Dr. N.G.P.ARTS AND SCIENCE COLLEGE (Autonomous)

REGULATIONS 2019-20 for Post Graduate Programme
(Outcome Based Education model with Choice Based Credit System)

M.Sc. Computer Science with Data Analytics

(For the students admitted during the academic year 2020-21 and onwards)

Programme:M.Sc. Computer Science with Data Analytics

Eligibility

A candidate who has passed in B.Sc. Computer Science / B.C.A. /B.Sc. Computer Technology / B.Sc. Information Technology /B.Sc. nformation Sciences/B.Sc. Information Systems/ B.Sc. Software Systems/B.Sc. Software Sciences/B.Sc. Applied Sciences (Computer Science/Computer Technology) / B.Sc. Electronics of any University in Tamil Nadu or an Examination accepted as equivalent thereto by the Academic Council, subject to such conditions as may be prescribed thereto are permitted to appear and qualify for the **Master of Computer Science with Data Analytics Degree Examination** of this College after a programme of study of two academic years.

Programme Educational Objectives

The Curriculum is designed to attain the following learning goals which students shall accomplish by the time of their graduation:

1. To secure future developments in the field of Computer Science.
2. To gain agility in the advanced programming languages and build software for big data analytics.
3. To show professional competence in the areas of Artificial Intelligence and Machine Learning.
4. To develop and equip the employability skills to meet the requirement of the Information Technology sector.
5. To engage in research oriented fields and life-long learning to adapt with the continuously evolving technology.

PROGRAMME OUTCOMES:

On the successful completion of the program, the following are the expected outcomes.

PO Number	PO Statement
PO1	Ability to apply knowledge of Computer Science, Mathematics and Statistics to solve problems
PO2	Ability to model, analyze, design, visualize and realize physical systems or processes of increasing size and complexity
PO3	An ability to select appropriate methods and tools for data analysis in specific organizational contexts
PO4	An ability to analyze very large data sets in the context of real world problems and interpret results
PO5	Ability to work in a team and understand professional and societal responsibilities

Guidelines for Programmes offering Part I& Part II for Two Semesters:

Part	Subjects	No.of Papers	Credit	Semester No.
III	Core	7	5x4=20 2x2=4	I
	Core	6	4x4=16 2x2=4	II
	DSE-I	1	1x4=4	II
	Core	6	4x4=16 2x2=4	III
	DSE-II	1	1x4=4	III
	Project Work	2	2+16=18	III &IV
TOTAL CREDITS			**90**	

CURRICULUM

M.Sc. COMPUTER SCIENCE WITH DATA ANALYTICS

Course Code	Course Category	Course Name	L	T	P	Exam (h)	Max Marks			Credits
							CIA	ESE	Total	
First Semester										
204DA2A1CA	Core - I	Foundations of Data Science and Python Programming	5	-	-	3	25	75	100	4
204DA2A1CE	Core – II	Probability and Statistics	5	-	-	3	25	75	100	4
204DA2A1CB	Core - III	Data Structures andAlgorithm Design	4	-	-	3	25	75	100	4
204DA2A1CC	Core - IV	Advanced Operating Systems	4	-	-	3	25	75	100	4
204DA2A1CD	Core-V	Advanced Java Programming	4	-	-	3	25	75	100	4
204DA2A1CP	Core Practical-I	Advanced Java Programming	-	-	4	3	40	60	100	2
204DA2A1CQ	Core Practical-II	Python Programming	-	-	4	3	40	60	100	2
Total			**22**	**-**	**8**	**-**	**-**	**-**	**700**	**24**

BoS Chairman/HoD
Department of Computer Science with Data Analytics
Dr. N. G. P. Arts and Science College
Coimbatore – 641 048

Academic Council
Dr.V. Rajendran
CHAIRMAN
Dr. N.G.P. Arts and Science College

Course Code	Course Category	Course Name	L	T	P	Exam (h)	Max Marks			Credits
							CIA	ESE	Total	
Second Semester										
204DA2A2CA	Core - VI	Artificial Intelligence	5	-	-	3	25	75	100	4
204DA2A2CB	Core – VII	Data Mining and Business Intelligence	5	-	-	3	25	75	100	4
204DA2A2CC	Core-VIII	Data Base Administration and Management	4	-	-	3	25	75	100	4
204DA2A2CD	Core-IX	Information and Network Security	4	-	-	3	25	75	100	4
204DA2A2CP	Core Practical-III	Data Mining	-	-	4	3	40	60	100	2
204DA2A2CQ	Core Practical-IV	R Programming	-	-	4	3	40	60	100	2
204DA2A2DA	DSE-I	Advanced Statistics	4	-	-	3	25	75	100	4
204DA2A2DB		Image and Video Analytics								
204DA2A2DC		Web Intelligence								
Total			**22**	**-**	**8**	-	-	-	**700**	**24**

Course Code	Course Category	Course Name	L	T	P	Exam (h)	Max Marks			Credits
							CIA	ESE	Total	
Third Semester										
204DA2A3CA	Core –X	Wireless Networks	4	-	-	3	25	75	100	4
204DA2A3CB	Core - XI	Cloud Computing	4	-	-	3	25	75	100	4
204DA2A3CC	Core –XII	Software Project Management	4	-	-	3	25	75	100	4
204DA2A3CD	Core -XIII	Machine Learning	4	-	-	3	25	75	100	4
204DA2A3CV	Core- XIV	Project Work-I	-	-	2	3	40	60	100	2
204DA2A3CP	Core Practical-IV	Big Data Analytics	-	-	4	3	40	60	100	2
204DA2A3CQ	Core Practical-V	Machine Learning	-	-	4	3	40	60	100	2
204DA2A3DA	DSE – II	Principles of Deep Learning	4	-	-	3	25	75	100	4
204DA2A3DB		Predictive Data Analytics								
204DA2A3DC		Health Care Data Analytics								
Total			**20**	**-**	**10**	**-**	**-**	**-**	**800**	**26**

Course Code	Course Category	Course Name	L	T	P	Exam (h)	Max Marks			Credits
							CIA	ESE	Total	
Fourth Semester										
204DA2A4CV	Core - XV	Project Work-II	-	-	30	3	50	150	200	16
		Total	-	-	30	-	-	-	200	16
						Grand Total			**2400**	**90**

DISCIPLINE SPECIFIC ELECTIVE

Students shall select the desired course of their choice in the listed elective course during Semesters V & VI

Semester II (Elective I)

List of Elective Courses

S. No.	Course Code	Name of the Course
1.	204DA2A2DA	Advanced Statistics
2.	204DA2A2DB	Image and Video Analytics
3.	204DA2A2DC	Web Intelligence

Semester III (Elective II)

List of Elective Courses

S. No.	Course Code	Name of the Course
1.	204DA2A3DA	Principles of Deep Learning
2.	204DA2A3DB	Predictive Data Analytics
3.	204DA2A3DC	Health Care Data Analytics

EXTRA CREDIT COURSES

The following are the courses offered under self study to earn extra credits:

S. No.	Course Code	Course Name
1	204DA2ASSA	Professional Ethics
2	204DA2ASSB	Human Computer Interactions

Regulation
(2019-2020)

PG Programme

Effective from the academic year 2019-20 and applicable to the students admitted to the Degree of Master of Arts/Commerce/Management/Science.

1. NOMENCLATURE

1.1 Faculty: Refers to a group of programmes concerned with a major division of knowledge. Eg. Faculty of Computer Science consists of Programmes like Computer Science, Information Technology, Computer Technology, Computer Applications etc.

1.2 Programme: Refers to the Master of Arts/Management/Commerce/Science Stream that a student has chosen for study.

1.3 Batch: Refers to the starting and completion year of a programme of study. Eg. Batch of 2015–2017 refers to students belonging to a 2-year Degree programme admitted in 2015 and completing in 2017.

1.4 Course: Refers to a component (a paper) of a programme. A course may be designed to involve lectures / tutorials / laboratory work / seminar / project work/ practical training / report writing / Viva voce, etc or a combination of these, to effectively meet the teaching and learning needs and the credits may be assigned suitably.

a) **Core Courses**

   A course, which should compulsorily be studied by a candidate as a core requirement is termed as a Core course.

b) **Extra Departmental Course (EDC)**

   A course chosen generally from a related discipline/subject, with an intention to seek exposure in the discipline relating to the core domain of the student.

c) **Discipline Specific Elective Course (DSE)**: DSE courses are the courses offered by the respective disciplinary/ interdisciplinary programme.

d) Project Work:

It is considered as a special course involving application of knowledge in problem solving/analyzing/exploring a real-life situation. The Project work will be given in lieu of a Core paper.

e) Extra credits

Extra credits will be awarded to a student for achievements in co-curricular activities carried out outside the regular class hours. The guidelines for the award of extra credits are given in section two, these credits are not mandatory for completing the programme.

e) Advanced Learner Course (ALC):

ALC is doing work of a higher standard than usual for students at that stage in their education. Research work carried out in University/ Research Institutions/ Industries of repute in India or abroad for a period of 15 to 30 days.

2. EXTRA CREDITS

- Earning extra credit is mandatory. However, it is not essential for programme completion.
- Extra Credits will be awarded to a student for achievement in co-curricular/ extracurricular activities carried other than the regular class-hours.
- A student is permitted to earn a maximum of 10 extra Credits during the programme duration of PG from I to IV Semester.
- Candidate can claim a maximum of 1 credit under each category listed.

The following are the guidelines for the award of Extra credits:

2.1 Proficiency in Foreign Language

Qualification	Credit
A pass in any foreign language in the examination conducted by an authorized agency	1

2.2 Proficiency in Hindi

Qualification	Credit
A pass in the Hindi examination conducted by Dakshin Bharat Hindi Prachar Sabha	1

Examination passed during the programme period only will be considered for extra credit

2.3 Self-study Course

Qualification	Credit
A pass in the self-study courses offered by the department	1

The candidate should register in the self-study course offered by the department only in the III semester

2.4 Typewriting/Short hand

A Pass in shorthand /typewriting examination conducted by Tamil Nadu Department of Technical Education (TNDTE) and the credit will be awarded.

Qualification	Credit
A pass in the type writing /short hand examination offered by TNDTE	1

2.5 Diploma / Certificate

Courses offered by any recognized University / NCVRT

Qualification	Credit
A pass in any Certificate /Diploma/PG Diploma Course	1

2.6 CA /ICSI/ CMA

Qualification	Credit
Qualifying foundation/Inter level/Final in CA/ICSI/CMA etc.	1

2.7 Sports and Games

The Student can earn extra credit based on their achievement in sports as given below:

Qualification	Credits
Achievement in University/State /National/ International	1

2.8 Online Courses

Pass in any one of the online courses

Qualification	Credit
SWAYAM/NPTEL/Spoken Tutorial etc.,	1

2.9 Publications / Conference Presentations (Oral/ Poster) /Awards

Qualification	Credit
Research Publications in Journals/oral/poster presentation in Conference	1

2.10 Innovation / Incubation / Patent / Sponsored Projects / Consultancy

Qualification	Credit
Development of model/ Products/ Prototype/ Process/App/Registration of Patents/ Copyrights/ Trademarks/Sponsored Projects/Consultancy	1

2.11 Representation

Qualification	Credit
Participation in State / National level celebrations such as Independence day, Republic day Parade, National Integration camp etc.,	1

3. EXAMINATIONS

The following are the distribution of marks for External and Internal i.e., Comprehensive examination and Continuous Internal Assessment and passing minimum marks for theory papers of PG programmes.

