



**Rajagiri College of Social Sciences (Autonomous),
Kalamassery**

**M.Sc. Computer Science (Data Analytics)
Syllabus and Scheme
2022 Admission Onwards**

**BOARD OF STUDIES (COMPUTER SCIENCE)
RAJAGIRI COLLEGE OF SOCIAL SCIENCES (AUTONOMOUS)
KALAMASSERY, KOCHI, 683104
KERALA, INDIA**



Board of Studies

Name and Address	Designation
Dr. Bindiya M Varghese, Rajagiri College of Social Sciences (Autonomous), Kalamassery.	(Chairperson)
Dr. M V Rajesh Maliyeckal, Associate Professor, Department of Electronic Engineering Govt. Model Engineering College, Thrikkakara	(University Nominee)
Dr. Sherly K K, Associate Professor, DEPARTMENT OF INFORMATION TECHNOLOGY, Rajagiri School of Engineering and Technology	Subject Expert nominated by the Academic Council
Prof Biju Skaria, Associate Professor & HOD, Dept. of Computer Applications, M A College of Engineering, Kothamangalam	Subject Expert nominated by the Academic Council
Mr. Jayakrishnan, Co-Founder, Xtend Technologies	Industry Representative
Bharath Jagadeesh, Senior Engineer, Ernst & Young, Carnival Infopark, Phase IV, Kakkanad	Meritorious Alumni
Mr. George Joseph, Rajagiri College of Social Sciences (Autonomous), Kalamassery.	Member
Dr. Jaya Vijayan, Rajagiri College of Social Sciences (Autonomous), Kalamassery.	Member
Ms. Sunu Mary Abraham, Rajagiri College of Social Sciences (Autonomous), Kalamassery.	Member
Dr. Ann Baby, Rajagiri College of Social Sciences (Autonomous), Kalamassery.	Member
Mr.Shiju Thomas M. Y, Rajagiri College of Social Sciences (Autonomous), Kalamassery.	Member
Mr.Sabeen Thomas, Rajagiri College of Social Sciences (Autonomous), Kalamassery.	Member



Preface

Introduction and Scope of the Programme

Data analytics is an essential field that brings together Data, technology, information, statistical analysis all in one platform. Every organization in private/ public sector creates a large volume of data from almost every area. Analyzing that data has huge potential to predict the future of the organization. A good amount of knowledge is very necessary in the field of data management, machine learning, natural language processing as they are the key factors in Data Science. Data analytics will provide the graduates of computer science with the essential requirements that are needed for data science.

These are few of the domains in which data analytics is going to be prominent in:

- Data security: Analytics are already transforming intrusion detection, differential privacy, digital watermarking and malware countermeasures.
- Internet of Things (IoT): Analytics tools and techniques for dealing with the massive amounts of structured and unstructured data generated by IoT will continue to gain importance.
- Finance Domain: Creating newer business models or frameworks that leverages the available data allows financial institutions to monetize data to deliver superior customer value.
- Health Care: Health care analytics allows for the examination of patterns in various healthcare data in order to determine how clinical care can be improved while limiting excessive spending.

Master of Science Programme in Computer Science with specialization on Data Analytics

Trends indicate the dream job of the future is a Data Scientist. The current state of master's programme in computer science is more generalized in nature. The design of the proposed programme is done on the basis of specializing the graduates who have an aptitude in computer science/ mathematics to focus on the data analytics domain. Many Software organizations specifically recruit candidates trained in the tools and algorithms of Data Science.

The two-year course concentrates on the core subjects of computer science in the first two semesters and emphasizes on Data analytics subjects in the second year. The main project which is to be carried in the fourth semester gives the student a live industry experience before they dive into their career.



Revision Suggested for 2022 Syllabus

The existing MScCS(DA) syllabus was proposed in 2017 and got approved by the parent university in 2018. The college got the programme sanctioned in 2020. Hence a pressing need is to revise the syllabus of M.Sc Computer Science(Data Analytics) in order to bring it into line with the industry needs. The revision was deliberated in the BOS of Computer Science and was recommended to the Academic Council.

