hours
1	**Introduction and Basic Concepts** Two approaches to Formal Reasoning, Basic Definitions, Normal forms and their properties, The theoretical point of view, Expressiveness vs. Decidability, Boolean structure in Decision Problems.	6
2	**Decision Procedures for Propositional Logic** Propositional Logic, SAT Solvers, Binary Decision Diagrams	8
3	**Equality Logic and Uninterpreted Functions** Introduction, Uninterpreted Functions, From Uninterpreted Functions to Equality Logic, Functional Consistency is not enough, Two examples of the use of Uninterpreted Functions.	7
4	**Decision Procedures for Equality Logic and Uninterpreted Functions** Congruence Closure, Basic Concepts, Simplification of the formula, A Graph-Based Reduction to Propositional Logic, Equalities and Small Domain Instantiations, Ackermann's vs. Bryant's Reduction.	8
5	**Linear Arithmetic**	11

	Introduction, The Simplex Algorithm, The Branch and Bound Method, Fourier-Motzkin Variable Elimination, The Omega Test, Preprocessing, Difference Logic.	

Course Plan

No	Topic	No. of Lectures (40 hrs)
1	**Module 1: Introduction and Basic Concepts**	**6**
1.1	Two approaches to formal reasoning – Proof by deduction, Proof by enumeration, Deduction and enumeration	1
1.2	Basic definitions	1
1.3	Normal forms and their properties (Lecture 1)	1
1.4	Normal forms and their properties (Lecture 1)	1
1.5	The theoretical point of view	1
1.6	Expressiveness vs. Decidability, Boolean structure in decision problems	1
2	**Module 2: Decision Procedures for Propositional Logic**	**8**
2.1	Propositional Logic – Introduction	1
2.2	SAT solvers – Introduction, The *Davis-Putnam-Loveland-Logemann (DPLL)* framework	1
2.3	Boolean Constraints Propagation (BCP) and Implication Graph	1
2.4	Conflict Clauses and Resolution	1
2.5	Decision Heuristics	1
2.6	The Resolution Graph and the Unsatisfiable Core	1
2.7	Binary Decision Diagrams (Lecture 1)	1
2.8	Binary Decision Diagrams (Lecture 2)	1
3	**Module 3: Equality Logic and Uninterpreted Functions**	**7**
3.1	Introduction – Complexity and Expressiveness, Boolean Variables, Removing the Constraints	1
3.2	Uninterpreted Functions – How they are used	1
3.3	Proving equivalence of programs	1
3.4	From Uninterpreted Functions to Equality Logic – Ackermann's Reduction	1

3.5	From Uninterpreted Functions to Equality Logic – Bryant's Reduction	1
3.6	Functional Consistency is not enough	1
3.7	Two examples of the use of Uninterpreted Functions – Proving equivalence of circuits, Verifying a compilation process with Translation Validation	1
4	**Module 4: Decision Procedures for Equality Logic and Uninterpreted Functions**	**8**
4.1	Deciding a conjunction of Equalities and Uninterpreted Functions with Congruence Closure	1
4.2	Basic Concepts	1
4.3	Simplifications of the formula	1
4.4	A Graph-Based Reduction to Propositional Logic	1
4.5	Equalities and Small-Domain Instantiations – Some Simple Bounds, Graph-Based Domain Allocation	1
4.6	The Domain Allocation Algorithm	1
4.7	A Proof of Soundness	1
4.8	Ackermann's vs. Bryant's Reduction: Where does it matter?	1
5	**Module 5: Linear Arithmetic**	**11**
5.1	Basic Definitions, Solvers for Linear Arithmetic	1
5.2	The Simplex Algorithm – The basics	1
5.3	Simplex with Upper and Lower Bounds, Incremental Problems	1
5.4	The Branch and Bound Method	1
5.5	Cutting-Planes	1
5.6	Fourier-Motzkin Variable Elimination	1
5.7	The Omega Test (Lecture 1)	1
5.8	The Omega Test (Lecture 2)	1
5.9	The Omega Test (Lecture 3)	1
5.10	Preprocessing	1
5.11	Difference Logic	1

References

1. Daniel Kroening andOferStrichman. Decision Procedures – An Algorithmic Point of View, Springer, 2008.
2. ChristelBaier and Joost-Pieter Katoen. Principles of Model Checking, The MIT Press.
3. Michael Huth and Mark Ryan. Logic in Computer Science – Modelling and Reasoning about Systems, Cambridge University Press.

Model Question paper

Course Code: R25CST113

Automated Verification

Time: 2.5 Hours **Max. Marks: 60**

Part A

(Answer all questions. Each question carries 5 marks) 5x5 = 25 Marks

1. Use Tseitin's encoding to convert the formula $x_1 \Rightarrow (x_2 \wedge x_3)$ to Conjunctive Normal Form (CNF).
2. Consider a formula that contains the following set of clauses.

$$C_1 = (\neg x_1 \vee x_2),$$

$$C_2 = (\neg x_1 \vee x_3 \vee x_5),$$

$$C_3 = (\neg x_2 \vee x_4),$$

$$C_4 = (\neg x_3 \vee \neg x_4),$$

$$C_5 = (x_1 \vee x_5 \vee \neg x_2),$$

$$C_6 = (x_2 \vee x_3),$$

$$C_7 = (x_2 \vee \neg x_3),$$

$$C_8 = (x_6 \vee \neg x_5)$$

   a. Draw a partial implication graph for decision level 6, after a decision $x_1 = 1$.
   b. Draw a partial implication graph after learning a conflict clause $C_9 = (x_5 \vee \neg x_1)$ and backtracking to level 3.
3. Prove the equivalence between the following two programs by replacing the multiplications with uninterrupted functions.

```
int power3(int in) {
int i, out_a;
    out_a = in;
    for (i = 0; i < 2; i++)
        out_a = out_a * in;
    return out_a; }
```

```
int power3_new(int in) {
int out_b;

    out_b = (in * in) * in;
    return out_b; }
```

4. Give the algorithm to convert a formula in Equality Logic to an equisatisfiable formula in Propositional Logic.
5. Consider the following system of constraints.

Part B

(Answer any 5 questions. Each question carries 7 marks) 5x7 = 35 Marks

6. a. Let T1 and T2 be two theories whose satisfiability problem is decidable and in the same complexity class. Is the satisfiability problem in a T1 formula reducible to a satisfiability problem in T2 formula? Why?

b. Let T1 T2 be two theories whose satisfiability problems are reducible to one another. Are they in the same complexity class? Why?

7. Explain any three decision heuristics used in SAT solvers.

8. Show a formulation in propositional logic of the following problem: given a directed graph, does it contain a Hamiltonian cycle?

9. Illustrate how uninterpreted functions can be used for verifying a compilation process with translation validation.

10. Give the algorithm to simplify an equality logic formula. Illustrate with an example.

11. Develop a SAT solver that verifies a compilation process with Translation Validation.

12. A 0-1 integer linear system is an integer linear system in which all variables are constrained to be either 0 or 1. Show how a 0-1 integer linear system can be translated to a Boolean formula. What is the complexity of the translation?

SEMESTER I

PROGRAM ELECTIVE II

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
R25CST123	ADVANCED COMPUTER NETWORKS	PROGRAM ELECTIVE 2	3	0	0	3

Course Objective

1. To provide a strong foundation in advanced networking concepts, including congestion control, quality of service, multimedia communication, and large-scale routing.
2. To develop analytical skills to evaluate protocols, traffic engineering methods, and emerging paradigms such as Software Defined Networking (SDN) and Network Function Virtualization (NFV).
3. To familiarize learners with security challenges in modern networks and techniques for ensuring confidentiality, authentication, and privacy.
4. To encourage innovation by enabling students to design and develop solutions for next-generation networks such as IoT, cloud, and data center networking.

Course Outcomes: After the completion of the course the student will be able to

	Course Outcome	Bloom's Knowledge Level(KL
CO 1	Examine the problem of scalability for routing and also identify the challenges in mobile and multicast routing.	K3
CO 2	Choose the technique that provides the Quality-of-Service needs of a particular application.	K4
CO 3	Survey various wired and wireless networking technologies including wireless cellular technologies.	K3
CO 4	Classify the multimedia applications in the Internet and compile the various protocols handling these applications.	K4

CO 5	Describe examples of current networking trends and identify the technological gaps.	K5

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	3		2	2	3	3	
CO 2	3		2	3	3	3	
CO 3	3		2	2	3	3	
CO 4	3		3	3	3	3	
CO 5	3	2	3	3	3	2	2

(1-Weak, 2-Medium, 3- Strong)

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	19
Analyze	19
Evaluate	22

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred): 15 marks

Course based task/Seminar/Data collection and interpretation: 15 marks

Test paper, 1 no.: 10 marks

Test paper shall include minimum 80% of the syllabus.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College. There will be two parts; Part A and Part B. Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question (such questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students).

Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to the theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

Note: The marks obtained for the ESE for an elective course shall not exceed 20%over the average ESE mark % for the core courses. ESE marks awarded to students for each elective course shall be normalized accordingly. For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60%.

SYLLABUS

Module	Description	Contact Hours
1	Advanced Internetworking The Global Internet, Routing Areas, Interdomain Routing -BGP, IP Version 6, Multicast, Multicast Addresses, Multicast Routing -DVMRP-PIM-MSDP, Routing to a mobile node, Mobile IP, TCP and Mobility, Mobile TCP	8
2	Internetwork Quality of Service QoS Architectural Framework - Integrated Services Architecture – RSVP - Differentiated Services, Multiprotocol Label Switching- Destination-Based Forwarding - Explicit Routing Virtual Private Networks and Tunnels, Performance issues in networks, Delay Tolerant Networking	8
3	Wired: DSL, Cable Networks, SONET, ATM, VLAN, Wireless: Satellite Networks, (3) (4) (3) (5) (2) (3) (4) (7) (4) (3) WiMAX.Cellular Networks: Introduction-Wireless links and Network characteristics -CDMA, Cellular Internet access - An overview of cellular network architecture, 3G cellular data networks, 4G LTE Cellular networks - LTE Protocol Stacks -LTE Radio Access Network - Additional LTE functions, 5G Cellular networks, Managing mobility in cellular networks, Wireless and Mobility-Impact on higher level protocols, Personal Area Networks: Bluetooth, Zigbee.	9
4	Multimedia in the Internet: Streaming stored audio/video, Streaming live audio/video, Real time interactive audio/video, Real time Interactive Protocols: RTP- RTCP-SIP-H.323, SCTP Compression: Audio Compression, Image compression- JPEG, Video Compression- MPEG	7

5	Overlay Networks: Routing overlays -Resilient overlay networks, Peer-Peer Networks – Bit Torrent Distributed Hash Tables, Content Distribution networks, Software Defined Networks: Architecture – Control and Data Planes – Open Flow – SDN Controllers, Network Function Virtualization, Data Center Networking	**8**

COURSE PLAN

No	Topic	No. of Lectures (40 hrs)
1	Module 1: Advanced Internetworking	8
1.1	The Global Internet, Routing Areas	1
1.2	Inter-domain Routing -BGP	1
1.3	IP Version 6	1
1.4	Multicast, Multicast Addresses	1
1.5	Multicast Routing – DVMRP	1
1.6	PIM, MSDP	1
1.7	Routing to a mobile node, Mobile IP	1
1.8	TCP and Mobility, Mobile TCP	1
2	Module 2: Internetwork Quality of Service	8
2.1	QoS Architectural Framework	1
2.2	Integrated Services Architecture	1
2.3	RSVP - Differentiated Services	1
2.4	Multiprotocol Label Switching,	1
2.5	Virtual Private Networks and Tunnels	1
2.6	Destination-Based Forwarding - Explicit Routing	1
2.7	Performance issues in networks	1
2.8	Delay Tolerant Networking	1
3	Module 3: Networking Technologies	**7**
3.1	Wired: DSL, Cable Networks, SONET,	**1**
3.2	ATM, VLAN	**1**
3.3	Wireless: Satellite Networks, WiMAX	**1**
3.4	Cellular Networks: Introduction-Wireless links and Network characteristics CDMA,	**1**
3.5	Cellular Internet access-An overview of cellular network architecture, 3G cellular data networks,	**1**
3.6	4G LTE Cellular networks - LTE Protocol Stacks -LTE Radio Access Network Additional LTE functions	**1**
3.7	5G Cellular networks	**1**

3.8	Managing mobility in cellular networks, Wireless and Mobility-Impact on higher level protocols	**1**
3.9	Personal Area Networks: Bluetooth, Zigbee	**1**
4	Module 4: Networking Applications	**7**
4.1	Multimedia in the Internet: Streaming stored audio/video, Streaming live audio/video,	**1**
4.2	Real time interactive audio/video	**1**
4.3	Real time Interactive Protocols: RTP- RTCP	**1**
4.4	H-323	**1**
4.5	SIP, SCTP	**1**
4.6	Compression: Audio Compression, Image compression-JPEG,	**1**
4.7	Video Compression- MPEG	**1**
5	Module 5: Current Topics in Networking	**8**
5.1	Overlay Networks: Routing overlays	**1**
5.2	-Resilient overlay networks	**1**
5.3	Peer-Peer Networks – Bit Torrent – Distributed Hash Tables,	**1**
5.4	Content Distribution networks	**1**
5.5	Software Defined Networks: Architecture – Control and Data Planes	**1**
5.6	Open Flow, SDN Controllers	**1**
5.7	Network Function Virtualization	**1**
5.8	Data Center Networking	**1**

References

1. Larry Peterson and Bruce Davie, Computer Networks - A Systems Approach, Morgan Kaufmann, 6th edition, 2022
2. James F. Kurose and Keith W. Ross, Computer Networking A Top-Down Approach, Pearson, 8th edition, 2022
3. Jochen Schiller, Mobile Communications, Addison-Wesley, 2nd edition, 2003
4. William Stallings, Data and Computer Communications, Pearson, 5th edition, 2017 5. Andrew Tanenbaum and David Wetherall, Computer Networks, Pearson, 5th edition, 2010
6. Behrouz A Forouzan, Data Communications and Networking, McGraw Hill, 5th edition, 2017 7. Thomas D. Nadeau and Ken Gray, SDN – Software Defined Networks, O'Reilly, 2013

MODEL QUESTION PAPER

Course Code: R25CST123

ADVANCED COMPUTER NETWORKS

Max. Marks: 60 Duration: 2.5 hours

PART A

(Answer *ALL* questions; each question carries 5 marks)

1. Illustrate with an example how standard TCP can be enhanced to support mobile users.
2. Explain the architectural framework for supporting Quality of Service in packet networks.
3. Examine the role of core network in 3G cellular data network.
4. There is one sender and eight receivers in a real time multimedia communication system. If the sender is sending multimedia data at 2 Mbps, how many RTCP packets can be sent by the sender and each receiver in a second? The system allocates 75 percent of the RTCP bandwidth to the receivers and 25 percent to the sender. The average size of each RTCP packet is 125 bytes.
5. Define OpenFlow specification used in SDN.

PART B

(Answer *any FIVE* questions; each question carries 7 marks)

6. (a) X, Y, Z are three Ass. X and Z are connected through Y. X has a peering agreement with Y and Y with Z. Z moves all traffic from Y but does not forward traffic from X. Can Z use BGP to implement this policy? (4)

(b) How does PIM solve the scalability problem of existing multicast protocols. (3)

7. (a) Derive the hexadecimal form of representation of the following link local multicast address: (4)

(i) a permanently-assigned multicast group address of 66

(ii) a transient multicast group address of 316

(b) A foreign network has a foreign agent. Explain if it is possible for two mobile nodes in the foreign network to use the same care-of address in mobile IP. (3)

8. (a) Justify the need for Resource Reservation in multicast transmission. Explain the methodologies used for reducing the branch cost with prediction in instruction level parallelism. (4)

(b) How is VPN implemented using MPLS? (3)

9. (a) Elaborate on the various elements of 4G LTE network and the interaction between them. (5)

(b) Calculate the minimum time required to download 2*106 bytes using ADSL modem with minimum rate. (2)

10. (a) Sketch the superframe format of Zigbee 802.15.4 standard. (3)

(b) Name some applications which use Zigbee standard and justify its use. (4)

11. Describe H323 architectural model for Internet Telephony. (7)

12. (a) Comment on the statement "Distributed Hash Tables are said to build structured P2P networks". (4)

(b) Explain Data Center Networking. (3)

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
R25CST123	PATTERN RECOGNITION	PROGRAM ELECTIVE 2	3	0	0	3

Course Objective

1. To introduce fundamental concepts of statistical, syntactic, and structural approaches to pattern recognition.
2. To develop analytical skills for applying supervised and unsupervised learning algorithms to real-world recognition problems.
3. To familiarize learners with feature extraction, dimensionality reduction, and classification techniques used in machine learning.
4. To enable evaluation of recognition models in terms of accuracy, robustness, and computational efficiency.