TOTAL MARKS	EXTERNAL		Internal Max. marks	Overall Passing Minimum for total marks (Internal + External)
	Max. marks	Passing Minimum for External alone		
100	75	38	25	50
50	50	25	----	25

The following are the Distribution of marks for the Continuous Internal Assessment in the theory papers of PG programmes.

S. No.	For Theory– PG courses	Distribution of Marks
1	TESTS I (2 hours)	5
2	TESTS II / End semester Model test (3 hours)	10
3	OBE- Rubrics	10
	TOTAL MARKS	**25**

The following are the distribution of marks for the External Assessment in PG Theory courses

S. No.	For Theory– PG courses	Distribution of Marks	
1	Comprehensive (Written) Examination	65	50
2	Online MCQ Examination	10	--
	TOTAL MARKS	**75**	**50**

The following are the distribution of marks for External examinations (CE) and Continuous Internal Assessment (CIA) and passing minimum marks for the practical courses of PG programmes.

TOTAL MARKS	EXTERNAL		Internal Max. marks	Overall Passing Minimum for total marks (Internal + External)
	Max. marks	Passing Minimum for External alone		
100	60	30	40	50
200	120	60	80	100

The following are the distribution of marks for the Continuous Internal Assessment (CIA) in PG practical courses

S. No.	For Theory – PG Practical courses	Distribution of Marks	
1	Tests: Two tests out of which one shall be during the mid semester and the other to be conducted as model test at the end of the semester.)	24	48
2	OBE- Rubrics	16	32
	TOTAL MARKS	**40**	**80**

The following are the distribution of marks for the External Assessment in PG practical courses

S. No.	For Theory – PG Practical courses	Distribution of Marks	
1	Experiment-I	25	50
2	Experiment-II	25	50
3	Record & Viva-Voce	10	20
	TOTAL MARKS	**60**	**120**

The following are the distribution of marks for Project and Viva voce examinations/Industrial Training and Continuous Internal Assessments and passing minimum marks for the project courses/Industrial Training of PG programmes

TOTAL MARKS	EXTERNAL		Internal Max. marks	Overall Passing Minimum for total marks (Internal + External)
	Max. marks	Passing Minimum for External alone		
100	60	30	40	50
200	120	60	80	100

The following are the distribution of marks for the Continuous Internal Assessment in PG Project/ Industrial Training courses.

S. No.	For- PG Project courses/ Industrial Training	Distribution of Marks	
1	Review-I	10	20
2	Review-II	10	20
3	Review-III	10	20
4	Documentation	10	20
	TOTAL MARKS	**40**	**80**

The following are the distribution of marks for the External Examination (CE) in PG Project /Industrial Training courses

S. No.	For- PG Project courses/ Industrial Training Courses	Distribution of Marks	
1	Record Work and Presentation	40	80
2	Viva-Voce	20	40
	TOTAL MARKS	**60**	**120**

- The end semester examinations shall normally be conducted after completing 90 working days for each semester.

- The maximum marks for each theory and practical course (including the project work and Viva-Voce examination in the final Semester) shall be 100 with the following breakup.

(i) **Theory Courses**

Continuous Internal Assessment (CIA) : 25 Marks

End Semester Exams (ESE) : 75 Marks

(Online Exam: 10 Marks &Written Exam: 65 Marks)

(ii) **For Practical Courses**

Continuous Internal Assessment (CIA) : 40 Marks

End Semester Exams (ESE) : 60 Marks

Continuous Assessment OBE Rubrics Score Sheet

Degree: _____ Branch: _____ Semester: _____

Course Code: _____ Course: _____

Max. Marks: Internal:_____ External: _____ Total: _____

S. No.	REG. NO.	THEORY / PRACTICAL & LIBRARY CLASS PARTICIPATION (15) (Compulsory)				RUBRICS ASSESSMENT (SELECT ANY ONE)									Total Marks out of : 30	Total Marks out of : 16 / 10 / 08 / 04
						PAPERS / REPORTS (15)			ASSIGNMENTS (15)			CLASS PRESENTATION (15)				
		Library	Integration of Knowledge	Interaction & Participation	Demonstration of Knowledge	Organization & Knowledge	Format & Spelling	Reference / Experiments	Demonstration of Knowledge	Format & Spelling	Reference	Content & Coherence	Creativity and Speaking Skills	Duration of Presentation		
		6	3	3	3	5	5	5	5	5	5	5	5	5		
1																

a) Utilization of Library

Marks will be awarded to the student based on the hours spent in the library after the working hours and submission of report by the student.

Hours spent in Library	Marks	Type of Document submitted
2	1	Report/ Assignment/ Class presentation
4	2	
6	3	
8	4	
10	5	
12	6	

- During the Library hour, the student must spend time in reading the articles, books, journals of their subject of interest
- Each student should borrow minimum three books during the semester

b) Class Participation

Active participation in classroom discussion by the student will be evaluated based on Integration of knowledge, Interaction and Participation and demonstration of knowledge.

c) Papers / Reports/ Assignments/ Class Presentation

The student will be evaluated based on his ability to do analysis of application of theory to real world problems or creative extension of class room learning and his/her ability to communicate the given topic effectively and clearly. The following are the distribution of marks for the continuous internal assessment in PG practical courses

4. FOR PROGRAMME COMPLETION

Programme Completion (for students admitted during the A.Y.2019-20 and Onwards)

Student has to complete the following:

i) Core, EDC, DSE, Project as mentioned in the scheme
ii) Internship / Industrial/ Institutional training as mentioned in the scheme

Students must undertake industrial / institutional training for a minimum of 15 days and not exceeding 30 days during the II semester summer vacation. The students will submit the report for evaluation during III semester.

Based on the performance Grade will be awarded as follows:

Marks Scored	**Grade to be awarded**
75 and above	A
60-74	B
50-59	C
< 50	Re-Appearance

Course Code	Course Name	Category	L	T	P	Credit
204DA2A1CA	FOUNDATIONS OF DATA SCIENCE AND PYTHON PROGRAMMING	CORE	5	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- Concepts of Data Science
- Understand about Python Programming
- Web Scraping and Visualization in Python

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Understand the principles of data science	K2
CO2	Understand the techniques for Data Handling	K2
CO3	Apply the concepts of Python for Data Aggregation and Wrangling	K3
CO4	Apply the numerical concepts using Numpy	K3
CO5	Create the visualization concepts in Python	K6

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	S	S	L
CO2	S	M	S	S	L
CO3	M	S	S	S	L
CO4	S	M	S	S	L
CO5	S	S	S	S	M

S Strong **M** Medium **L** Low

204DA2A1CA	FOUNDATIONS OF DATA SCIENCE AND PYTHON PROGRAMMING	SEMESTER I

Total Credits: 4

Total Instruction Hours: 60 h

Syllabus

Unit I Data Science 10 h

Introduction : The current landscape of Data Science- Data science process – Roles - Stages in data science project Data Science Components–Working with data: Retrieving data – Data preparation – Data exploration – Data modeling – Presentation

Unit II Handling Data 10 h

Problems when handling large data – General techniques for handling large data – Case study – Steps in big data – Distributing data storage and processing with Frameworks

Unit III Python: Data Aggregation and Wrangling 14 h

Introduction : GroupBy Mechanics – Data Aggregation – Groupwise Operations and Transformations – Pivot Tables and Cross Tabulations – Date and Time Date Type tools – Time Series Basics – Data Ranges - Frequencies and Shifting Combining and Merging DataSets – Reshaping and Pivoting – Data Transformation – String Manipulation, Regular Expressions

Unit IV Python Numpy 12 h

Introduction to NumPy - Understanding the N - dimensional data structure - Creating NumPy arrays - Basic operations and manipulations on N-dimensional arrays - Indexing and Slicing-Advanced Indexing –Mathematical Functions – Statistical Functions – Search,Sorting and Counting Functions –Matrix Library

Unit V Web Scraping and Visualization 14 h

Introduction: Data Acquisition by Scraping web applications –Submitting a form - Fetching web pages – Downloading web pages through form submission – CSS Selectors. Visualization: Visualization In Python :Matplotlib package – Plotting Graphs – Controlling Graph – Adding Text – More Graph Types – Getting and setting values

Text Books

1. Wes Mc Kinney,(2012), "Python for Data Analysis",(1st Edn.), O'Reilly.
2. Mark Lutz, (2010), "Programming Python", (4th Edn.), O'Reilly.

References

1. John V Guttag,(2016), "Introduction to Computation and Programming Using Python",(2nd Edn.), MIT press.
2. Timothy A. Budd, (2011), "Exploring Python, (1st Edn.), Mc-Graw Hill.

Course Code	Course Name	Category	L	T	P	Credit
204DA2A1CE	PROBABILITY AND STATISTICS	CORE	5	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- Statistical tools for analyzing data
- Concepts of various Probability and Statistics techniques
- Hypothesis Testing and ANOVA techniques

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Explain theory of application behind the descriptive statistics	K2
CO2	Demonstrate the properties of random variable and expectation	K3
CO3	Analyse various distributions and their usage	K4
CO4	Apply concepts in testing of hypothesis	K4
CO5	Interpret real world problems into statistical models	K6

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	M	M	M	L
CO2	S	M	M	M	L
CO3	S	S	M	M	L
CO4	S	M	M	M	L
CO5	S	S	S	S	S

S Strong **M** Medium **L** Low

204DA2A1CE	PROBABILITY AND STATISTICS	SEMESTER I

Total Credits: 4
Total Instruction Hours: 60 h

Syllabus

Unit I Descriptive Statistics 14 h

Measures of central value- Arithmetic mean- Median and Mode- Measures of dispersion – range- quartile deviation- standard deviation and coefficient of variation for grouped and ungrouped data- Skewness – Karl Pearson and Bowley's measure of skewness. Correlation analysis- Spearman's rank correlation coefficient- Regression analysis – Simple linear regression

Unit II Probability and Expected Value 12 h

Definitions of probability- Importance of the concepts of probability- Calculation of probability- Addition and Multiplication rules of probability- Conditional probability- Baye's theorem- Mathematical Expectation- Random variables and probability distribution.