Semester 1

Course No:	Subject	Change	Change in %
CSDA101	Operating Systems with Linux as Case study	Each Module is Incorporated with Linux Case Studies	5%
CSDA102	Data Structures Using C	No Change	0
CSDA103	Statistics for Data Analytics	Added few more topics to existing syllabus	1%
CSDA104	Database Management Systems with SQL/PL-SQL	Changed the Title	0
CSDA105	Business Intelligence	Changes in the topics	2%
CSDA106	Data Structures Lab	No Change	0
CSDA107	DBMS Lab	No Change	0
Total Change			8%

Semester 2

Course No:	Subject	Change	Change in %
CSDA201	Java programming	Title Change; Removed Applets	1%
CSDA202	Data Communication and Computer networks	No Change	0
CSDA203	Advanced Software Engineering	Contents changed with advanced concepts	5%
CSDA204	Artificial Intelligence	Removed PROLOG	
CSDA205	Data Mining	Added more topics to 5 th module	1%
CSDA206	Java lab	Applet questions removed	1%
CSDA207	Data Mining lab	No change	0
Total Change			8%

Semester 3

Course No:	Subject	Change	Change in %
CSDA301	Data Visualization	No Change	0
CSDA302	Big Data Analytics	Title Change	0
CSDA303	Elective I	No Change	0
CSDA304	Business Modelling & Applied Analytics Using R	Elective II shifted to Semester 4	0
CSDA305	Web Technologies	Introduced new course	100
CSDA306	Python Programming	No change	0

Total Change	100
---------------------	-----

Course No:	Subject	Change	Change in %
CSDA401	Deep Learning	Introduced the new course	100%
CSDA402	Elective II	Shifted from Semester 3	0
CSDA403	Comprehensive Viva-Voce	No change	0
CSDA404	Main project	Duration changed from 4 months to 40 working days.	20%
Total Change			120%

Average Change= Total Percentage Changes made in the syllabus/ Total Number of Courses

$$= 236/24$$

$$=9.8\%$$

Approximately 10% of Syllabus is revised against the existing M.Sc. CS(DA) syllabus was recommended by the BOS and has been approved by the Academic Council.

-sd-

Bindiya M Varghese Ph.D
 Chairperson,
 Board of Studies (Computer Science)
 Rajagiri College of Social Sciences (Autonomous), Kalamassery



Table of Contents

Board of Studies.....	2
Preface	3
Introduction and Scope of the Programme.....	3
Master of Science Programme in Computer Science with specialization on Data Analytics ...	3
Revision Suggested for 2022 Syllabus.....	4
Rajagiri College of Social Sciences (Autonomous)	9
Vision.....	9
Mission.....	9
Department of Computer Science	9
Vision.....	9
Mission.....	9
M.Sc. Computer Science (Data Analytics)	9
Programme Educational Objectives (PEO) of the MSC CS (DA) Programme	10
Programme Outcome (PO)	10
Program Specific Objectives (PSO).....	11
Mapping of PO to PEO	11
Eligibility Criteria.....	12
Admission	12
Programme Structure and Duration.....	12
Attendance	13
Condonation.....	13
Promotion.....	13
Grading	13
Evaluation.....	13
Direct Grading.....	14
Grade Point Average (GPA).....	14
Internal Evaluation for Regular Programme.....	14
Components of Internal (CE) and External Evaluation (ESE)	14
For Theory (CE) [Internal].....	14
For Theory (ESE) [External]	14
Pattern of question for practical	15
For Practical (CE) [Internal]	15
For Practical (ESE) [External]	15
For Internship/ Main Project (CE) [Internal].....	15