Course Outcomes: After the completion of the course the student will be able to

Course Outcome		Bloom's Knowledge Level(KL
CO 1	Apply probability and numerical methods in statistical pattern recognition.	K3
CO 2	Apply statistical methods in feature selection.	K3
CO 3	Apply linear algebra and statistical methods in parameter and non-parameter estimation.	K3
CO 4	Apply the technique of decision trees in pattern recognition.	K3

CO 5	Analyze the use of deep learning networks and artificial neural networks in pattern recognition.	K4
CO6	Design, Develop, Implement and Present innovative ideas in problem solving with various pattern recognition techniques.	K5

Mapping of course outcomes with program outcomes

	PO 1	**PO 2**	**PO 3**	**PO 4**	**PO 5**	**PO 6**	**PO 7**
CO 1	3		2	2	3	3	
CO 2	3		2	3	3	3	
CO 3	3		2	2	3	3	
CO 4	3		3	3	3	3	
CO 5	3	2	3	3	3	2	2

(1-Weak, 2-Medium, 3- Strong)

Assessment Pattern

Bloom's Category	**End Semester Examination**
Apply	19
Analyze	19
Evaluate	16

Create	8

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred): 15 marks

Course based task/Seminar/Data collection and interpretation: 15 marks
Test paper, 1 no.: 10 marks

Test paper shall include minimum 80% of the syllabus.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College. There will be two parts; Part A and Part B. Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question (such questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students).

Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to the theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

Note: The marks obtained for the ESE for an elective course shall not exceed 20%over the average ESE mark % for the core courses. ESE marks awarded to students for each elective course shall be normalized accordingly. For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60%.

SYLLABUS

Module	Description	Contact Hours
1	Introduction to Pattern Recognition No Basics of pattern recognition systems, various applications, Machine Perception, classification of pattern recognition systems. Design of Pattern recognition system, Pattern recognition Life Cycle. Statistical Pattern Recognition: Review of probability theory, Gaussian distribution. Normal density and discriminant functions.	7
2	Feature selection – Outlier removal – Data normalization – Missing data, The Peaking phenomenon, Feature selection using statistical hypothesis testing- Hypothesis testing basics – Application of t-Test in feature selection. Class separability measures-Divergence-Chernoff bound and Bhattacharya distance-Scatter matrices, Feature subset selection –Scalar feature selection, Feature vector selection.	10
3	Clustering Algorithms - Unsupervised learning and clustering - Criterion functions for clustering. Cluster validation. Fuzzy clustering algorithms- Point representatives- quadratic surfaces and representatives – hyper plane representatives. Binary morphology clustering algorithms (BMCAs) – Discretization – Morphological operations - Determination of clusters in a discrete binary set- Assignment of feature vectors to clusters – The algorithmic scheme, Boundary detection algorithms.	9

4	Dimensionality reduction: Principal component analysis - its relationship to Eigen analysis. Fisher discriminant analysis - Generalised Eigen analysis. Eigen vectors/Singular vectors as dictionaries. Factor Analysis, Total variability space - a dictionary learning method.Non negative matrix factorisation - a dictionary learning method. Linear discriminant functions: Gradient descent procedures, Perceptron.	**8**
5	Artificial neural networks: Review of Artificial neural network concepts, convolutional neural networks, recurrent neural networks. Non-metric methods for pattern classification: non-numeric data or nominal data. Decision trees: Classification and Regression Trees (CART).	**6**

COURSE PLAN

No	Topic	No. of Lectures (40 hrs)
1	**Module 1: Introduction to Pattern Recognition**	7
1.1	Basics of pattern recognition systems, applications	1
1.2	Machine Perception, Classification of pattern recognition systems	1
1.3	Design of Pattern recognition system	1
1.4	Pattern recognition Life Cycle	1
1.5	Statistical Pattern Recognition	1
1.6	Review of probability theory	1
1.7	Normal density and discriminant functions	1
2	**Module 2: Feature Selection**	10
2.1	Feature selection – Outlier removal	1
2.2	Data normalization – Missing data	1
2.3	The peaking phenomenon	1

2.4	Feature selection using statistical hypothesis testing	1
2.5	Hypothesis testing basics – Application of tTest in feature selection	1
2.6	Class separability measures-Divergence	1
2.7	Chernoff bound and Bhattacharya distance	1
2.8	Scatter matrices	1
2.9	Feature subset selection –Scalar feature selection	1
2.10	Feature vector selection	1
3	**Module 3: Clustering Algorithms**	9
3.1	Unsupervised learning and clustering	1
3.2	Criterion functions for clustering. Cluster validation.	1
3.3	Fuzzy clustering algorithms- Point representatives	1
3.4	Quadratic surfaces and representatives – hyper plane representatives.	1
3.5	Binary morphology clustering algorithms (BMCAs)	1
3.6	Discretization	1
3.7	Morphological operations - Determination of clusters in a discrete binary set	1
3.8	Assignment of feature vectors to clusters	1
3.9	The algorithmic scheme, Boundary detection algorithms.	1
4	**Module 4: Dimensionality reduction**	7
4.1	Principal component analysis - its relationship to Eigen analysis	1
4.2	Fisher discriminant analysis	1
4.3	Generalised Eigen analysis	1
4.4	Eigen vectors/Singular vectors as dictionaries	1
4.5	Total variability space - a dictionary learning method	1
4.6	Non negative matrix factorisation - a dictionary learning method	1

4.7	Linear discriminant functions: Gradient descent procedures	1
4.8	Perceptron	1
5	**Module 5: Artificial neural networks and Pattern Classification**	8
5.1	Review of Artificial neural networks, Introduction to deep neural networks	1
5.2	Convolutional neural networks	1
5.3	Recurrent neural networks	1
5.4	Non-metric methods for pattern classification: Non-numeric data or nominal data	1
5.5	Decision trees: Classification and Regression Trees (CART) lecture 1	1
5.6	Decision trees: Classification and Regression Trees (CART) lecture 2	1

References

1. S.Theodoridis and K.Koutroumbas, "Pattern Recognition", 4th Ed., Academic Press, 2009
2. C.M.Bishop, "Pattern Recognition and Machine Learning", Springer, 2006
3. R.O.Duda, P.E.Hart and D.G.Stork, "Pattern Classification", John Wiley, 2001
4. Hastie, T., Tibshirani, R. and Friedman, J. "The Elements of Statistical Learning". Springer. 2001.

MODEL QUESTION PAPER

Course Code: R25CST123

PATTERN RECOGNITION

Max. Marks: 60 Duration: 2.5 hours

PART A

(Answer *ALL* questions; each question carries 5 marks)

1. In a town it was estimated that 3% of people have a particular disease. Adiagnosis test was conducted for all the people, which yielded 8% false positiveand 92% true positive results. A person is found as positive after the test. What isthe probability that this person is truly having the disease?

2. How does morphological operations play a role in pattern recognition?

3. How can visual imagery be analysed using convolutional neural networks?

4. How does a decision tree handle continuous attributes?

5. Define the terms: weights, bias, activations with respect to neural networks

Part B

(Answer any five questions. Each question carries 7 marks)

6. Illustrate the design principles of pattern recognition system with an example.

7. Derive the fuzzy C spherical shells (FCSS) algorithm for the case that spherical clusters are to be identified.

8. Discuss the significance of pre-processing in feature selection. Illustrate any two methods used for pre-processing.

9. How can artificial neural networks be applied in Pattern recognition? Also illustrate the features of recurrent neural networks.

10. Explain the role of dimensionality reduction in pattern recognition. Compare any two commonly used techniques with examples.

11. Discuss the importance of feature extraction in classification tasks. Illustrate with an example how improper feature extraction can affect classifier performance.

12. Construct a decision tree using the following data.

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

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
R25CST123	ADVANCED COMPUTER ARCHITECTURE	PROGRAM ELECTIVE 2	3	0	0	3

Course Objective

1. To provide a solid foundation in advanced computer architecture by introducing principles of design, performance evaluation, and the hardware/software interface.