Unit III Probability Distributions 12 h

Probability distributions – Discrete Distribution -Binomial- Poisson Distribution – Continuous distribution- Normal distributions- Sampling distribution and standard error- Universe distributions- Sample distribution-Utility the concepts of standard error

Unit IV Testing of Hypothesis 12 h

Introduction –Procedure of hypothesis testing- Type I and Type II errors- One and two tailed tests- Tests of significance for large samples- Difference between small and large samples- Two tailed test for difference between of two samples- Standard error of the difference between two standard deviations- Test of significance for small samples- Assumption of normality- Students t-distribution- Properties of t-distribution- Application of the t-distribution

Unit V F-Test and Analysis of Variance 10 h

F-test or variance ratio test- Application of F-test- Analysis of variance- Assumptions in analysis of variance - Technique of analysis of variance - Analysis of variance in one way classification model – Analysis of variance in two way classification model

Note: Theory 20% and Problem 80%

Text Book

1 Gupta, S.P, 2017, "Statistical Methods", 16th Edition, Sultan Chand and Sons, New Delhi.

References

1 Ronald E. Walpole, 2018, "Probability & Statistics", 9th Edition, Pearson education, South Asia

2 Sheldon M.Ross, 2017, "Introductory Statistics", 4th Edition, Academic Press, United States

3 Vijay K. Rohatgi A.K, MD. Ehsanes Saleh, 2015, "An Introduction to Probability and Statistics", 3rd Edition, John Wiley and Sons, New Delhi

4 Sheldon Ross, 2017, "A First Course in Probability", 5th Edition, PHI, New Jersey

Course Code	Course Name	Category	L	T	P	Credit
204DA2A1CB	DATA STRUCTURES AND ALGORITHM DESIGN	CORE	4	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- Algorithms in various domains
- Techniques for designing efficient algorithms
- The concepts of Trees and Graphs

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Analyze the time and space complexity of algorithms	K4
CO2	Apply Divide and Conquer techniques for solving problems	K3
CO3	Apply the Dynamic Programming Techniques for solving the Shortest Path Problems	K3
CO4	Understand the Concepts of trees and their techniques	K2
CO5	Apply graphs traversals to solve the Shortest Path Problems	K3

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	M	S	L
CO2	S	S	M	S	L
CO3	S	S	M	S	L
CO4	S	M	M	S	L
CO5	S	S	M	S	M

S Strong M Medium L Low

204DA2A1CB	DATA STRUCTURES AND ALGORITHM DESIGN	SEMESTER I

Total Credits: 4

Total Instruction Hours: 48 h

Syllabus

Unit I Algorithm 09 h

Introduction: Definition - Structure and Properties of algorithms – Development of an algorithm – Data Structures and algorithms – Data Structure definition and classification - Analysis of algorithms: Efficiency of algorithms – Apriori analysis – Asymptotic notations –Time and Space complexity of an algorithm using O notation

Unit II Divide and Conquer 10 h

Introduction: Integer Multiplication – Strassen's Matrix Multiplication – Closest Pair, The Greedy Method: Minimum Cost Spanning Trees - Optimal Storage On Tapes – Optimal Merge Patterns - Single Source Shortest Paths

Unit III Dynamic Programming 09 h

Introduction: The General Method – Multistage Graphs – All-Pairs Shortest Paths – Single-Source Shortest Paths - 0/1 Knapsack - Reliability Design - The Traveling Salesperson Problem

Unit IV Trees 10 h

Trees: Introduction - Binary Trees - Binary Search Trees : Searching – Insertion and Deletion of elements - Height Balancing Techniques-AVL Trees: Height – searching –Insertion and deletion of elements- AVL rotations -Splay Trees-Multiway Search Trees – B Trees

Unit V Graphs 10 h

Graphs- Graphs terminology, Graph Representations, graph traversals-search methods-Depth First Traversal -Breadth First Traversal - Applications of Graphs-Minimum cost spanning tree using Kruskal's algorithm, Dijkstra's algorithm for Single Source Shortest Path Problem

Text Books

1. Thomas H. Cormen, Charles E. Leiserson and Ronald L. Rivest,(2009), "Introduction to Algorithms",(3rd Edn.), MIT Press.
2. Mark Allen Weiss,(2013), "Data Structures and Algorithm Analysis in C++",(4th Edn.), Addison-Wesley.

References

1. Robert L. Kruse and Clovis L.Tondo, (2007),"Data Structures and Program design in C",(2nd Edn.), Pearson Education.
2. Michael T. Goodrich, Roberto Tamassia,"(2001),Algorithm Design, Foundations, Analysis, and Internet Examples",(1st Edn.), Wiley.
3. SahniSartaj,(2004),"Data Structures, Algorithms and Applications in C++", (2nd Edn.),Silicon Press.

Course Code	Course Name	Category	L	T	P	Credit
204DA2A1CC	ADVANCED OPERATING SYSTEMS	CORE	4	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- Advanced Concepts in operating Systems
- The architecture of operating system
- The components of distributed systems

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Understand the structure and functions of operating Systems.	K2
CO2	Analyze deadlock situations, recovery mechanisms and memory management.	K4
CO3	Understand about distributed operating systems and its Issues.	K2
CO4	Understand the Concepts of Process synchronization and Scheduling in Distributed System	K2
CO5	Evaluate the Real time operating systems and its functionality	K5

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	M	S	L
CO2	S	S	M	S	L
CO3	S	S	M	S	L
CO4	S	M	M	S	L
CO5	S	S	M	S	M

S Strong M Medium L Low

204DA2A1CC	ADVANCED OPERATING SYSTEMS	SEMESTER I

Total Credits: 4

Total Instruction Hours: 48 h

Syllabus

Unit I Operating System 09 h

Overview - Introduction –Operating System Structure – Process Management – Memory Management – Storage Management – Distributed System – Operating System Services - User Operating System Interface – System Calls – System Programs - Operating System Design and Implementation – Operating System Structures

Unit II Deadlock and Memory Management 10 h

Introduction: Methods of Handling Deadlocks - Deadlock Prevention - Deadlock Avoidance -Deadlock Detection and Recovery – Memory Management: Swapping - Contiguous Memory Allocation- Segmentation- Demand Paging- Page Replacement Algorithms- Allocation of Frames - Thrashing - Memory Mapped Files

Unit III Distributed Systems 10 h

Introduction to Distributed Systems- Goals of Distributed Systems - Hardware Concepts - Software Concepts- Design issues of Distributed Systems - Communication in Distributed Systems - The Client Server Model- Remote Procedure Call - Group Communication – Remote Procedure Call - Remote Procedure call Protocols

Unit IV Process Scheduling and Synchronization 10 h

Processes and Processors in Distributed Systems: Threads- System Models Processor Allocation- and Scheduling in Distributed Systems: Fault Tolerance - Synchronization in Distributed Systems – Clock Synchronization- Mutual Exclusion- Election Algorithms- Atomic Transactions – Deadlock in Distributes Systems

Unit V Real Time Systems 09 h

Introduction :Introduction – System Characteristics – Features of Real Time Kernels – Implementation of Real Time Operating Systems - Real Time CPU Scheduling Case Studies :Linux Systems - Design Principles – Kernel Modules – Process Management – Scheduling – Inter Process Communications

Text Books

1. Abraham Silberschatz, Peter Baer Galvin, Greg Gagne, (2006),"Operating System Principles", (7th Edn.), John Wiley & Sons Inc.
2. AndrewS.Tanenbaum,(1995),"Distributed Operating Systems",(1st Edn.),Pearson

References

1. Mukesh Singhal and Niranjan G. Shivaratri, (2011),"Advanced Concepts in Operating Systems – Distributed Database- and Multiprocessor Operating Systems",(2nd Edn.), Tata McGraw-Hill.
2. Rajib Mall, (2008), "Real-Time Systems: Theory and Practice", (2nd Edn.), Pearson Education India.

Course Code	Course Name	Category	L	T	P	Credit
204DA2A1CD	ADVANCED JAVA PROGRAMMING	CORE	4	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- Advance Java concepts to develop applications
- Database Connectivity using JDBC
- The concepts of Applets and Servlets

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Understand about Java beans and swing	K2
CO2	Develop and apply event in applets	K3
CO3	Understand the life cycle of Java Servlet	K2
CO4	Learn the architecture and design of Enterprise Java Bean	K1
CO5	Establishing Database Connectivity using JDBC	K3

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	M	S	M	L
CO2	M	S	M	L	M
CO3	S	M	M	L	S
CO4	S	M	L	M	S
CO5	M	M	M	L	S

S Strong M Medium L Low

204DA2A1CD	ADVANCED JAVA PROGRAMMING	SEMESTER I

Total Credits: 4

Total Instruction Hours: 48 h

Syllabus

Unit I Java Beans and Swings 09 h

Introduction: Advantages – Design patterns for Properties – Events – Methods and Design Patterns - Java Beans API – Swing : Introduction – Swing Is Built on the AWT - Two Key features of Swing – MVC Connections – Components and Containers – The Swing Packages – Simple Swing Applications - Exploring Swing

Unit II Applet 10 h

Introduction:– Applet Basics – Types of Applet - Applet Architecture – An Applet Skeleton – Simple Applet Display Methods – Requesting and Painting – Using the status window – The HTML Applet tag – Passing parameters to applet – getDocumentBase() and getcodeBase() – Applet context and showDocument()

Unit III Java Servlet 10 h

Introduction :Background - The life cycle of a Servlet – Using Tomcat for Servlet development – A Simple Servlet – The Javax.Servlet Packages – Reading Servlet Parameters – The javax.servlet.http packages – Handling Http request and responses – cookies - Session Tracking – Applying applets and Servlets

Unit IV Enterprise Java Bean 10 h

Introduction: EJB Architecture and Design – EJB component Model – Roles ,Relationships and Responsibilities – Enterprise Java Bean – Understanding EJB Container Functionality – Integrating with CORBA – Session Bean :Writing a Session in EJB – Implementing the session Synchronization interface - Choosing the stateless and Stateful Beans

Unit V Database Connectivity 09 h

Introduction : JDBC Driver types – Creating the First JDBC Programs – Perform Batch Updates – Using Save Points – Configuring the JDBC-ODBC Bridge – Explaining the Database connection Pools and Data Sources – Revisiting the DB Processor – Using the Rowset Interface – Understanding the J2EE connector Architecture

Text Books

1. Herbert Schildt, (2002), "Java The Complete Reference", (5th Edn.), Tata McGraw Hill.
2. James McGovern, Jason Gordan, Ethan Henry, (2003), "Java 2 Enterprise Edition 1.4 Bible", (2nd Edn.), Wiley publishing.

References

1. Herbert Schildt, (2018), "Java, A Beginner Guide", (8th Edn.),Oracle Press..
2. Bert Bates, KarthySierra, Eric Freeman, Elisabeth Robson, (2009), "Head First Design Patterns", (1st Edn.), O'Reilly.