For Internship/ Main Project (ESE) [External]	15
Comprehensive viva – voce (CE) [Internal]	16
Comprehensive viva – voce (ESE) [External]	16
External Evaluation	16
Direct Grading System	16
Performance Grading.....	17
Bridge Courses.....	19
Semester Courses	19
Elective courses	21
Bridge courses.....	23
Programming concepts using C language	23
Basic Mathematics and Statistics.....	23
Digital Logic and Computer Organization	23
Basic Principles of Management	24
Semester I	25
CSDA101 Operating Systems with Linux as Case study	25
CSDA102 Data Structures using C	26
CSDA103 Statistics for Data Analytics	27
CSDA104 Database Management Systems with SQL/PL-SQL	28
CSDA105 Business Intelligence.....	29
CSDA106 Data Structures Lab	30
CSDA107 DBMS Lab.....	31
Semester II.....	32
CSDA201 Java Programming.....	32
CSDA202 Data Communications and Computer Networks.....	33
CSDA203 Advanced Software Engineering	34
CSDA204 Artificial Intelligence	35
CSDA205 Data Mining.....	35
CSDA206 Java Programming Lab.....	37
CSDA207 Data Mining lab.....	38
Semester III.....	39
CSDA301 Data Visualization	39
CSDA302 Big Data Analytics	40
CSDA303 (1) Data Warehousing.....	41
CSDA303 (2) Digital Image Processing	42



CSDA304 Business Modelling & Applied Analytics Using R.....	43
CSDA305 Web Technology Lab.....	44
CSDA306 Data Analytics using Python	45
Semester IV	47
CSDA401 Deep Learning.....	47
CSDA402 (1) Information Retrieval Techniques	48
CSDA402 (2) Social Media Mining.....	49
CSDA403 Comprehensive Viva Voce	50
MCA404 Main Project	50



Rajagiri College of Social Sciences (Autonomous)

Rajagiri College of Social Sciences (Autonomous) was established as a result of the indefatigable industry and foresight of the CMI. The various axioms of the institution maintain the axiomatic spirit of Rajagiri - 'Relentlessly Towards Excellence'. Rajagiri College of Social Sciences (RCSS) is the eldest child of Rajagiri Vidyapeetham (Rajagiri group of educational institutions). It is located on two picturesque campuses- the Hill Campus at Kalamassery and the Valley campus at Kakkanad in Kochi, in the state of Kerala. RAJAGIRI literally means "The hill of the King" and derivatively it refers to the hillock where Jesus Christ is accepted as the King or the model, as the human embodiment of the virtues of love, truth and justice.

Vision

To become a centre par excellence of learning, unique in experience, value based in its approach, and pioneering in its efforts for enriching and fulfilling LIFE.

Mission

To facilitate comprehensive and integrated development of individuals imbued with righteousness and courage of conviction, to effectively function as social beings

Department of Computer Science

The Department commenced its Master of Computer Applications (MCA) Programme in 2001, Affiliated to Mahatma Gandhi University, Kottayam, Kerala, India, and approved by the All India Council for Technical Education, New Delhi. The programme is designed to meet the ever-growing demand for well qualified and trained computer programmers. The Department has been consecutively rated "Platinum" in every AICTE-CII survey, for its illustrious placement track record and distinctive industry interaction programmes.

The Department offers a two-year MCA programme and an MSc. Computer Science programme with data analytics specialization

Vision

To create technically competent individuals, who are innovative and uphold human values.

Mission

To develop globally recognised competent and innovative IT professionals, committed to lifelong learning, and blended with social commitment through comprehensive programmes.

M.Sc. Computer Science (Data Analytics)

The programme adheres strictly to an academic schedule that creates a strong knowledge base in the programme. Students are given training in the current technologies and an in-depth understanding of the current environment through various activities. The activities include extracurricular pursuits and value-added



programmes that foster awareness in contemporary technologies which help in the holistic development of students.

The Rajagiri Immersive Learning Experience Methodology is a pedagogical innovation of the institute. This methodology aims to evolve the Rajagiri student into a socially responsible professional. The Immersive Learning methodology is based on four dimensions – conceptual learning, experiential engagement, executive modelling and corporate competency.