2. To develop analytical and problem-solving skills in memory hierarchy, pipelining, multiprocessor systems, and parallelism (thread-level and data-level).

3. To familiarize learners with emerging architectural trends such as SIMD, GPUs, and modern processor designs to enable innovation in hardware/software solutions.

Course Outcomes: After the completion of the course the student will be able to

Course Outcome		Bloom's Knowledge Level(KL)
CO 1	Identify and solve the advanced issues in design of computer processors, caches and memory.	K3
CO 2	Analyze the memory hierarchy design, performance improvement techniques and cache optimization techniques.	K4
CO 3	Analyze the working and features of branching and exception handling in pipeline architecture.	K4
CO 4	Analyze the operation of multiprocessors, thread-level parallelism, and the concepts of data-level parallelism including SIMD and GPU processors.	K5
CO 5	Design, Develop, Implement and Present innovative ideas on advanced computer architecture and techniques.	K6

Note: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	3		2	2	3		
CO 2	3		2	3	3		
CO 3	3		2	2	3		
CO 4	3		3	3	3		
CO 5	3	2	3	3	3	2	2

(1-Weak, 2-Medium, 3- Strong)

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	16
Analyse	20
Evaluate	12
Create	12

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation Pattern:

Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred): 15 marks

Course based task/Seminar/Data collection and interpretation: 15 marks

Test paper, 1 no.: 10 marks

Test paper shall include minimum 80% of the syllabus.

End Semester Examination Pattern:

Total Marks: 60

The end semester examination will be conducted by the respective College. There will be two parts; Part A and Part B. Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question (such questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students).

Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to the theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks.

Note: The marks obtained for the ESE for an elective course shall not exceed 20%over the average ESE mark % for the core courses. ESE marks awarded to students for each elective course shall be normalized accordingly. For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60%.

SYLLABUS

MODULE	DESCRIPTION	CONTACT HOURS
1	**Design and Analysis:** Principles of computer design, Fallacies and Pitfalls, Instruction Set Principles- Classifying instruction set architecture, Memory addressing, Type and size of operands, Operations in the instruction set, Instruction for control flow, Encoding an instruction set, Role of compiler.	8
2	**Memory Hierarchy:** Introduction, Cache performance, Basic cache optimizations, Virtual memory–Techniques for fast address translation, Protection via virtual memory, Fallacies and Pitfalls, Case study of Pentium/Linux memory system-Pentium address translation.	8
3	**Pipelining:** Introduction, Pipeline hazards, Static branch prediction and dynamic branch prediction, Implementation of MITS, Basic pipeline of MITS, Implementing the control in MITS pipeline, Dealing with branches in pipeline, Dealing with	8

	exceptions, Handling of multi-cycle operations, Maintaining precise exceptions, Case study of MITS R4000 pipeline.	
4	**Thread Level Parallelism:** Introduction, Centralized Shared-Memory Architectures, Performance of Symmetric Shared Memory Multiprocessors, Distributed Shared-Memory and Directory-Based Coherence, Synchronization: The Basics, Models of Memory Consistency: An Introduction, Crosscutting Issues, Case study of Sun T1 Multiprocessor.	8
5	**Data Level Parallelism:** Vector architecture, SIMD instruction set, Extension for multimedia, Graphic Processing Units, Case study Envida GPU instruction set architecture, GPU memory structure, Innovations in GPU architecture, Comparisons between vector architecture and GPUs, Comparisons between multimedia SIMD computers and GPUs, Loop level parallelism, Finding dependencies, Eliminating dependencies.	8

COURSE PLAN

No	Topic	No. of Lectures
1	**Module 1: Design and Analysis(8)**	
1.1	Principles of computer design	1
1.2	Fallacies and Pitfalls	1
1.3	Instruction Set Principles- Classifying instruction set architecture	1
1.4	Memory addressing, Type and size of operands	1
1.5	Operations in the instruction set	1
1.6	Instruction for control flow	1
1.7	Encoding an instruction set	1
1.8	Role of compiler	1
2	**Module 2: Memory Hierarchy (8)**	
2.1	Introduction	1
2.2	Pipeline hazards	1

2.3	Static branch prediction and dynamic branch prediction	1
2.4	Implementation of MITS, Basic pipeline of MITS	1
2.5	Implementing the control in MITS pipeline	1
2.6	Dealing with branches in pipeline, Dealing with exceptions	1
2.7	Handling of multi-cycle operations, Maintaining precise exceptions	1
2.8	Case study of MITS R4000 pipeline	1
3	**Module 3: Multiprocessors and Thread level Parallelism (8)**	
3.1	Introduction	1
3.2	Centralized Shared-Memory Architectures	1
3.3	Performance of Symmetric Shared-Memory Multiprocessors	1
3.4	Distributed Shared-Memory and Directory-Based Coherence	1
3.5	Synchronization: The Basics	1
3.6	Models of Memory Consistency: An Introduction	1
3.7	Crosscutting Issues	1
3.8	Case study Sun T1 Multiprocessor	1
4	**Module 4: Multiprocessors and Thread level Parallelism (8)**	
4.1	Introduction	1
4.2	Centralized Shared-Memory Architectures	1
4.3	Performance of Symmetric Shared-Memory Multiprocessors	1
4.4	Distributed Shared-Memory and Directory-Based Coherence	1
4.5	Synchronization: The Basics	1
4.6	Models of Memory Consistency: An Introduction	1
4.7	Crosscutting Issues	1
4.8	Case study Sun T1 Multiprocessor	1
5	**Module 5: Data Level Parallelism (8)**	
5.1	Vector architecture, SIMD instruction set	1
5.2	Extension for multimedia, Graphic Processing Units	1
5.3	Case study Envida GPU instruction set architecture	1

5.4	GPU memory structure	1
5.5	Innovations in GPU architecture, Comparisons between vector architecture and GPUs	1
5.6.	Comparisons between multimedia SIMD computers and GPUs	1
5.7	Loop level parallelism	1
5.8	Finding dependencies, Eliminating Dependencies	

Reference Books

12. E. M. Phillips and D. S. Pugh, "How to get a PhD - a handbook for PhD students and their supervisors", Viva books Pvt Ltd
13. G. L. Squires, "Practical physics", Cambridge University Press.
14. Antony Wilson, Jane Gregory, Steve Miller, Shirley Earl, Handbook of Science Communication, Overseas Press India Pvt Ltd, New Delhi, 1st edition 2005
15. C. R. Kothari, Research Methodology, New Age International, 2004
16. Panneerselvam, Research Methodology, Prentice Hall of India, New Delhi, 2012
17. Leedy P. D., Practical Research: Planning and Design, McMillan Publishing Co
18. Day R. A., How to Write and Publish a Scientific Paper, Cambridge University Press, 1989.
19. William Strunk Jr., Elements of Style, Fingerprint Publishing, 2020
20. Peter Medawar, 'Advice to Young Scientist', Alfred P. Sloan Foundation Series, 1979.
21. E. O. Wilson, Letters to a Young Scientist, Liveright, 2014.
22. R. Hamming, You and Your Research, 1986 Talk at Bell Labs.

MODEL QUESTION PAPER

Course Code: R25CST123

ADVANCED COMPUTER ARCHITECTURE

Max. Marks: 60 **Duration: 2.5 hours**

PART A

(Answer *ALL* questions; each question carries 5 marks)

1. Impact of optimization improves the performance of complier. Justify your answer. Also mention types of optimizations.

2. Determine whether a 32 KB four-way set associative L1 cache has a faster memory access time than a 32 KB two-way set associative L1 cache. Assume the miss penalty to L2 is 15 times the access time for the faster L1 cache. Ignore misses beyond L2. Which has the faster average memory access time? (Miss rate for two-way set associative cache is 0.038 and four-way set associative cache is 0.037).

3. Analyze the type of hazards may occur in the following code.

    LW R1, 0(R2)

    SUB R4, R1

    AND R6, R1, R7

    OR R8, R1, R9

4. Suppose we have an application running on a 32-processor multiprocessor, which has a 200ns time to handle reference to a remote memory. For this application, assume that all the references except those involving communication hit in the local memory hierarchy, which is slightly optimistic. Processors are stalled on a remote request, and the processor clock rate is 2GHz. If the base CPI (assuming that all references hit in the cache) is 0.5, evaluate how much faster is the multiprocessor if there is no communication versus if 0.2% of the instructions involve a remote communication reference?