204DA2A1CP	CORE PRACTICAL: ADVANCED JAVA PROGRAMMING	SEMESTER I

Total Credits: 2
Total Instructions Hours: 48 h

S.No	List of Experiments
1	Programs using Java control statements
2	Programs using simple applets
3	To create applets incorporating features such as images, shapes, background and foreground color
4	Create an application using simple GUI
5	Program to Perform some applications using Java Bean
6	Create an application using Swing
7	Program to perform some operations using Threads
8	Program to perform Session Tracking
9	Java servlet program to implement sendredirect () Method (using Http servlet class).
10	Servlet Program using HTTP Servlet
11	Program to perform client server interface using RMI
12	JDBC to interact with database.

Note: Ten Programs are Mandatory

204DA2A1CQ	CORE PRACTICAL : PYTHON PROGRAMMING	SEMESTER I

Total Credits: 2
Total Instructions Hours: 48 h

S.No	List of Experiments
1	Programs to test simple Python statements
2	Program to Implement a sequential search
3	Program to Explore string functions
4	Program to Read and Write into a file
5	Program to Demonstrate use of List
6	Program to Demonstrate use of Dictionaries
7	Program to Demonstrate use of Tuples
8	Program to Create Comma Separate Files (CSV), Load CSV files into internal Data
9	Program using Pandas: Extract items at given positions from a series
10	Program using Pandas: Filter valid emails from a series
11	Program using Pandas: Find the outliers from a series or data frame column
12	Program to Perform Analysis for given data set using Pandas

Note: Ten Programs are mandatory

Course Code	Course Category	Course Name	L	T	P	Exam (h)	Max Marks			Credits
							CIA	ESE	Total	
Second Semester										
204DA2A2CA	Core - VI	Artificial Intelligence	5	-	-	3	25	75	100	4
204DA2A2CB	Core – VII	Data Mining and Business Intelligence	5	-	-	3	25	75	100	4
204DA2A2CC	Core-VIII	Data Base Administration and Management	4	-	-	3	25	75	100	4
204DA2A2CD	Core-IX	Information and Network Security	4	-	-	3	25	75	100	4
204DA2A2CP	Core Practical-III	Data Mining	-	-	4	3	40	60	100	2
204DA2A2CQ	Core Practical-IV	R Programming	-	-	4	3	40	60	100	2
204DA2A2EA	DSE-I	Advanced Statistics	4	-	-	3	25	75	100	4
204DA2A2EB		Image and Video Analytics								
204DA2A2EC		Web Intelligence								
Total			**22**	**-**	**8**	**-**	**-**	**-**	**700**	**24**

BoS Chairman/HoD
Department of Computer Science with Data Analytics
Dr. N. G. P. Arts and Science College
Coimbatore – 641 048

Academic Council
Dr.V. Rajendran
CHAIRMAN
Dr.N.G.P. Arts and Science College

Dr.NGPASC
COIMBATORE | INDIA

M.Sc.Computer Science with Data Analytics (Students admitted during the AY 2020-21)

Course Code	Course Name	Category	L	T	P	Credit
204DA2A2CA	ARTIFICIAL INTELLIGENCE	CORE	5	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- Concepts in Artificial Intelligence
- AI approaches towards problem solving, knowledge representation, and learning
- The applications of AI

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Understand the foundations of artificial intelligence	K2
CO2	Model problems by applying suitable search methods	K4
CO3	Construct techniques for Constraint Satisfaction problem and Adversarial search	K3
CO4	Explain knowledge representation and reasoning	K3
CO5	Discuss the major approaches in planning	K5

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	M	S	M
CO2	S	S	M	S	M
CO3	S	M	M	S	M
CO4	S	S	M	M	M
CO5	S	M	M	S	M

S Strong M Medium L Low

204DA2A2CA	ARTIFICIAL INTELLIGENCE	SEMESTER II

Total Credits: 4

Total Instruction Hours: 60 h

Syllabus

Unit I Artificial Intelligence 12 h

Foundations of Artificial Intelligence - Intelligent Agents: Agents and Environments- Structure of Agents – Problem Solving: Problem Solving Agents- Problem Formulation - Uninformed Search Strategies: Breadth-First Search – Depth-First Search – Depth-Limited Search – Iterative Deepening Depth-First Search – Bidirectional Search - Comparing Uninformed Search Strategies

Unit II Heuristic and Classical Search 12 h

Greedy Best-First Search – A * Search – Memory-Bounded Heuristic Search – Heuristic Functions – Local Search Algorithms – Hill Climbing Search – Simulated Annealing –Local Beam Search – Genetic Algorithms – Online Search Agents and Unknown Environments: Online Search Problems - Online Search Agents - Online Local Search –Learning in Online Search

Unit III Constraint Satisfaction Problems 12 h

Introduction : Constraint Satisfaction Problems(CSP): Backtracking search for CSP- Local Search for CSP - Structure of Problems - Adversarial Search: Introduction – Games – Optimal Decision in Games - The Min Max Algorithm – Alpha–Beta Pruning – Games that Include an Element of Chance: Card Games

Unit IV Knowledge Representation and Reasoning 12 h

Knowledge Based Agents – Logic – Propositional Logic: Syntax – Semantics – A simple knowledge base – Inference – Equivalence, Validity and Satisfiability – Reasoning Patterns in Propositional Logic: Resolution – Forward and Backward Chaining - First Order Logic : Syntax and Semantics of First Order Logic – Using First Order Logic

Unit V Planning and Uncertainty 12 h

Language of Planning Problems – Planning with State-Space Search: Forward and Backward State-Space Search – Heuristics– Partial-Order Planning: Partial-order planning with unbound variables- Heuristics for partial-order planning- Uncertainty: Basic Probability Notation-Bayes' Rule and its Use

Text Books

1 Russell, S.J. and Norvig, P.,2014, "Artificial Intelligence: A Modern Approach", 3rd Edition, Pearson Education

References

1 Nilsson, N.J,2011,"Artificial Intelligence and New Systems", 1stEdition, Elsevier

2 Patterson, D.W.,2012,"Introduction to Artificial Intelligence and Expert Systems", Prentice Hall of India

Course Code	Course Name	Category	L	T	P	Credit
204DA2A2CB	DATA MINING AND BUSINESS INTELLIGENCE	CORE	5	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- Concepts and applications of data mining
- Data mining techniques
- Concepts and issues related to business intelligences and decision support systems

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Discuss the scope and necessity of Data Mining	K2
CO2	Apply clustering techniques	K3
CO3	Develop association rules for mining data	K5
CO4	Identify the foundations and capabilities of BI.	K3
CO5	Model Business intelligence applications	K4

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	S	S	M
CO2	S	S	S	S	M
CO3	S	S	S	S	S
CO4	M	S	M	S	M
CO5	M	S	S	S	M

S Strong M Medium L Low

204DA2A2CB	DATA MINING AND BUSINESS INTELLIGENCE	SEMESTER II

Total Credits: 4

Total Instruction Hours: 60 h

Syllabus

Unit I Data Mining Tasks 12 h

Introduction: Basic Data Mining Tasks – Data Mining Versus Knowledge Discovery in Databases – Data Mining Issues – Metrics – Classification: Introduction – Statistical Based Algorithms: Regression – Bayesian Classifications – Distance Based Algorithms: K Nearest Neighbors - Decision Tree Based Algorithms

Unit II Clustering 12 h

Similarity and Distance Measures – Outliers – Hierarchical Algorithms : Agglomerative – Divisive Clustering – Partition Algorithms : Minimum Spanning Trees – Squared Error Clustering Algorithms – K means Clustering – Nearest Neighbor Algorithm– Clustering with Genetic Algorithms – Clustering with Neural Networks

Unit III Association Rules 12 h

Association Rules - Introduction - Large Itemsets – Apriori Algorithm – Sampling algorithm – Partitioning – Parallel and Distributed Algorithms: Data and Task Parallelism – Incremental Rules – Advanced Association Rule Techniques: Generalized Association Rule – Multi Level Association Rules – Quantitative Association Rules

Unit IV Business Intelligence 13 h

Business Intelligence :Introduction – The Role of Mathematical Models - Business Intelligence Architectures: Business Intelligence Analysis Cycle - Enabling Factors in Business Intelligence Projects - Development of Business Intelligence System – Decision Support Systems: Evolution – Representation – Development - Ethics and Business Intelligence.

Unit V Applications 11 h

Visualization: Performance Dashboards - Balanced Scorecard - Business Intelligence Applications: Fraud Detection - Market Segmentation - Retail Industry- Telecommunications Industry - Banking and Finance – Customer Relationship Management

Text Books

1. Margaret H. Dunham, 2003, "Data Mining : Introductory and Advanced Topics", 1st Edition, Pearson Education(Unit I-III)
2. Carlo Vercellis,2009,"Business Intelligence, Data Mining Optimization for decision Making"1st Edition, Wiley,(Unit IV,V)

References

1. Arun.K.Pujari,2013,"Data Mining Techniques",3rd Edition,University Press (India) Limited
2. Jiawei Han, Micheline Kamber and Jian Pei,2012,"Data Mining Concepts and Techniques", 3rd Edition, Elsevier Publications

Course Code	Course Name	Category	L	T	P	Credit
204DA2A2CC	DATABASE ADMINISTRATION AND MANAGEMENT	CORE	4	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- DBMS components and their function
- Concepts of Database Administration
- Components of Structured Query Language

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Discuss the elements of Relational Database Management Systems	K2
CO2	Design database queries usingStructured Query Language.	K5
CO3	Understand the concept of a database transaction and recovery	K2
CO4	Explain features of Parallel and distributed database systems.	K4
CO5	Examine the role of the database administrator.	K3

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	S	S	M
CO2	S	S	S	S	M
CO3	S	M	M	M	M
CO4	S	S	S	S	M
CO5	M	M	M	S	S

S Strong M Medium L Low

204DA2A2CC	DATABASE ADMINISTRATION AND MANAGEMENT	SEMESTER II

Total Credits: 4

Total Instruction Hours: 48 h

Syllabus

Unit I Database Systems 8 h

Database Languages – Relational Databases – Database Design – Data Storage and Querying – Transaction Management – Database Architecture – Data Mining and Information Retrieval – Specialty Databases - Database Users and Administrators - Structure of Relational Databases – Schema – Keys- Relational Query Language – Relational Operations

Unit II Query Language 10 h

Introduction: Structured Query Language (SQL) : Data Definition – Aggregate Functions - Nested Subqueries – Join Expressions – Views – Integrity constraints – Functions and Procedures – Triggers - Embedded and Dynamic SQL

Unit III Transaction and Recovery 10 h

Concept of Transaction – Transaction Model – Storage Structure – Transaction Atomicity and Durability – Transaction Serizability – Concurrency Controls: Lock Based Protocols – Deadlock Handling – Multiple Granularity – Time-Stamp Based Protocols –Validity Based Protocols – Recovery and Atomicity – Recovery Algorithm – Buffer Management

Unit IV Parallel and Distributed Databases 10 h

Architecture for Parallel Databases – Parallel Query Evaluation – Parallelizing Individual Operations - Distributed Database : Distributed DBMS Architecture – Storing Data in Distributed DBMS – Distributed Catalog Management – Updating Distributed Data – Distributed Transactions – Distributed Concurrency control – Distributed Recovery