The faculty keep themselves abreast with the latest trends in technology through regular online and onsite /offsite training programmes. Prominent industry professionals and academicians conduct workshops and lectures for the faculty. Faculty are regularly sent for Faculty Development Programmes and other courses to enhance their knowledge base. Rajagiri governance policy encourages a creative and flexible atmosphere where all individuals are respected and valued by the community.

Programme Educational Objectives (PEO) of the MSC CS (DA) Programme

- PEO1 : Graduates of the program will be computer professionals of probity, positive attitude and scientific temper
- PEO2 : Graduates of the program will have sound theoretical knowledge and skill for development and implementation of Software pertaining to the Data Analytics
- PEO3 : Graduates of the program will possess good communication, technical and innovative skills
- PEO4 : Graduates of the program will have a sense of social awareness

Programme Outcome (PO)

1. Computational Knowledge:

Apply knowledge of computing fundamentals, computing specialisation, mathematics, and domain knowledge appropriate for the computing specialisation to the abstraction and conceptualisation of computing models from defined problems and requirements.

2. Problem Analysis:

Identify, formulate, research literature, and solve *complex* computing problems reaching substantiated conclusions using fundamental principles of mathematics, computing sciences, and relevant domain disciplines.

3. Design /Development of Solutions:

Design and evaluate solutions for *complex* computing problems, and design and evaluate systems, components, or processes that meet specified needs with appropriate consideration for public health and safety, cultural, societal, and environmental considerations.

4. Conduct Investigations of Complex Computing Problems:



Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. Modern Tool Usage:

Create, select, adapt and apply appropriate techniques, resources, and modern computing tools to complex computing activities, with an understanding of the limitations.

6. Professional Ethics:

Understand and commit to professional ethics and cyber regulations, responsibilities, and norms of professional computing practice.

7. Life-long Learning:

Recognise the need, and have the ability, to engage in independent learning for continual development as a computing professional.

8. Communication Efficacy:

Communicate effectively with the computing community, and with society at large, about *complex* computing activities by being able to comprehend and write effective reports, design documentation, make effective presentations, and give and understand clear instructions.

9. Societal and Environmental Concern:

Understand and assess societal, environmental, health, safety, legal, and cultural issues within local and global contexts, and the consequential responsibilities relevant to professional computing practice.

10. Individual and Team Work:

Function effectively as an individual and as a member or leader in diverse teams and in multidisciplinary environments.

Program Specific Objectives (PSO)

PSO1: Data Analytics: Acquire knowledge of Data pre-processing and Data quality, Modelling and design of data warehouses, Algorithms for data mining, skills to design, analyse and develop algorithms and implement using high-level programming languages and to define and critically analyse mining approaches for various domains.

Mapping of PO to PEO

Program Educational Objectives	PEO1	PEO2	PEO3	PEO4
Program Outcomes				
PO1: Computational Knowledge		√		



PO2: Problem Analysis			√	
PO3: Design /Development of Solutions	√			
PO4: Conduct Investigations of Complex Computing Problems			√	
PO5: Modern Tool Usage			√	
PO6: Professional Ethics	√			√
PO7: Life-long Learning	√			
PO8: Communication Efficacy			√	
PO9: Societal and Environmental Concern				√
PO10: Individual and Team Work	√			√
PSO1: Data Analytics		√	√	

Eligibility Criteria

The eligibility for admission to M Sc Computer Science (Data Analytics) programme is a B Sc Degree with Mathematics/Computer Science /Electronics as one of the subjects (Main or Subsidiary) or BCA/B.Tech degree with not less than 55% marks in optional subjects.

Note: Candidates having degree in computer science/Computer Application/IT/Electronics shall be given a weightage of 20% in their qualifying degree examination marks considered for ranking for admission to M Sc. Computer Science (Data Analytics).

Reservation policy will be as regulated by parent University.