5. Differentiate between thread-level parallelism and data-level parallelism. Illustrate with an example each.

PART B

(Answer *any FIVE* questions; each question carries 7 marks)

6. A processor has a CPI of 1.1 without memory stalls. Assume 40% of instructions access memory. The miss rate of L1 data cache is 5% with a miss penalty of 40 cycles. Calculate the effective CPI. How does adding a write buffer change performance?

7. “Fully associative caches do not have conflict misses”. Examine the statement.

8. Consider an in-order execution computer. Assume that the cache miss penalty is 200 clock cycles, and all instructions normally take 1.0 clock cycles. Assume that the average

miss rate is 2%, there is an average of 1.5 memory references per instruction, and the average number of cache misses per 1000 instructions is 30. What is the impact on performance when behavior of the cache is included? Calculate the impact using both misses per instruction and miss rate.

9. Explain the methodologies used for reducing the branch cost with prediction in instruction level parallelism.

10. Illustrate and explain extending the MIPS pipeline to handle multicycle operations.

11. Explain multiprocessor cache coherence.

12. Propose a heterogeneous computing architecture integrating CPUs, GPUs, and AI accelerators for autonomous vehicle control. Justify the design with expected benefits.

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
R25CST123	NATURAL LANGUAGE PROCESSING AND TEXT MINING	PROGRAM ELECTIVE 2	3	0	0	3

Course Objectives:

1. To provide an exposure to the concepts and techniques in Natural language processing and Text mining.
2. Fundamental concepts and practical applications of Natural Language Processing (NLP) are covered in this course.
3. To help the learners to analyze and interpret textual data.

Course Outcomes: After the completion of the course the student will be able to

Course Outcome		Bloom's Knowledge Level(KL)
CO 1	Apply different approaches of syntax and semantics in NLP.	K3
CO 2	Employ approaches to generate dialogue and summarisation within NLP.	K3
CO 3	Apply different statistical approaches to machine translation.	K3
CO 4	Research, analyze and deploy appropriate machine learning techniques in NLP including hidden Markov models and unsupervised methods.	K4
CO 5	Use text mining concepts and methods to model real-world problems and develop technical solutions.	K4
CO6	Design, develop and implement NLP and text mining methods to solve real world problems.	K6

Note: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	2				2	1	
CO 2	2		3	3	2	1	
CO 3	2		3	3	2	1	
CO 4	3		3	3	2	2	
CO 5	3		3	3	2	2	
CO 6	3	2	3	3	3	2	2

(1-Weak, 2-Medium, 3- Strong)

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	70%-80%
Analyse	30%-40%
Evaluate	
Create	

Mark distribution

Total Marks	CIA	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks ii. Course based task / Seminar/ Data collection and interpretation : 15 marks iii. Test paper (1 number) : 10 marks Test paper shall include minimum 80% of the syllabus. Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students. End Semester Examination Pattern: The end semester examination will be conducted by the respective College. There will be two parts; Part A and Part B. Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks Total duration of the examination will be 150 minutes. Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly. For example if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

SYLLABUS

Module	Description	Contact Hours
1	**Introduction** Natural Language Processing (NLP) - Syntax, semantics, pragmatics, and ambiguity in NLP, Regular Expressions, Text Normalisation, Edit Distance. N-gram Language Models-N-Grams, Evaluating Language Models, Generalisation and Zeros, Smoothing, Kneser-Ney Smoothing, The Web and Stupid Backoff, Perplexity's Relation to Entropy.	6
2	**Neural Language Models, Vector Semantics and Embeddings** Neural Networks and Neural Language Models-Units, Feed-Forward Neural Networks, Training Neural Nets, Neural Language Models. Vector Semantics and Embeddings-Lexical Semantics, Vector Semantics, Words and Vectors, Cosine for measuring similarity, TF-IDF: Weighing terms in the vector, Applications of the tf-idf vector model, Word2vec, Visualizing Embeddings, Semantic properties of embeddings, Bias and Embeddings, Evaluating Vector Models.	8
3	**Sentiment Classification and Part-of-Speech Tagging** Sentiment Classification –Sentiment classification. Machine Learning for Sentiment Classification - Training the Classifier (Naive Bayes, Logistic Regression, Support Vector Machine, Decision Tree, Random Forest), Optimising for Sentiment Analysis - Other text classification tasks – Evaluation of classification models: Precision, Recall, F-measure, Test sets and Cross-validation, Statistical Significance Testing. Part-of-Speech Tagging-English Word Classes, The Penn Treebank Part-of-Speech Tagset, Part-of-Speech Tagging, HMM Part-of-Speech Tagging, Maximum Entropy Markov Models, Bi-directionality, Part-of Speech Tagging for Morphological Rich Languages. Information Extraction-Named Entity Recognition, Relation Extraction, Extracting Times, Extracting Events and their Times, Template Filling.	10

4	**Sequence Processing with Recurrent Networks** Sequence Processing with Recurrent Networks-Simple Recurrent Neural Networks, Applications of Recurrent Neural Networks, Deep Networks: Stacked and Bidirectional RNNs, Managing Context in RNNs: LSTMs and GRUs, Words, Subwords and Characters Neural Language Models and Generation Revisited, Encoder-Decoder Networks, Attention, Applications of Encoder-Decoder Networks. Case study: Machine translation, Question Answering	**7**
5	**Text Mining** Document representation - representing unstructured text documents with appropriate format and structure, automated text mining algorithms. Text Mining: Text categorization, Text clustering, Topic modeling, Applications - classification, image annotation, collaborative filtering, and hierarchical topical structure modeling. Document summarization - Extraction- based summarization methods Sentiment analysis - concept, sentiment polarity prediction, review mining, aspect identification. Text visualization - introduction to mathematical and programming tools.	**9**

COURSE PLAN

No	Topics	No. of Lectures (40)
1	**Introduction**	
1.1	Natural Language Processing (NLP) - Syntax, semantics, pragmatics, and ambiguity in NLP	**1**
1.2	Regular Expressions	**1**
1.3	Text Normalisation, Edit Distance	**1**

1.4	N-gram Language Models-N-Grams, Evaluating Language Models	**1**
1.5	Generalisation and Zeros, Smoothing	**1**
1.6	Kneser-Ney Smoothing, The Web and Stupid Backoff, Perplexity's Relation to Entropy	**1**
2	**Neural Language Models, Vector Semantics and Embeddings**	

2.1	Units, Feed-Forward Neural Networks	**1**
2.2	Training Neural Nets	**1**
2.3	Neural Language Models	**1**
2.4	Lexical Semantics, Vector Semantics	**1**
2.5	Words and Vectors, Cosine for measuring similarity	**1**
2.6	TF-IDF: Weighing terms in the vector, Applications of the tf-idf vector model	**1**
2.7	Word2vec, Visualizing Embeddings	**1**
2.8	Semantic properties of embeddings, Bias and Embeddings, Evaluating Vector Models	**1**
3	**Sentiment Classification and Part-of-Speech Tagging**	
3.1	sentiment classification. Machine Learning for Sentiment Classification	**1**
3.2	Training the Classifier (Naive Bayes, Logistic Regression)	**1**
3.3	Training the Classifier (Support Vector Machine, Decision Tree, Random Forest)	**1**
3.4	Optimising for Sentiment Analysis - Other text classification tasks	**1**
3.5	Evaluation of classification models: Precision, Recall, F-measure, Test sets and Cross-validation, Statistical Significance Testing.	**1**
3.6	English Word Classes, The Penn Treebank Part-of-Speech Tagset	**1**

3.7	Part-of-Speech Tagging, HMM Part-of-Speech Tagging,	**1**
3.8	Maximum Entropy Markov Models, Bi-directionality, Part-of Speech Tagging for Morphological Rich Languages.	**1**
3.9	Information Extraction-Named Entity Recognition, Relation Extraction	**1**
3.10	Extracting Times, Extracting Events and their Times, Template Filling	**1**
4	**Sequence Processing with Recurrent Networks**	
4.1	Simple Recurrent Neural Networks, Applications of Recurrent Neural Networks	**1**
4.2	Deep Networks: Stacked and Bidirectional RNNs	**1**
4.3	Managing Context in RNNs: LSTMs	**1**
4.4	GRUs, Words, Subwords and Characters	**1**
4.5	Neural Language Models and Generation Revisited	**1**
4.6	Encoder-Decoder Networks, Attention, Applications of Encoder-Decoder Networks.	**1**
4.7	Case study: Machine translation, Question Answering	**1**
5	**Text Mining**	
5.1	Representing unstructured text documents with appropriate format and Structure	**1**
5.2	automated text mining algorithms	**1**
5.3	Text categorization	**1**
5.4	Text clustering	**1**
5.5	Topic modelling, Applications - classification, imagine annotation, collaborative filtering, and hierarchical topical structure modeling	**1**