Unit V Database Administration 10 h

Types of Oracle Database Users - User Creation and Management - Tasks of a Database Administrator- Submitting Commands and SQL to the Database - Database Administrator Security and Privileges - Database Administrator Authentication - Creating and Maintaining a Password File - Data Utilities

Text Books

1. A.Silberchartz, H.F.Korth and S.Sudarshan, 2011, "Database System Concepts", 6th Edition, McGraw Hill (Unit I-III)
2. Raghu Ramakrishnan, Johannes Gehrke, 2003,"Database Management System",3rd Edition, McGraw Hill (Unit IV)
3. https://docs.oracle.com/cd/B19306_01/server.102/b14231/dba.htm(Unit V)

References

1. Elmasri Ramez and Navathe Shamkant.B,2010,"Fundamentals of Database System Concepts",6th Edition,2010, Addison Wesley

Course Code	Course Name	Category	L	T	P	Credit
204DA2A2CD	INFORMATION AND NETWORK SECURITY	CORE	4	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- The concepts of Information and Network security
- The issues regarding Confidentiality, Integrity and Availability of a data
- Various protocols forInformation and Network security

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Discuss the concepts of information security, threats, security services, and countermeasures.	K2
CO2	Illustrate the information security blue prints and major components	K3
CO3	Analyze various cryptographic algorithms	K4
CO4	Compare network security services and mechanisms	K4
CO5	Illustrate Internet security protocols for protecting data	K3

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	M	M	M	S
CO2	S	S	M	M	S
CO3	S	S	S	S	S
CO4	S	M	M	M	S
CO5	S	M	M	M	S

S Strong M Medium L Low

204DA2A2CD	INFORMATION AND NETWORK SECURITY	SEMESTER II

Total Credits: 4

Total Instruction Hours: 48 h

Syllabus

Unit I Information Security 8 h

Information Security: Introduction - Concepts – Components of Information Systems – Approaches to Information Security Implementation – System Development Life Cycle – Security Systems Development Life Cycle – Security Professionals and Organization - Needs of Security: Business Needs – Threats – Attacks – Secure Software Development

Unit II Security Planning and Technology 10 h

Information Security Planning and Governance – Information Security Policy, Standards and Practices – Information Security Blueprint - Continuity Strategies – Security Technology : Access Control – Firewalls – Protecting Remote Connections – Intrusion, Detection and Prevention Systems – Honeypots,Honeynets and Padded Cell System – Scanning and Analysis Tools

Unit III Cryptography 10 h

Introduction : OSI Security Architecture – Security Services – Classical Encryption Techniques :Symmetric Cipher Model – Substitution Techniques – Transposition Techniques – Rotor Machines - Steganography –Block Cipher and Data EncryptionStandards–Advanced Encryption Standards :AES Structure – AES Transformation Functions-RSA

Unit IV Network Security 10 h

Network Access control – Extensible Authentication Protocol – Cloud Computing – Cloud Security risks and counter measures – Data protection in the cloud – Cloud security as a service - Addressing cloud computing security concerns – Transport Level Security :Web security Considerations - HTTPS –Secure Shell

Unit V Internet Security 10 h

Introduction : Internet Mail Architecture – Email Threats and comprehensive email security – S/MIME –Pretty Good Privacy –DNS Based Authentication of Named Entities – Sender Policy Framework –IP Security : Overview –Policy –Encapsulating security payload –Combining Security Associations – Internet key Exchange

Text Books

1. Michael E.Whitman and Herbert J Mattord, 2011, "Principles of Information security", 4th Edition, Cengage Learning(Unit I,II)
2. William Stallings,2017,"Cryptography and Network Security Principles and Practices",7th Edition ,Prentice hall(Unit III-V)

References

1. Nina Godbole,2017,"Information Systems Security: Security Management, Metrics, Frameworks and Best Practices", 2nd Edition, Wiley
2. Micki Krause ,Harold F.Tipton,2008,"Handbook of Information security management", Vol 1-3, CRC press

204DA2A2CP	CORE PRACTICAL: DATA MINING	SEMESTER II

Total Credits: 2
Total Instructions Hours: 48h

S.No	List of Experiments
1	Explore available Data Sets in WEKA
2	Create data sets using WEKA
3	Program to Demonstrate Preprocessing on data sets
4	Program to Demonstrate Normalization
5	Program to Apply Association rules on data sets
6	Program to Demonstrate Classification Rule on data Sets
7	Program to Implement the Bayesian Classification
8	Program to Demonstrate Clustering Techniques on data Sets
9	Program to Implement K Means Clustering Techniques
10	Program to create a Decision tree for given datasets

Note: Eight Programs are Mandatory

204DA2A2CQ	CORE PRACTICAL: R PROGRAMMING	SEMESTER II

Total Credits: 2
Total Instructions Hours: 48h

S.No	List of Experiments
1	Program to Create and Manipulation of Vectors in R
2	Program for Matrix Manipulation
3	Program to perform operations on Factors
4	Operations on Data Frame in R
5	Programs to test Lists in R
6	Programs to test mathematical functions in R
7	Programs to implement the array object
8	Program to Import data sets and Perform read and write in R
9	Program to Create data sets and Plot graph in R
10	Program to Plot functions to customize the graph

Note: Eight Programs are Mandatory

Course Code	Course Name	Category	L	T	P	Credit
204DA2A2EA	ADVANCED STATISTICS	DSE	4	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- Statistical tools for analyzing data
- Statistical quality control measures
- Concept of nonparametric tests

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Examine the experimental designs	K2
CO2	Develop concept of acceptance sampling plans	K5
CO3	Analyse various statistical decision theory	K4
CO4	Model solutions using partial and multiple correlation	K4
CO5	Analyse the concept of non-parametric tests	K4

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	M	M	S	M
CO2	S	S	M	S	M
CO3	S	M	M	S	M
CO4	S	S	M	S	M
CO5	S	S	M	S	M

S Strong M Medium L Low

204DA2A2EA	ADVANCED STATISTICS	SEMESTER II

Total Credits: 4

Total Instruction Hours: 48 h

Syllabus

Unit I Experimental Designs 8 h

Introduction- Randomized block design - advantages of a completely randomized experimental design - Latin Square - Significance of Latin Square - Assumptions in the analysis of Latin Square - Steps in construction of Latin square – Randomized Block vs Latin Square –Latin Cubes - Factorial experiment

Unit II Statistical Quality Control (SQC) 10 h

Introduction-.Control Charts- X Chart – R Chart – Control chart for number of defects per unit - Control chart for fraction defectives -Advantages and limitations of SQC-Acceptance sampling: Role of acceptance sampling –Types of acceptance plans - Advantages of double sampling plan - Selection of sampling plan –OC curve: Construction - Characteristics – AQL and LTPD

Unit III Partial And Multiple Correlation 10 h

Partial correlation - Zero order First order and Second order coefficients- Partial correlation coefficients in case of four variables – Second order partial correlation coefficients- The significance of a partial correlation coefficients - Multiple correlation - Coefficients of multiple correlation- Multiple regression analysis - Reliability of estimates - Coefficient of Multiple determination

Unit IV Statistical Decision Theory 10 h

Statistical Decision Theory : Introduction - Terminologies - Ingredients of decision problem - optimal decisions - Decision Tree analysis - Steps in Decision Tree Analysis - Advantages and Drawbacks of Decision Tree approaches - Decision Tree methods: Applications for Classification and Prediction

Unit V Non-Parametric Tests 10 h

Introduction- Advantages of non-parametric tests - The sign Test - The paired sample sign test – A Rank Sum Test: The Mann-Whitney U Test- The one sample run test - The Kruskal – Wallis or H test - The Spearman Rank Correlation – Spearman limitations of non-parametric test.

Note: 20% Theory and 80% Problem

Text Books

1 Gupta, S.P, 2014, "Statistical Methods", 44th Edition, Sultan Chand and Sons, New Delhi

References

1 Ronald E. Walpole, 2018, "Probability & Statistics", 9th Edition, Pearson education, South Asia

2 Sheldon M.Ross, 2017, "Introductory Statistics", 4th Edition, Academic Press, United States

3 Gupta S.C and Kappoor V.K., " Fundamentals Applied Statistics", 9th Edition, Sultan Chand and Sons, New Delhi.

4 Robert. V. Hogg and Allen T.G. Craig, "Introduction to Mathematical Statistics", 5th edition, Pearson Education, 2006

Course Code	Course Name	Category	L	T	P	Credit
204DA2A2EB	IMAGE AND VIDEO ANALYTICS	DSE	4	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- The concepts of Image analytics
- The concepts of Video analytics
- The applications of Image and Video analytics

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Understand the concepts of image processing	K2
CO2	Experiment with image filtering	K3
CO3	Apply color processing and compression techniques	K3
CO4	Understand the concepts of object detection and tracking	K2
CO5	Apply image and video analytics in real life problems	K3

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	S	S	S
CO2	S	S	S	S	S
CO3	S	M	S	M	M
CO4	S	M	S	S	M
CO5	S	M	S	S	S

S Strong M Medium L Low

204DA2A2EB	IMAGE AND VIDEO ANALYTICS	SEMESTER II

Total Credits: 4

Total Instruction Hours: 48 h

Syllabus

Unit I Image Representation and Processing 8 h

Digital image representation- Visual Perception- Sampling and Quantization- Basic Relations between Pixels- Mathematical Tools Used in Digital Image Processing: Fundamental Operations –Vector and Matric Operations- Image Transforms (DFT, DCT,DWT, Hadamard).

Unit II Image Filtering 10 h

Fundamentals of spatial filtering: spatial correlation and convolution-smoothing, blurring-sharpening- edge detection - Basics of filtering in the frequency domain: smoothing-blurring- sharpening--Histograms and basic statistical models of image

Unit III Colors and Compression 10 h

Color Models: Introduction - Color models and Transformations – Image and Video segmentation - Image and video demonizing - Image and Video enhancement- Image and Video compression.