Admission

The admission to the M.Sc. programme shall be based on one-hour Entrance Examination conducted by Rajagiri College of Social Sciences, Kalamassery, Academic performance and Personal Interview.

Programme Structure and Duration

The duration of the programme shall be 4 semesters. The duration of each semester shall be 90 working days. Odd semesters from June to October and even semesters from November to March.

A student may be permitted to complete the programme, on valid reasons, within a period of 8 continuous semesters from the date of commencement of the first semester of the programme.

Attendance

The minimum requirement of attendance for each course during a semester for appearing at the end-semester examination shall be 75%. Condonation of shortage of attendance to a maximum of 15 days in a semester subject to a maximum of two times during the whole period of the programme may be granted by the Principal, Rajagiri College of Social Sciences (Autonomous), Kalamassery.

Those who could not register for the examination of a particular semester due to shortage of attendance may repeat the semester along with junior batches, without considering sanctioned strength, subject to the existing Rules of the institution.

A Regular student who has undergone a programme of study under earlier regulation/scheme and could not complete the Programme due to shortage of attendance may repeat the semester along with the regular batch subject to the condition that he has to undergo all the examinations of the previous semesters as per the 2020 Regulations

A student who had sufficient attendance and could not register for fourth semester examination can appear for the end semester examination in the subsequent years with the attendance and progress report from the Principal.

Condonation

As per the regulations of Examination Manual, Rajagiri College of Social Sciences, Kalamassery.

Promotion

A student who registers for a particular semester examination shall be promoted to the next semester.

A student having 75% attendance for each course and who fails to register for examination of a particular semester will be allowed to register notionally and is promoted to the next semester, provided application for notional registration shall be submitted with 15 days from the commencement of the next semester.

Grading

There shall be a Semester Examination at the end of each semester for all credit courses of duration of 3 hours. A question paper for a theory course may contain short answer type/annotation and long essay type questions. Different types of questions shall have different weightage.

Evaluation

The evaluation scheme for each course shall contain two parts; (a) End Semester Evaluation (SE) [External Evaluation] and (b) Continuous Evaluation (CE) [Internal Evaluation]. 25% weightage shall be given to internal evaluation and the remaining 75% to external evaluation and the ratio and weightage between internal and external is 1:3.



Both End Semester Evaluation (SE) and Continuous Evaluation (CE) shall be carried out using direct grading system.

Direct Grading

The direct grading for CE (internal) and ESE (external evaluation) shall be based on 6 letter grades (A+, A, B, C, D and E) with numerical values of 5, 4, 3, 2, 1 and 0 respectively.

Grade Point Average (GPA)

Internal and External components are separately graded and the combined grade point with weightage 1 for internal and 3 for external shall be applied to calculate the Grade Point Average (GPA) of each course. Letter grade shall be assigned to each course based on the categorization provided in 12.16.

Internal Evaluation for Regular Programme

The internal evaluation shall be based on predetermined transparent system involving periodic written tests, assignments, seminars, lab skills, records, viva-voce etc.

Components of Internal (CE) and External Evaluation (ESE)

Grades shall be given to the evaluation of theory / practical / project / comprehensive viva-voce and all internal evaluations are based on the Direct Grading System. There shall be no separate minimum grade point for internal evaluation.

The model of the components and its weightages for Continuous Evaluation (CE) and the End Semester Evaluation (ESE) are shown in below:

For Theory (CE) [Internal]

	Components	Weightage
i.	Assignment	1
ii.	Seminar	2
iii.	Two test papers	2 (1 each)
Total		5

(For test papers all questions shall be set in such a way that the answers can be awarded A+, A, B, C, D, E grade).

For Theory (ESE) [External]

Evaluation is based on the pattern of question specified as follows.

Questions shall be set to assess knowledge acquired, standard, and application of knowledge, application of knowledge in new situations, critical evaluation of knowledge and the ability to synthesize knowledge. Due weightage shall be given to each module based on content/teaching hours allotted to each module. The question setter shall

ensure that questions covering all skills are set. The question shall be prepared in such a way that the answers can be awarded A+, A, B, C, D, E grades.