5.6	Document summarization - Extraction- based summarization methods	**1**
5.7	Sentiment analysis - concept, sentiment polarity prediction	**1**
5.8	Review mining, aspect identification.	**1**
5.9	Text visualization - introduction to mathematical and programming tools.	**1**

Reference Books

1. Daniel Jurafsky and James H. Martin. Speech and Language Processing (2nd ed), Pearson International edition, 2008
2. Manning C, Schuetze H. Foundations of Statistical Natural Language Processing, MIT Press, 1999
3. James Allen, "Natural Language Understanding", 2/E, Addison-Wesley, 1994
4. Charu C. Aggarwal and Cheng Xiang Zhai, “Mining Text Data”, Springer, 2012.
5. Text Mining Classification, Clustering, and Applications - Ashok N. Srivastava, Mehran Sahami, CRC Press.

MODEL QUESTION PAPER

Course Code: R25CST123

NATURAL LANGUAGE PROCESSING AND TEXT MINING

Max. Marks: 60 **Duration: 2.5 hours**

PART A

(Answer ALL questions; each question carries 5 marks)

1. For each sentence, identify whether the different meanings arise from structural ambiguity, semantic ambiguity or pragmatic ambiguity?
   1. Time flies like an arrow
   2. He crushed the key to my heart
2. The confusion matrix for a binary classifier is given below. Compute Precision, Recall, F- score, Specificity and Accuracy of the model

Actual	Predicated	
	Class 1	Class 2
Class 1	14	2
Class 2	8	40

3. Justify whether a model/ program designed to perform sentimental analysis can be adapted for document grouping

4. Describe the structure of GRU cell. How the output is computed at each gate?

5. Produce the extractive summary by creating a document of your own with the title as "Kerala- at a glance". Evaluate the performance of your summarizer by using the appropriate evaluation measures.

PART B

(Answer any FIVE questions; each question carries 7 marks)

6. (a) Define minimum edit distance

(b) Compute the minimum edit distance between the strings INTENTION and EXECUTION using dynamic programming

7 (a) Consider a document collection having 37 documents

Find the TF-IDF embedding of the following words, “Battle, Good, Fool, Wit” using the tables given below

WordsWords	Document Frequency
Romeo	1
Salad	2
Falstaff	4
Forest	12
Battle	21
Wit	34
Fool	36
Good	37
Sweet	37

Raw Word	Count	Doc1	Doc2	Doc3
Battle	1	0	7	13
Good	114	80	62	89
Fool	36	58	1	4
Wit	20	15	2	3

(b) Find the cosine similarity off the words ”Fool & Wit”

8. Explain the Skip-gram and CBOW methods of word embedding generation

9. (a) Identify the Named Entities in the text given below and tag them with appropriate NE

tags

Citing high fuel prices, Indian Airlines said Friday it has increased by 56 per round trip on flights to some cities also served by lower-cost carriers. Spice Jet, a unit of AMR Corp, immediately matched the move, Spokesman Ram Goel said Indian Airlines, a unit of TATA Corp, said the increase took effect Thursday and applies to most routes where it competes against discount carriers, such as cochin to Delhi and Chennai to Calcutta.

(b) Why NE tagging is considered hard when compared to POS tagging? Give one method to overcome it. Describe the architecture of a RNN that can perform sequence classification. Explain in detail how LSTM can be used for language translation. 12. (a) What application of text mining is coming under the Social Media Platform YouTube? Justify your answer. (b) Consider that you are running a business and you are having a social media page to promote your business. Is it possible to design a sentiment analyser by using the raw data extracted from the promotion page? Construct an architecture for the same.

10. Describe the architecture of a RNN that can perform sequence classification.

11. Explain in detail how LSTM can be used for language translation.

12. (a) What application of text mining is coming under the Social Media Platform YouTube? Justify your answer.

(b) Consider that you are running a business and you are having a social media page to promote your business. Is it possible to design a sentiment analyser by using the raw data extracted from the promotion page? Construct an architecture for the same.

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
R25CST123	ADVANCED COMPILER DESIGN	PROGRAM ELECTIVE 2	3	0	0	3

Course Objective

1. To analyze the different phases of compiler /techniques for designing a compiler.

2. To introduce students to the advanced concepts of compilation phases such as lexical analysis, syntax analysis, semantic analysis, intermediate code generation, code optimization and code generation.

3. This course helps the learners to design and develop compilers for programming languages.

Course Outcomes: After the completion of the course the student will be able to

Course Outcome		Bloom's Knowledge Level(KL)
CO 1	Illustrate lexical rules and grammars for a representative programming language	K2
CO 2	Construct intermediate code representations and code optimization techniques.	K3
CO 3	Experiment with register allocation strategies and code scheduling	K3
CO 4	Inspect programming language design, target machine language design and runtime environment of compilers.	K4
CO 5	Assess recent trends in compiler design and build a compiler for a hypothetical language.	K6

Note: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create

Mapping of course outcomes with program outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	2		2	2	2	1	
CO 2	2		2	3	2	1	
CO 3	2		3	3	2	1	
CO 4	3		3	3	3	2	
CO 5	3		3	3	3	2	2

(1-Weak, 2-Medium, 3- Strong)

Assessment Pattern

Bloom's Category	End Semester Examination
Understand	14
Apply	20
Analyse	16
Create	10

Mark distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 hours

Continuous Internal Evaluation: 40 marks

i. Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

ii. Course based task / Seminar/ Data collection and interpretation : 15 marks

iii. Test paper (1 number) : 10 marks

Test paper shall include minimum 80% of the syllabus.

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the Students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College.

There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

SYLLABUS

MODULE	DESCRIPTION	CONTACT HOURS
1	**Overview of Compiler Design:** Introduction - The phases of Compiler - Lexical Analysis - Role of Lexical Analyzer -Specification and Recognition of Tokens - Context Free Grammar – Symbol - Table Structure, Symbol Attributes and Symbol - Table Entries, Local Symbol - Table Management, Global Symbol - Table Structure, Storage Binding and Symbolic Registers, Approaches to Generating Loads and Stores. Intermediate representation – Issues – High level, medium level, low level intermediate languages –MIR, HIR, LIR – ICANfor	8

	Intermediate code.	
2	**Intermediate Representations:** The value - number method for constructing DAGs - Addresses and instructions – Quadruples - Triples - Storage organization – Static versus dynamic storage allocation - stack allocation of space -Activation trees - Activation records - Garbage collection - Design goals for Garbage collectors - Reference counting garbage collectors - Introduction to trace - based collection - A basic mark - and-sweep collector. Translation of expressions Translation of expressions – Operations within expressions – Incremental translation – Addressing array elements – Translation of array references Control flow – Boolean expressions – Short - circuit code – flow – of - control statements – Control flow translation of Boolean expressions – Avoiding redundant Gotos – Boolean values and jumping code – Backpatching – One - passcode generation using back patching – Back patching for Boolean expressions –Flow- of-control statements – Break, continue and Goto statements Translation of switch statements – syntax directed translation of switch statements - intermediate code for procedures.	9
3	**Code Optimization:** Principal sources of optimization - causes of redundancy preserving transformations - Global common subexpressions - Copy Propagation - Dead code elimination – Code motion –Upward code motion – Downward code motion – Induction variables and reduction in strength -Introduction to data flow analysis - Loops in Flow graphs – Dominators - Introduction to global data flow analysis - Points and Paths - Reaching definitions - Live variable analysis - Data flow analysis of structured programs.	8
4	**Register Allocation and Code Scheduling:** Register allocation and assignment – graph coloring – control flow and low-level optimizations - Inter - procedural analysis and optimization – call graph — register allocation– global	7

	References: – Optimization for memory hierarchy. Code Scheduling – Instruction scheduling – peculative scheduling – Software pipelining – trace scheduling –percolation scheduling.	
5	**Parallelism and Case study**: Instruction - level parallelism - Instruction pipelines and branch delays – pipelined execution –data dependence – dependencies among memory accesses. Case Studies – Sun Compilers for SPARC–IBMXL Compilers – Alpha compilers – PA – RISC assembly language – COOL– (Classroom Object oriented language) – Compiler testing tools – SPIM	8