Unit IV Object Detection and Tracking 10 h

Object detection and recognition in images - Object detection and recognition in video -Texture models - Image and Video classification models - Introduction to Object tracking - Object tracking in Video

Unit V Applications 10 h

Applications and Case Studies - Industrial Sector - Retail Stores - Transportation and Travel Sector - Remote sensing- Healthcare domain - Government and law enforcement sector - IoT Surveillance

Text Books

1. Rafael C.Gonzalez, Richard E.Woods, 2008, “Digital Image Processing”, 3rd Edition, Pearson Education(Unit I-III)
2. Caifeng Shan, FatihPorikli, Tao Xiang, Shaogang Gong, 2002, "Video Analytics for Business Intelligence", Springer(Unit IV,V)

References

1. A.L. Bovik,2005,,”Handbook of Image and Video Processing“, 2nd Edition, Elsevier Academic Press
2. AsierPerallos, Unai Hernandez-Jayo, Enrique Onieva, Ignacio Julio García Zuazola, 2015, “Intelligent Transport Systems: Technologies and Applications”, Wiley,
3. Jean-Yves Dufour, 2013, “Intelligent Video Surveillance Systems”, Wiley

Course Code	Course Name	Category	L	T	P	Credit
204DA2A2EC	WEB INTELLIGENCE	DSE	4	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- The concepts of web mining
- The techniques in opinion mining
- The concepts of social network Analysis

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Understand the concepts of web mining	K2
CO2	Analyze social networks and web crawling	K4
CO3	Experiment with opinion mining and sentiment analysis	K5
CO4	Understand Google Analytics	K2
CO5	Design Applications using web intelligence	K5

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	S	S	S
CO2	S	S	S	S	S
CO3	S	S	S	S	M
CO4	S	S	S	S	S
CO5	S	S	S	S	S

S Strong M Medium L Low

204DA2A2EC	WEB INTELLIGENCE	SEMESTER II

Total Credits: 4

Total Instruction Hours: 48 h

Syllabus

Unit I Fundamentals of Web Mining 8 h

Introduction - Web Mining: Information Retrieval and Web Search - Basic Concepts of Information Retrieval - Information Retrieval Models - Relevance Feedback - Evaluation Measures - Text and Web Page Pre-Processing - Web Search - Meta Search: Combining Multiple Rankings - Web Spamming

Unit II Social Network Analysis and Web Crawling 10 h

Social Network Analysis - Co-Citation and Bibliographic Coupling - Page Rank - Web Crawling: Basic Crawler Algorithm - Implementation Issues - Universal Crawlers - Focused Crawlers - Topical Crawlers : Topical Locality and Cues - Best-First Variations - Adaptation - Evaluation - Crawler Ethics and Conflicts

Unit III Opinion Mining and Sentiment Analysis 10 h

The Problem of Opinion Mining - Document Sentiment Classification - Sentence Subjectivity and Sentiment Classification – Opinion Lexicon Expansion - Aspect-Based Opinion Mining – Mining Comparative Opinions - Opinion Search and Retrieval - Opinion Spam Detection - Utility of Reviews

Unit IV Google Analytics 10 h

Google Analytics: Introduction - Cookies - Accounts vs Property - Tracking Code - Tracking Unique Visitors - Demographics - Page Views and Bounce Rate Acquisitions -Custom Reporting

Unit V Applications 10 h

Applications: Filters - Ecommerce Tracking - Real Time Reports - Customer Data-Alert - Adwords Linking - Adsense Linking - Attribution Modeling - Segmentation - Campaign Tracking - Multi-Channel Attribution

Text Books

1 Bing Liu ,2011, "Web Data Mining Exploring Hyperlinks, Contents, and Usage Data", 2nd Edition, Springer(Unit I-III)

2 Ning Zhong , Jiming Liu and Yiyu Yao, 2010, "Web Intelligence", Springer(Unit IV,V)

References

1 Ricardo Baeza -Yates and BerthierRibeiro-Neto,2011,"Information Retrieval: The Concepts and Technology behind Search,2nd Edition,ACM Press

2 Juan D. Velasquez, Lakhmi C. Jain (Eds.),2010,"Advanced Techniques in Web Intelligence - 1",1st Edition,Springer

3 Mark Levene,2010,"An Introduction to Search Engines and Web Navigation",2nd Edition,Wiley

4 Eric Fettman, Shiraz Asif, Feras Alhlou , 2016 "Google Analytics Breakthrough", Wiley

13/7/2020

BoS Chairman/HoD
Department of Computer Science with Data Analytics
Dr. N. G. P. Arts and Science College
Coimbatore – 641 048

Academic Council
Dr.V.Rajendran
CHAIRMAN
Dr. N.G.P. Arts and Science College

Dr.NGPASC
COIMBATORE | INDIA

M.Sc.Computer Science with Data Analytics (Students admitted during the AY 2020-21)

Course Code	Course Category	Course Name	L	T	P	Exam (h)	Max Marks			Credits
							CIA	ESE	Total	
Third Semester										
204DA2A3CA	Core –X	Wireless Networks	4	-	-	3	25	75	100	4
204DA2A3CB	Core – XI	Cloud Computing	4	-	-	3	25	75	100	4
204DA2A3CC	Core –XII	Software Project Management	4	-	-	3	25	75	100	4
204DA2A3CD	Core –XIII	Machine Learning	4	-	-	3	25	75	100	4
204DA2A3CV	Core- XIV	Project Work-I	-	-	2	3	40	60	100	2
204DA2A3CP	Core Practical-IV	Big Data Analytics	-	-	4	3	40	60	100	2
204DA2A3CQ	Core Practical-V	Machine Learning	-	-	4	3	40	60	100	2
204DA2A3DA	DSE – II	Principles of Deep Learning	4	-	-	3	25	75	100	4
204DA2A3DB		Predictive Data Analytics								
204DA2A3DC		Health Care Data Analytics								
		Total	**20**	**-**	**10**	**-**	**-**	**-**	**800**	**26**

Course Code	Course Name	Category	L	T	P	Credit
204DA2A3CA	WIRELESS NETWORKS	CORE	4	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- Concepts of wireless networks
- Wireless architecture, protocols and applications
- Wireless sensor networks

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Understand the concepts of wireless networks.	K5
CO2	Describe the structure of cellular systems.	K3
CO3	Discuss Mobility Management and security.	K2
CO4	Explain Wireless Application Protocols	K4
CO5	Describe functions of wireless sensor networks	K6

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	M	S	M	M
CO2	S	M	S	M	M
CO3	S	M	S	M	M
CO4	S	S	S	M	M
CO5	S	S	S	S	M

S Strong M Medium L Low

204DA2A3CA	WIRELESS NETWORKS	SEMESTER III

Total Credits: 4

Total Instruction Hours: 48 h

Syllabus

Unit I Overview of Wireless Systems 10 h

Introduction - First and Second-Generation Cellular Systems - Cellular Communications from 1G to 3G - Wireless 4G Systems- Future Wireless Networks - Teletraffic Engineering: Service Level - Traffic Usage - Traffic Measurement Units - Call Capacity - Mobile System Terms - Data Collection - Digital Communication and Transmission: Baseband Systems - Messages, Characters and Symbols - Sampling Process - Voice Communication.

Unit II Cellular Communications and WWAN 10 h

Cellular Systems - Hexagonal Cell Geometry - Cochannel Interference Ratio - Cochannel Interference Reduction - Cell Splitting - Adjacent Channel Interference - Architecture of Wireless Wide-Area Network (WWAN): WWAN Subsystem Entities - Logical Channels - Channel and Frame Structure -Basic Signal Characteristics -Speech Processing- Power Levels in Mobile Station.

Unit III Mobility Management and Security 09 h

Mobility Management Functions - Mobile Location Management - Mobile Registration – Handoff-Security in Wireless Systems: Security and Privacy Needs of a Wireless System - Required Features for a Secured Wireless Communications System - Methods of Providing Privacy and Security in Wireless Systems - Wireless Security and Standards.

Unit IV Wireless Application Protocol (WAP) and Bluetooth 09 h

Introduction: WAP and the World Wide Web - The WAP Programming Model - WAP Architecture, Advantages and Disadvantages – Applications of WAP - Wireless Personal Area Network - Bluetooth (IEEE 802.15.1) - Terms Used in Bluetooth - Bluetooth Protocol Stack - Bluetooth Link Types- Bluetooth Security.

Unit V Wireless Sensor Network and WLAN 10 h

Introduction - Wireless Sensor Network - Usage of Wireless Sensor Networks - Wireless Sensor Network Model - Sensor Network Protocol Stack - ZigBee Technology - Wireless Local Area Networks: WLAN Equipment - WLAN Topologies - WLAN Technologies - Other WLAN Standards.

Text Books

1 Vijay K Garg , 2007 "Wireless Communications and Networking", Morgan Kaufmann Publishers.

References

1 Nicopolitidis P, Obadiah M S, Papadimitriou G S and Pomportsis A S,2009, "Wireless Networks", John Wiley and Sons

2 Cory Beard, William Stallings, "Wireless Communication Networks and Systems", 2016,Pearson Higher Education, Inc

Course Code	Course Name	Category	L	T	P	Credit
204DA2A3CB	CLOUD COMPUTING	CORE	4	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- Architecture of cloud computing
- Major cloud services and its Applications
- Security methodology and Virtualization process over Networks

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Understand the architecture of cloud computing	K2
CO2	Describe various service providers and cloud services	K3
CO3	Discuss virtualization data storage in Network	K3
CO4	Analyze security measures in cloud computing	K4
CO5	Review future directions in mobile cloud ecosystems	K5

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	S	M	M
CO2	M	M	S	S	M
CO3	S	S	S	M	M
CO4	S	S	S	S	M
CO5	M	M	M	S	M

S Strong M Medium L Low

204DA2A3CB	CLOUD COMPUTING	SEMESTER III

Total Credits: 4

Total Instruction Hours: 48 h

Syllabus

Unit I Introduction 9 h

Introduction: Cloud Computing Basics: Cloud Computing Overview - Applications of cloud computing – Intranets and the cloud – First movers in the cloud - Benefits - limitations of cloud computing – Security Concerns – Cloud Computing Services.

Unit II Types of Cloud Services 10 h

Developing cloud services- -Advantages and Disadvantages –Types of cloud services development- Software as a Service – Platform as a service – On-Demand computing-Discovering cloud service development and tools.

Unit III Virtualization 9 h

Introduction - Understanding Virtualization - History of Virtualization – Server Virtualization – Desktop Virtualization – Virtual Networks – Data Storage Virtualization. Data Storage in Cloud: Evolution of Network Storage – Cloud based data Storage – Advantages and disadvantages - Cloud based Backup systems

Unit IV Cloud Security 10 h

Introduction – General security advantages – Introducing business continuity and disaster recovery: Data storage wiping- Distributed Denial of Service Attacks-Packet Sniffing- Man-in the Middle Attack-Monitoring Device Screens-Malicious Employees- Hypervisor Attack- Guest - Hopping Attack-SQL-Injection Attack-Physical Security

Unit V Future Directions 10 h

Cloud services for individuals – Migration : Applications needed- sending the existing data to the cloud- Mobile Cloud Computing: Evolution of Mobile Computing – Mobile Cloud EcoSystem – Mobile Players

Text Books

1 1 Anthony T. Velte, Toby J. Velte, Robert Elsenpeter, "Cloud Computing: A Practical Approach", McGraw Hill Edition.(Unit-I)

2 Kris Jamsa,(2013), "Cloud Computing SaaS, PaaS, IaaS, Virtualization, Business Models, Mobile,Security and More" Jones and Barlett Learning(Unit II-IV) .

3 Michael Miller ,2008,"Cloud Computing Web based application that change the way you work and collaborate online". Pearson Education(Unit V)

References

1 John W .Rittinghouse and James F.Ransome(2009), "Cloud Computing: Implementation, Management, and Security", CRC Press.