Sl. No	Type of questions	Weight	Number of questions to be answered
1.	Short Answer type questions	1	10 out of 12
2.	Long essay type questions	4	5 EITHER/OR Questions. (One each from 5 modules)
		5	Total Weightage =30

Pattern of question for practical

The pattern of questions for external evaluation of practical shall be prescribed by the Board of Studies.

For Practical (CE) [Internal]

Components	Weightage
Written /Lab test	2
Lab involvement and record	1
Viva	2
Total	5

For Practical (ESE) [External]

Components	Weightage
Written /Lab test	7
Lab involvement and record	3
Viva	5
Total	15

For Internship/ Main Project (CE) [Internal]

Components	Weightage
Interim presentation on Internship	2
Internship Interim Report	2
Internship Evaluation at the Organization by Internal Faculty	1
Total	5

For Internship/ Main Project (ESE) [External]

Components	Weightage
Final Presentation	3
Internship Final Report	7
Internship Evaluation at the Organization by Organization	5



Total	15
--------------	-----------

Comprehensive viva – voce (CE) [Internal]

Components	Weightage
Comprehensive viva-voce (all courses from first semester to fourth semester)	5
Total	5

Comprehensive viva – voce (ESE) [External]

Components	Weightage
Comprehensive viva-voce (all courses from first semester to fourth semester)	15
Total	15

All grade point averages shall be rounded to two digits.

To ensure transparency of the evaluation process, the internal assessment grade awarded to the students in each course in a semester shall be published on the notice board at least one week before the commencement of external examination.

There shall not be any chance of improvement for internal grade.

External Evaluation

The external examination in theory courses is to be conducted by the Examination Cell at the end of the semester. The answers may be written in English. The evaluation of the answer scripts shall be done by examiners based on a well-defined scheme of valuation. The external evaluation shall be done immediately after the examination preferably through Centralized valuation.

Photocopies of the answer scripts of the external examination shall be made available to the students on request as per the rules prevailing in the Examination Manual of the College.

The question paper should be strictly on the basis of model question papers set and directions prescribed by the BOS.

Direct Grading System

Direct Grading System based on a 6-point scale is used to evaluate the Internal and External examinations taken by the students for various courses of study.

Grade	Grade Points	Range
A+	5	4.50 to 5.00
A	4	4.00 to 4.49
B	3	3.00 to 3.99
C	2	2.00 to 2.99
D	1	0.01 to 1.99

E	0	0.00
---	---	------

Performance Grading

Students are graded based on their performance (GPA/ SGPA/CGPA) at the examination on a 7-point scale as detailed below:

CGPA	Grade	Indicator
4.50 to 5.00	S	Outstanding
4.00 to 4.49	A	Excellent
3.50 to 3.99	B	Very good
3.00 to 3.49	C	Good
2.75 to 2.99	D	Fair
2.50 to 2.74	E	Marginal (pass)
Upto 2.49	F	Deficient (fail)

No separate minimum is required for internal evaluation for a pass, but a minimum E grade is required for a pass in an external evaluation. A minimum E grade is required for pass in a course.

A student who fails to secure a minimum grade for a pass in a course will be permitted to write the examination along with the next batch.

Semester Grade Point Average (SGPA) and Cumulative Grade Point Average (CGPA) Calculations. The **SGPA** is the ratio of sum of the credit points of all courses taken by a students in the semester to the total credit for that semester, After the successful completion of a semester, Semester Grade Point Average (SGPA) of a student in that semester is calculated using the formula given below:

$$\text{Semester Grade Point Average - SGPA (S}_j\text{)} = \frac{\sum (C_i \times G_i)}{\sum C_i}$$

(SGPA = Total credit point awarded in a semester / Total credits of the semester)

Where '**S_j**' is the **jth** semester, '**G_i**' is the grade point scored by the student in the **ith** course '**c_i**' is the credit of the **ith** course.