COURSE PLAN

No	Topic	No. of Lectures
1	**Overview of Compiler Design**	
1.1	Introduction - The phases of Compiler. Lexical Analysis – Role of Lexical Analyzer	1
1.2	Specification and Recognition of Tokens – Context Free Grammar	1
1.3	Symbol – Table Structure - Symbol Attributes and Symbol – Table Entries	1
1.4	Local Symbol – Table Management, Global Symbol – Table Structure,	1
1.5	Storage Binding and Symbolic Registers, Approaches to Generating Loads and Stores.	1
1.6	Intermediate representation – Issues	1
1.7	High level, medium level, low level intermediate languages	1
1.8	MIR, HIR, LIR – ICAN for Intermediate code.	1
2	**Intermediate Representations**	
2.1	The value – number method for constructing DAGs - Addresses and Instructions – Quadruples - Triples.	1
2.2	Storage organization – Static versus dynamic storage allocation - stack Allocation of space. Activation trees – Activation records.	1

2.3	Garbage collection – Design goals for Garbage collectors - Reference Counting garbage collectors. Introduction to trace – based collection -A basic mark – and – sweep collector.	1
2.4	Translation of expressions Translation of expressions –Operations Within expressions–Incremental translation–Addressing array elements.	1
2.5	Translation of array references.	1
2.6	Control flow –Boolean expressions – Short – circuit code - flow- of -Control statements – Control flow translation of Boolean expressions.	1
2.7	Avoiding redundant Go to – Boolean values and jumping code.	1
2.8	Back patching - One-pass code generation using back patching – Backpatching for Boolean expressions.	1
2.9	Flow-of-control statements – Break, continue and Goto statements, Translation of switch statements – syntax directed translation of Switch statements - intermediate code for procedures.	1
3	**Code Optimization**	
3.1	Principal sources of optimization – causes of redundancy - Semantics Preserving transformations.	1
3.2	Global common subexpressions – Copy Propagation – Dead code elimination.	1
3.3	Code motion – Upward code motion - Downward code motion.	1
3.4	Induction variables and reduction in strength.	1
3.5	Introduction to data flow analysis.	1
3.6	Loops in Flow graphs - Dominators	1
3.7	Introduction to global data flow analysis – Points and Paths - Reaching Definitions.	1
3.8	Live variable analysis-Data flow analysis of structured program	1
4	**Register Allocation and Code Scheduling**	
4.1	Register allocation and assignment – graph coloring .	1
4.2	Control flow and low-level optimizations.	1
4.3	Inter-procedural analysis and optimization.	1

4.4	Call graph—register allocation.	1
4.5	Global References: – Optimization for memory hierarchy.	1
4.6	Code Scheduling –Instructions scheduling – Speculative scheduling	1
4.7	Software pipelining. Trace scheduling – percolation scheduling	1
5	**Parallelism and Case study**	
5.1	Optimization of basic blocks.	1
5.2	The DAG representation of basic blocks.	1
5.3	Peep hole optimization.	1
5.4	Instruction- level parallelism-Instruction pipelines and branch Delays – pipelined execution.	1
5.5	Data dependence – dependencies among memory accesses.	1
5.6	Case Studies – Sun Compilers for SPARC.	1
5.7	IBMXL Compilers –Alpha compilers.	1
5.8	PA–RISC assembly language, COOL– (Classroom Object oriented language), Compiler testing tools –SPIM.	1

Reference Books

1. Steven S Muchnik, “Advanced Compiler Design and Implementation”, Morgan Kaufmann publishers, Elsevier Science, India, Indian Reprint 2003.

2. Compilers: Principles, Techniques and Tools (2nd edition), Alfred V. Aho, Monica S.

Lam, Ravi Sethi, and Jeffery D. Ullman., Addison Wesley, Boston, MA, 2006.

3. Compilers: Principles, Techniques and Tools, Aho, A. V, Sethi, R. and Ullman, J. D. Pearson Education, 1986.

4. D. M. Dhamdhere, “Compiler Construction” (2/e), Macmillan.

5. Cooper &Torczon, “Engineering a Compiler”Elsevier.

6. K C. Louden, “Compiler Construction: Principles and Practice”Cengage.

MODEL QUESTION PAPER

Course Code: **R25CST123**

ADVANCED COMPILER DESIGN

Max. Marks: 60 Duration: 2.5 hours

PART A

(Answer All Questions. Each Question Carries 5 Marks)

1. Explain different implementation methods of the Symbol table.
2. With an example explain the following loop optimization techniques:

a) Code motion.

b) Induction variable elimination and

c) Strength reduction

3. Explain graph colouring global register allocator.
4. Explain reference counting garbage collector. (5)
5. Illustrate the role of peep hole optimization in the compilation process (5)

Part B

(Answer any 5 questions. Each question carries 7 marks)

6. Translate the following statements into three address statements and construct the flow graph.

```
for(i=0;i<n;i++)
for(j=0;j<n;j++)
for(k=0;k<n;k++)
C[i][j]=c[i][j]+a[i][k]*b[k][j]
```

7. Explain the translation of Boolean expressions into three-address code.
8. Illustrate the liveness analysis on the flow graph with suitable example?
9. Explain Dominators and Algorithms for finding Dominators.
10. Explain static allocation and heap allocation strategies. Construct the Directed Acyclic Graph for the basic block given below and simplify the three-address code

d = b *

ce=a+b

b =b *c

a = e –d

11. Illustrate loop invariant code motion with the following code segment.

b= 2

i = 1

L1: if i>100 gotoL3

a = b +1

c =2

If i mode 2 == 0 goto L2

d = c

f= a + i

i = i + 1

goto L1

L2: d = a + d

e = i + d

L3 : end

12. Generate a code sequence for the assignment d=(a-b)+(a-c)+(a-c). Also

CODE	COURSE NAME	CATEGORY	L	T	P	CREDIT
R25CST123	BIOINFORMATICS	PROGRAM ELECTIVE 2	3	0	0	3

Course Objective:

1. To provide foundational knowledge of biomolecules and molecular biology concepts such as DNA, RNA, proteins, genome organization, genetic code, and protein structure that form the basis of bioinformatics.
2. To familiarize students with biological databases and bioinformatics resources, including sequence, structure, and bibliographic databases, and to develop skills in accessing and retrieving biological information effectively.
3. To impart an understanding of sequence alignment, phylogenetic analysis, and computational algorithms for analyzing biological data, including pairwise/multiple sequence alignment, BLAST, FASTA, and phylogenetic tree construction methods.
4. To introduce advanced approaches in protein structure prediction and machine learning applications in bioinformatics, emphasizing protein folding, homology modeling, and the use of AI/ML techniques for sequence and structural analysis.

Course Outcomes: After the completion of the course the student will be able to

Course Outcome		Bloom's Knowledge Level
CO 1	Make use of fundamental concepts of molecular biology to provide computational solutions	K2
CO 2	Utilize bioinformatics tools and databases for retrieving, analysing and understanding biological data	K3
CO 3	Analyze multiple sequences and find conserved regions.	K3
CO 4	Find the relationships between species by constructing phylogenetic tree.	K3

CO 5	Predict unknown protein structures and apply concepts of Machine learning and their applications in Bioinformatics.	K3
CO 6	Design, Develop, Implement and Present innovative ideas on Bioinformatics techniques.	K4

Mapping of Course Outcomes with Program Outcomes

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7
CO 1	3		1	2	1	1	
CO 2	3		1	2	1	3	
CO 3	3		1	2	1	2	
CO 4	3		1	2	1	1	
CO 5	3		1	2	1	1	
CO 6	3	1	1	2	3	2	1

Assessment Pattern

Bloom's Category	End Semester Examination
Apply	70-80%
Analyze	30-40%

Mark Distribution

Total Marks	CIE	ESE	ESE Duration
100	40	60	2.5 Hrs

Continuous Internal Evaluation Pattern:

Evaluation shall only be based on application, analysis or design-based questions (for both internal and end semester examinations).