2 Kailash Jayaswal , Jagannath Kallakurchi , Donald J. Houde , Deven Shah , Kogent Learning(2014), Cloud Computing Black Book, Dreamtech Press

Course Code	Course Name	Category	L	T	P	Credit
204DA2A3CC	SOFTWARE PROJECT MANAGEMENT	CORE	4	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- Software Project Planning methods.
- Risk Management and Evaluation techniques
- Software quality control

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Gain knowledge on project management concepts and process models	K1
CO2	Analyse software effort estimation techniques.	K2
CO3	Estimate the risks involved in various project activities.	K5
CO4	Describe Project Monitoring and Control.	K4
CO5	Apply quality control measures in software projects.	K5

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	M	M	M	M	M
CO2	S	S	M	M	M
CO3	S	S	S	M	S
CO4	M	M	S	M	S
CO5	S	S	S	S	M

S Strong M Medium L Low

204DA2A3CC	SOFTWARE PROJECT MANAGEMENT	SEMESTER III

Total Credits: 4

Total Instruction Hours: 48 h

Syllabus

Unit I Project Evaluation and Project Planning 09 h

Introduction to Software Project Management – Activities - Project planning: Overview-Steps involved in project planning – Project Evaluation: Strategic Assessment -Technical Assessment- Cost Benefit Analysis – Cash Flow Forecasting - Cost Benefit Evaluation Techniques - Risk Evaluation

Unit II Project Selection and Effort Estimation 10 h

Choosing Technologies – Technical plan content list - Choice of Process models – The Waterfall model – The V-Process model – The spiral model – Software prototyping-Software effort estimation: Problems with over and under estimates – The basis of software estimating – Software effort estimating techniques- Expert judgment - Albrecht function points Analysis – Constructive Cost Model

Unit III Activity Planning and Risk Management 10 h

Objectives of Activity planning – Project Schedules – Project and activities – Sequencing and scheduling activities – Network Planning models – Formulating a network model – Forward and Backward Pass– Identifying the Critical Path - Risk Management: Categories of risk - Risk Identification – Risk Assessment – Risk Planning –Risk Management – Project Evaluation and Review Techniques – Monte Carlo Simulation

Unit IV Project Monitoring and Control 09 h

Creating the framework – Collection of data – Visualizing progress – Cost monitoring – Earned Value – Prioritizing Monitoring – Change Control – Managing Contracts: Types of Contracts – Stages in Contract Placement – Terms of Contract – Contract Management - People management

Unit V Software Quality 10 h

Introduction – The place of software quality in Project planning – The Importance of software quality –Defining software quality – ISO 9126 – Product versus Process quality management – Quality Management Systems – Process Capability models – Techniques to enhance the software quality-Testing – Quality Plans.

Text Books

1 Bob Hughes, Mike Cotterell, (2010)," Software Project Management",(5th Edn), Tata McGraw Hill.

References

1 Robert K. Wysocki,(2011),"Effective Software Project Management", Wiley Publication

2 Walker Royce, (1998),"Software Project Management", Addison-Wesley

3 Gopalaswamy Ramesh,(2013), "Managing Global Software Projects", McGraw Hill Education

Course Code	Course Name	Category	L	T	P	Credit
204DA2A3CD	MACHINE LEARNING	CORE	4	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- Basic Concepts and Techniques in Machine Learning
- Regressions, Classification and Clustering
- Dimensionality reduction techniques

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Understand machine learning concepts	K2
CO2	Describe Bayesian Theory and Parametric Models	K2
CO3	Discuss Multivariate Methods and Dimensionality reduction	K3
CO4	Apply Clustering techniques	K4
CO5	Apply Hidden Markov and Graphical Models	K5

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	S	S	M
CO2	S	S	S	S	M
CO3	S	S	S	S	M
CO4	S	S	S	S	S
CO5	S	S	S	S	M

S Strong M Medium L Low

204DA2A3CD	MACHINE LEARNING	SEMESTER III

Total Credits: 4

Total Instruction Hours: 48 h

Syllabus

Unit I Introduction 8 h

Introduction to Machine Learning - Applications: Learning Associations - Classification- Regression - Unsupervised Learning - Reinforcement Learning - Supervised learning: Learning a Class - Probably Approximation Correct Learning - Noise – Learning Multiple Classes-Regression-Model selection and Generalization-Dimensions of supervised machine learning algorithm

Unit II Bayesian Decision theory and Parametric Methods 9 h

Introduction- Classification - Losses and Risks - Discriminant Functions - Utility theory - Association Rules - Parametric Methods: Maximum Likelihood estimation - Evaluating an Estimator - Bayesian Estimation - Parametric Classification - Regression - Model Selection Procedures

Unit III Multivariate Methods and Dimensionality Reduction 10 h

Multivariate Data-Parameter Estimation-Estimation of Missing values-Multivariate normal Distribution- Multivariate Classification - Tuning complexity - Discrete Features - Multivariate Regression - Dimensionality Reduction : Subset Selection - Principal Component Analysis - Factor Analysis - Multidimensional Scaling - Linear Discriminant Analysis – Isomap

Unit IV Clustering and Non Parametric Methods 11 h

Clustering: Mixture Densities - K-means-Expectation Maximization Algorithms - Mixtures of Latent Variable Models-Supervised Learning after Clustering-Hierarchical Clustering-Choosing the number of clusters-Nonparametric Methods: Non Parametric Density Estimation - Generalization to Multivariate Data-Nonparametric Classification-Condensed Nearest Neighbor- Nonparametric Regression

Unit V Hidden Markov Models(HMM) and Graphical Models 10 h

Discrete Markov Processes - Hidden Markov Models - Basic Problems-Evaluation Problem-Finding the State Sequence-Learning Model Parameters - HMM with input - Model selection in HMM-Graphical Models: Naive Bayes' Classifier-HMM-Linear Regression- d-Separation-Belief Propagation

Text Books

1 Ethem Alpaydin, (2010),"Introduction to Machine Learning", (2nd Edn.), MIT Press

References

1 Kevin P. Murphy(2002), "Machine Learning A Probabilistic Perspective", MIT Press

2 Shai Shalev-Shwartz and Shai Ben-David , (2014),"Understanding Machine Learning from Theory to Algorithms", (1st Edn), University Press.

204DA2A3CP	CORE PRACTICAL : BIG DATA ANALYTICS	SEMESTER III

Total Credits: 2
Total Instructions Hours: 48 h

S.No	LIST OF EXPERIMENTS
1	Install different modes of cluster with Hadoop Management.
2	Implement Hadoop file management like adding files and directories, retrieving files, deleting files
3	Move the Hadoop files from specified source to destination
4	Pig Latin scripts to group and find the desired results from data items
5	Pig Latin Scripts to find the total sales value from book data item
6	Pig Latin Scripts to find the maximum temperature from weather data for every month of year
7	Hive Scripts and to Perfom QL Manipulation
8	Hive Scripts to apply the Queries and Views

204DA2A3CQ	CORE PRACTICAL : MACHINE LEARNING	SEMESTER III

Total Credits: 2
Total Instructions Hours: 48 h

S.No	LIST OF EXPERIMENTS
1	Implement a real world problem using Linear Regression
2	Program to Implement Logistic Regression
3	Program to Implement Naïve Bayesian Classifier.
4	Program to implement back propagation algorithm
5	Program to implement the Multi-Class Classification
6	Program to construct a Bayesian network
7	Program to Implement candidate elimination algorithm
8	Program to implement K-Nearest neighbour algorithm
9	Programs for Anomaly Detection
10	Programs for Recommendation Systems

Course Code	Course Name	Category	L	T	P	Credit
204DA2A3DA	PRINCIPLES OF DEEP LEARNING	DSE	4	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- Concepts of Deep Learning
- Neural Networks
- Applications of Deep Learning

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Understand the basics of deep learning	K2
CO2	Explain Neural Networks	K2
CO3	Apply Convolution Neural Networks to model solutions	K3
CO4	Analyze Multimodality in Deep Learning	K4
CO5	Design Recent Trends in Deep Learning	K5

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	S	S	M
CO2	S	S	S	M	M
CO3	S	S	S	S	M
CO4	S	S	S	S	M
CO5	S	S	S	S	S

S Strong M Medium L Low

204DA2A3DA	PRINCIPLES OF DEEP LEARNING	SEMESTER III

Total Credits: 4

Total Instruction Hours: 48 h

Syllabus

Unit I Introduction 8 h

AI and Deep Learning - The history and rise of deep learning - Advantages - Impact of Deep Learning -Basics of Linear Algebra: Data Representation - Data Operations - Matrix Properties - Deep architecture: Neural viewpoints - Representation viewpoints – Applications - Deep Learning software frameworks

Unit II Neural Networks 9 h

Multilayer Perceptrons - Activation Functions - Network Learns: Weight Initialization - Forward Propagation - Back Propagation - Vanishing and Exploding gradients - Optimization algorithms - Regularization - Deep Learning Models: Convolution Neural Networks - Restricted Boltzmann Machines - Recurrent Neural Networks

Unit III Deep Learning in Computer Vision and NLP 10 h

Origin of Convolution Neural Networks (CNNs) - CNN: Data Transformation - Network Layers - Network Initialization – Regularizations – Loss Functions – Model Visualization -Natural Language Processing Vector Representation: Traditional NLP - Deep learning NLP: Word Embeddings - Word2Vec - Applications

Unit IV Multimodality 11 h

Multimodality learning - Challenges of Multimodality learning- Image captioning - Visual Question Answering - Multi-Source Based Self-driving - Reinforcement Learning: Introduction- Deep Reinforcement learning: Deep Q - Network – Double Deep Q-Network - Implementing reinforcement learning

Unit V Deep Learning Trends 10 h

Training Methods: Weight Initialization - Optimization- Choosing Loss Function - Preventing Overfitting - Fine-tuning - Model compression - Recent models for deep learning – Genomics – Predictive Medicine – Clinical Imaging – Visual Reasoning – Code Synthesis

Text Books

1 Wei Di,Anurag Bhardwaj,Jianing Wei,(2018),"Deep Learning Essentials",(1st Edn),Packt Publishing

References

1 Goodfellow I , Bengio Y, and Courville A,(2016), "Deep Learning",MIT Press

2 Charu C. Aggarwal,(2018),"Neural Networks and Deep Learning",Springer

Course Code	Course Name	Category	L	T	P	Credit
204DA2A3DB	PREDICTIVE DATA ANALYTICS	DSE	4	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- Concepts of data mining
- Statistical Analysis
- Applications of Predictive Analytics

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Understand basic concepts of data mining	K2
CO2	Discuss Exploratory Data Analysis	K2
CO3	Examine Statistical Approaches for model prediction	K3
CO4	Develop a predictive data model	K4
CO5	Discuss case studies in predictive modeling	K5

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	S	S	S	M	M
CO2	S	S	S	S	M
CO3	S	S	M	M	M
CO4	M	S	M	S	M
CO5	M	M	M	M	M

S Strong M Medium L Low

204DA2A3DB	PREDICTIVE DATA ANALYTICS	SEMESTER III

Total Credits: 4

Total Instruction Hours: 48 h

Syllabus

Unit I Introduction 10 h

Introduction to Data Mining and Predictive Analytics – Data Mining Tasks - Data Preprocessing: Data Cleaning - Handling Missing Data - Identifying Misclassification - Graphical Methods for Identifying Outliers - Measures of Center and Spread - Data Transformation - Min-Max Normalization - Z-Score Standardization - Decimal Scaling - Transformations - Numerical Methods for Identifying Outliers

Unit II Exploratory Data Analysis(EDA) 8 h

Hypothesis Testing Versus EDA- Exploring Categorical Variables - Numeric Variables - Multivariate Relationships - Uncover Anomalous Fields- Driving New Variables: Flag and Numerical Variables-Investigate Correlated Predictor Variables – Dimension Reduction Methods: Needs - Principal Component Analysis(PCA) - Applying PCA to Data Set – Profiling and Validation of PCA

Unit III Statistical Analysis 10 h

Univariate Statistical Analysis: Statistical Approaches to Estimation and Prediction –Confidence Interval Estimation of Mean and Proportion - Hypothesis Testing for Mean and Proportion, Multivariate Statistics: Two Sample Z-Test for Difference in Means and Proportions - Test for Homogenity and Chi-Square - Supervised and Unsupervised Methods – Statistical Methodology -Cross Validation - Overfitting - Classification and Regression Methods

Unit IV Predictive Data Model 10 h

Introduction: Listing The Business Objectives: Identifying Related Objectives-Collecting User Requirements - Processing Data: Identifying Data – Cleaning Data – Generating any Derived Data-Reducing Dimensionality of Data - Structuring Data - Building A Predictive Model: Developing and Testing the Model - Going Live with the Model- Visualization as a Predictive tool - Evaluating Visualization - Visualizing The Model's Analytical Results

Unit V Case Studies 10 h

Business Understanding - Customer Relationship Management: Churn Management - Health Care : Disease Prediction- Marketing : Purchase Prediction - Behavioral Analytics: Fraud Detection

Text Books

1. Daniel T.Larose, Chantal D.Larose, (2015), Data Mining and Predictive analytics ,(2nd Edn.),Wiley series.(Unit I-IV)
2. John D. Kelleher, Brian Mac Namee, Aoife D'Arcy,(2020), Fundamentals of Machine Learning for Predictive Data Analytics, (2nd Edn.),MIT Press(Unit V)

References

1. Anasse Bari, Mohamad Chaouchi, Tommy Jung,(2016),” Predictive Analytics For Dummies”, (2nd Edn.),
2. Vijay Kotu,Bale Deshpande,(2015),”Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner”,Margaun Koffmann edition

Course Code	Course Name	Category	L	T	P	Credit
204DA2A3DC	HEALTH CARE DATA ANALYTICS	DSE	4	-	-	4

PREAMBLE

This course has been designed for students to learn and understand

- Various forms of electronic health care information.
- Techniques adopted to analyze health care data.
- Predictive models for clinical data.

COURSE OUTCOMES

On the successful completion of the course, students will be able to

CO Number	CO Statement	Knowledge Level
CO1	Describe different types of data generated in health care.	K2
CO2	Discuss Biomedical Image Analysis.	K3
CO3	Describe Biomedical Signal Analysis.	K3
CO4	Apply mining techniques for clinical texts.	K2
CO5	Analyze health care data using Statistical Prediction Models.	K5

MAPPING WITH PROGRAMME OUTCOMES

COs/POs	PO1	PO2	PO3	PO4	PO5
CO1	M	M	S	S	M
CO2	S	M	S	M	M
CO3	M	M	S	M	S
CO4	M	S	S	M	M
CO5	S	S	S	S	M

S Strong M Medium L Low

204DA2A3DC	HEALTH CARE DATA ANALYTICS	SEMESTER III

Total Credits: 4

Total Instruction Hours: 48 h

Syllabus

Unit I Introduction 08 h

Healthcare Data Sources and Basic Analytics: Electronic Health Records - Advanced Data Analytics for Healthcare – Application and Practical Systems for Healthcare – Resources for Health Care Data Analytics – Electronic Health Records(EHR) : Introduction – History- Components – Coding Systems-Benefits of HER

Unit II Biomedical Image Analysis and Mining of Sensor Data 10 h

Introduction – Biomedical Imaging Modalities – Object Detection – Image Segmentation – Image Registration – Feature Extraction - Mining of Sensor Data in Healthcare-Mining Sensor Data in Medical Informatics Sensor Data :Scope and Challenge – Challenges in HDA –Sensor Data Mining Applications-Non Clinical Healthcare Applications.

Unit III Bio Medical Signal Analysis 10 h

Introduction – Types of Biomedical Signals –Electro Cardiography Signal Analysis - Electrode Contact Noise and Motion Artifacts – Denoising of Signals :Principal Component Analysis – Wavelet Filtering – Wavelet Wiener Filtering – Pilot Estimation Method – Multivariate Bio Medical Signal Analysis

Unit IV NLP and Data Mining for Clinical Text 10 h

Introduction – Natural Language Processing (NLP)-Mining Information from Clinical Text: Information Extraction – Current Methodologies – Clinical Text Corpora and Evaluation Metrics – Information for Integrating Biology and the Beside– Challenges of Processing Clinical Reports – Clinical Applications

Unit V Advanced Data Analytics for Health Care 10 h

Introduction – Basic Statistical Prediction Models: Linear Regression – Logistic Regression – Bayesian Models – Alternative Clinical Prediction Models: Decision Trees – Artificial Neural Networks – Cost Sensitive Learning – Advanced Predictive Models – Survival Models: Basic Concepts – Non Parametric Survival Analysis – Cox Propositional Hazards Model- Survival Trees- Evaluation and Validation

Text Books

1 Chandan K. Reddy and Charu C Aggarwal, (2015), "Healthcare Data Analytics", Taylor & Francis, CRC Press

References

1 Hui Yang and Eva K. Lee,(2016), "Healthcare Analytics: From Data to Knowledge to Healthcare Improvement", Wiley

204DA2ASSA	PROFESSIONAL ETHICS	SEMESTER III

Total Credit: 1

Syllabus

Unit I Overview of Ethics

Ethics-Introduction –Importance of Integrity- Difference between Morals , Ethics and Laws- Ethics in the Business world- Corporate Social Responsibility - Creating an Ethical work Environment- Ethical Considerations in Decision Making - Ethics in Information Technology

Unit II Ethics for Information Technology Workers

IT Professionals - Professional Codes of Ethics- Professional Organizations– Licensing – Malpractice - IT Users - Common Ethical Issues For IT Users - Supporting Ethical Practices

Unit III Computer and Internet Crime

IT Security Incidents - Types of Exploits -Types of Perpetrators - Implementing Trustworthy Computing - Risk Assessment - Establishing a Security Policy - Educating Employees - Prevention - Detection - Response

Unit IV Privacy

Privacy Protection and the Law- Information Privacy - Privacy Laws-Applications--Key Privacy-Anonymity Issues-Data Breaches-Electronic Discovery-Consumer Profiling-Workplace Monitoring-Freedom of Expression

Unit V Intellectual Property

Introduction - Copyrights -Patents -Trade Secrets -Key Intellectual Property Issues - Plagiarism- Reverse Engineering - Open Source code-Competitive Intelligence-Trademark Infringement-Cybersquatting

Text Books

1 George W. Reynolds,(2014), Ethics in Information Technology, (5th Edn), Strayer University, Cengage Learning

References

1 Van de Poel, I., and L. Royakkers, (2011). Ethics, Technology and Engineering: An Introduction. Wiley-Blackwell..

2 R.S.Nagaarazan,(2006),"Professional Ethics and Human Values", (1st Edn.), New Age International Pvt. Limited.

204DA2ASSB	HUMAN COMPUTER INTERACTIONS	SEMESTER III

Total Credit: 1

Syllabus

Unit I Foundations

The Human: Input-Output channels – Human Memory – Thinking: reasoning and problem solving - Emotion - Individual Differences - The Computer: Devices – Memory – Processing and Networks

Unit II Interaction and Paradigms

Interaction: Models – Frameworks – Ergonomics – Styles – Elements – Interactivity- Paradigms- Interaction Design: Basics – Process – Scenarios – Navigation – Screen Design – Iteration and Prototyping.

Unit III Software Process and Design Rules

HCI in software process: Software life cycle – Usability engineering – Prototyping in practice – Design rationale - Design Rules: Principles- Standards-Guidelines-Rules.

Unit IV Implementation support

Introduction - Elements of Windowing Systems - Programming the Application - Using toolkits - User Interface Management Systems

Unit V Evaluation Techniques

Goals of Evaluation - Evaluation through expert analysis - Evaluation through user participation - Choosing an evaluation method- Universal design: Principles - Multi-model interaction - Designing for diversity.

Text Books

1 Alan Dix, Janet Finlay, Gregory Abowd, Russell Beale,(2004),Human Computer Interaction‖, (3rd Edn), Pearson Education,

References

1 Brian Fling,(2009)Mobile Design and Development, 1st Edn, O'Reilly Media Inc.,

2 Bill Scott and Theresa Neil,(2009),Designing Web Interfaces, (1st Edn), O'Reilly,

Course Code	Course Category	Course Name	L	T	P	Exam (h)	Max Marks			Credits
							CIA	ESE	Total	
Fourth Semester										
204DA2A4CV	Core - XV	Project Work-II	-	-	30	3	50	150	200	16
		Total	-	-	30	-	-	-	200	16
						Grand Total			**2400**	**90**

204DA2A4CV	PROJECT WORK-II	SEMESTER IV

Total Credits: 16

GUIDELINES:

1. A Guide has been allotted to each student by the department. Student can select any topic in discussion with the supervisor. Students should maintain a work diarywere in weekly work carried out has to be written. Guide should review the workevery week and put his/her signature. The work diary along with project reportshould be submitted at the time of viva voce.

2. CA Marks Distribution: A minimum of three reviews have to be done, one at the time finalizing the project title, second at framing questionnaire/identifying the primary data and the third review at the time of commencement of report writing. They should be asked to present the work doneto the respective guide in the three reviews. The guide will give the marks for CIA asper the norms stated below:

First Review	10 Marks
Second Review	10 Marks
Third Review	10 Marks
Document, Preparation and Implementation	20 Marks
Total	**50 Marks**

3. End Semester Examination:The evaluation for the end semester examination should be as per the norms Given Below:

Record work and Presentation	100 Marks
Viva-Voce	50 Marks
Total	**150 Marks**

Note: (End Semester Examination marks jointly given by the external and internal examiner).

BoS Chairman/HoD
Department of Computer Science with Data Analytics
Dr. N. G. P. Arts and Science College
Coimbatore – 641 048

Academic Council
CHAIRMAN