Continuous Internal Evaluation : 40 marks

- Preparing a review article based on peer reviewed original publications (minimum 10 publications shall be referred) : 15 marks

- Course based task / Seminar/ Data collection and interpretation : 15 marks
- Test paper (1 number) : 10 marks

Course based task/test paper questions shall be useful in the testing of knowledge, skills, comprehension, application, analysis, synthesis, evaluation and understanding of the students.

End Semester Examination Pattern:

The end semester examination will be conducted by the respective College. There will be two parts; Part A and Part B.

Part A will contain 5 numerical/short answer questions with 1 question from each module, having 5 marks for each question. Students should answer all questions. Part B will contain 7 questions (such questions shall be useful in the testing of overall achievement and maturity of the students in a course, through long answer questions relating to theoretical/practical knowledge, derivations, problem solving and quantitative evaluation), with minimum one question from each module of which student should answer any five. Each question can carry 7 marks

Total duration of the examination will be 150 minutes.

Note: The marks obtained for the ESE for an elective course shall not exceed 20% over the average ESE mark % for the core courses. ESE marks awarded to a student for each elective course shall be normalized accordingly.

For example, if the average end semester mark % for a core course is 40, then the maximum eligible mark % for an elective course is 40+20 = 60 %.

Syllabus

Module	Description	Contact Hours
1	Biomolecules-DNA, RNA and proteins-Components and structure. Genome organization. Letter codes for amino acids. Central dogma of molecular Biology. Genetic code. Interatomic forces in proteins, different levels of protein structure, protein domains, motifs.	8
2	Definition and brief history. Bioinformatics vs Computational Biology. Scope and research areas of Bioinformatics. Data archives: Biological Databases-classification and importance; Nucleic acid databases: GenBank, DDBJ, EMB. Protein Sequence Databases: SwissProt, PIR. Derived databases: InterPro, Prosite, Pfam. Structure Databases: RCSB PDB, CATH, SCOP. Bibliographic Databases: PubMed, MEDLINE. Specialized databases. Gateways to archives: Entrez, SRS, ExPASy.	8

3	Concept of sequence alignment, Gaps in alignment, Scoring matrices: PAM and BLOSUM, Alignment of pairs of sequences: Dot Plot, Dynamic Programming, Alignment algorithms: The Needleman and Wunsch algorithm, Smith-Waterman algorithm. Search for homologous sequences using BLAST and FASTA programs. Statistical significance of database searches. Multiple sequence Alignment: Concept and Algorithms in MSA. Tools: Clustal, Mega.	8
4	Concept and its relation to Multiple Sequence Alignment. Representation of phylogeny. Concept of Outgroup. Gene and Species phylogeny. Phylogenetic tree construction methods: Distance based and Character based. Phylogenetic software's: PHYLIP, MrBayes.	8
5	Protein Stability and folding, Ramachandran plot, Homology modelling, Energy minimization, CASP; Protein structure prediction software's: ESyPred3D, Rosetta; PSI- BLAST; Introduction to machine learning techniques: Hidden Markov models, Genetic algorithms and artificial neural networks. Applications of machine learning techniques in multiple sequence alignment, structure prediction and phylogenetic analysis	8

Reference Books

1. Jin Xiong, Essential Bioinformatics, Cambridge University Press, 2006.
2. Brown T. A. Genomes 3, 3rd edition, BIOS Scientific Publishers Limited, 2007.
3. Jeremy M. Berg, John L. Tymoczko, Lubert Stryer, Biochemistry, 5th Edition, W. H. Freeman, 2002.
4. Arthur Lesk, Introduction to Bioinformatics, 5th Edition, Oxford University Press,2019.
5. Dan E. Krane, Fundamental concepts of Bioinformatics, 3rd Edition, Pearson Education India, 2002.
6. David W. Mount, Bioinformatics Sequence and Genome Analysis, 2nd Edition, Cold Spring Press, 2004.
7. Andreas D. Baxevanis, B. F. Francis Ouellette, Bioinformatics: A practical Guide to the Analysis of Genes and Proteins, John Wiley & Sons, 2004
8. Dov Stekel, Microarray Bioinformatics, Cambridge University Press, 2003.

COURSE PLAN

No	Topic	No. of Lectures (40 Hours)
1	Module 1: Molecular Biology	5
1.1	Biomolecules-DNA, RNA and proteins	1
1.2	Components and structure. Genome organization. Letter codes for amino acids.	1
1.3	Central dogma of molecular Biology	1
1.4	Genetic code. Interatomic forces in proteins	1
1.5	Different levels of protein structure, protein domains, motifs.	1
2	Module 2: Bioinformatics	7
2.1	Definition and brief history. Bioinformatics vs Computational Biology, Scope and research areas of Bioinformatics	1
2.2	Data archives: Biological Databases-classification and importance, Nucleic acid databases: GenBank, DDBJ, EMB	1
2.3	Protein Sequence Databases: SwissProt, PIR	1

2.4	Derived databases: InterPro, Prosite, Pfam	1
2.5	Structure Databases: RCSB PDB, CATH, SCOP	1
2.6	Bibliographic Databases: PubMed, MEDLINE	1
2.7	Specialized databases. Gateways to archives: Entrez, SRS, ExPASy.	1
3	Module 3: Sequence Alignment	10
3.1	Concept of sequence alignment, Gaps in alignment	1
3.2	Scoring matrices: PAM and BLOSUM	1
3.3	Alignment of pairs of sequences: Dot Plot, Dynamic Programming	1
3.4	Alignment algorithms: The Needleman and Wunsch algorithm	1
3.5	Smith-Waterman algorithm	1
3.6	Search for homologous sequences using BLAST and FASTA programs	1
3.7	Statistical significance of database searches	1
3.8	Multiple sequence Alignment: Concept	1

3.9	Algorithms in MSA	1
3.10	Tools: Clustal, Mega.	1
4	Module 4: Molecular Phylogenetics	8
4.1	Concept and its relation to Multiple Sequence Alignment	1
4.2	Representation of phylogeny	1
4.3	Concept of Outgroup	1
4.4	Gene and Species phylogeny	1
4.5	Phylogenetic tree construction methods: Distance based (Lecture 1)	1
4.6	Phylogenetic tree construction methods: Distance based (Lecture 2)	1
4.7	Phylogenetic tree construction methods: Character based	1
4.8	Phylogenetic software's: PHYLIP, MrBayes	1
5	Module 5: Protein Structure Prediction and Advanced Bioinformatics	10
5.1	Protein Stability and folding	1
5.2	Ramachandran plot, Homology modelling	1

5.3	Energy minimization	1
5.4	CASP; Protein structure prediction software's: ESyPred3D, Rosetta	1
5.5	PSI-BLAST	1
5.6	Introduction to machine learning techniques: Hidden Markov models	1
5.7	Genetic algorithms	1
5.8	Artificial neural networks	1
5.9	Applications of machine learning techniques in multiple sequence alignment	1
5.10	Structure prediction and phylogenetic analysis	1

MODEL QUESTION PAPER

Course Code: R25CST123

Course Name: Bioinformatics

Max. marks: 60 Duration: 2.5 hours

PART A

Answer All Questions. Each Question Carries 5 Marks

1	Explain the concept of base pairing. If the amount of thymine in genome is 30%, calculate the percentage of cytosine?	5
2	Compare and contrast DDBJ, Genbank.	5
3	Differentiate PAM and BLOSUM series.	5
4	Differentiate between rooted and unrooted phylogenetic tees. How many rooted and unrooted trees are possible for n species?	5
5	Explain the significance of protein folding.	5

Part B

(Answer any five questions. Each question carries 7 marks)

6a	If the sequence of one strand of DNA is written as follows: 5'-ATGCATGCATGCATGCATGCATGCATGC-3' Construct the sequence of complementary strand.	2
b	With the help of a diagram explain the concept of central dogma of molecular Biology.	5
7	Describe the importance of biological databases in bioinformatics. Explain Protein sequence databases.	7
8	Define scoring matrices? Explain how PAM is derived?	7
9	Using Smith Waterman method, construct the partial alignment scoring table and obtain the optimal local alignment of the following two sequences: _ ACGTATCGCGTATA GATGCTCTCGGAJAA	7
10	Illustrate the concept of a phylogenetic tree? Explain the steps of UPGMA method for phylogenetic tree construction with an example.	7
11	Explain the significance of Hidden Markov Model in bioinformatics. Discuss the advantages and disadvantages of using HMMs	7
12	Explain any two methods for Protein Secondary Structure Prediction.	7