Cumulative Grade Point Average (CGPA) of a Programme is calculated using the formula.

$$\text{Cumulative Grade Point Average (CGPA)} = \frac{\sum (C_i \times S_i)}{\sum C_i}$$

(CGPA = Total credit points awarded in all semesters / Total credits of the programme)

Where ' C_i ' is the credits for the i^{th} semester, ' S_i ' is the SGPA for the i^{th} semester. The **SGPA** and **CGPA** shall be rounded off to 2 decimal points.

For the successful completion of semester, a student shall pass all courses and score a minimum **SGPA** of 2.5. However, a student is permitted to move to the next semester irrespective of her/his **SGPA**.



Bridge Courses

#	Course Name	Hours
1	Programming concepts using C language	14
2	Basic Mathematics and Statistics	6
3	Digital Logic and Computer Organization	4
4	Basic Principles of Management	4
5	Skill and Personality Development Workshop	4 Full Working Days
	Evaluation after Bridge Course	2 Hour Theory Exam and 3 Hour Lab Exam

Semester Courses

Semester I							
Code	Course Name	Type	Hours			Exam (hours)	Credit
			Lecture	Tutorial	Practical		
CSDA101	Operating Systems with Linux as Case study	Core	3	1	0	3	3
CSDA102	Data Structures Using C	Core	3	1	0	3	3
CSDA103	Statistics for Data Analytics	Core	3	1	0	3	3
CSDA104	Database Management Systems with SQL/PL-SQL	Core	3	1	0	3	3
CSDA105	Business Intelligence	Core	3	1	0	3	4
CSDA106	Data Structures Lab	Lab			4	3	2
CSDA107	DBMS Lab	Lab			4	3	2
	Semester 1: Total Credits	20					

Semester II							
Code	Course Name	Type	Hours			Exam (hours)	Credit Total
			Lecture	Tutorial	Practical		
CSDA201	Java programming	Core	3	1	0	3	3
CSDA202	Data Communication and Computer networks	Core	3	1	0	3	3
CSDA203	Advanced Software Engineering	Core	3	1	0	3	3
CSDA204	Artificial Intelligence	Core	3	1	0	3	3
CSDA205	Data Mining	Core	3	1	0	3	4
CSDA206	Java Programming Lab	Lab			4	3	2
CSDA207	Data Mining lab	Lab			4	3	2
Semester 2 : Total Credits		20					

Semester III							
Code	Course Name	Type	Hours			Exam (hours)	Credit
			Lecture	Tutorial	Practical		
CSDA301	Data Visualization	Core	3	1	0	3	4
CSDA302	Big Data Analytics	Core	3	1	0	3	4
CSDA303(x)	Elective I	Elective	3	1	0	3	3
CSDA304	Business Modelling & Applied Analytics Using R	Core	3	1	0	3	4
CSDA305	Web Technologies Lab	Lab			6	3	3
CSDA306	Data Analytics using Python	Lab	2		4	3	2
Semester 3 : Total Credits		20					

Semester IV							
Code	Course Name	Type	Hours			Exam (hours)	Credit
			Lecture	Tutorial	Practical		
CSDA401	Deep Learning	Core	3		3	3	4
CSDA402(x)	Elective II	Core	3	1	0	3	4
CSDA403	Comprehensive Viva-Voce	Exam					2
CSDA404	Main project		40 Working days				10
	Semester 4: Total Credits	20					
	Grand Total Credits for MSc Computer Science (Data Analytics)	80					

Elective courses

CSDA303 (1) Data Warehousing

CSDA303 (2) Digital Image Processing

CSDA402 (1) Information Retrieval Techniques

CSDA402 (2) Social Media Mining



Bridge courses

Programming concepts using C language

- Work with primitive types and expressions
- Understanding the basic structure of a C Program, the main function and using standard I/O
- Understand C Character Set and Tokens, Data Types, Variables and Constants
- Work with Operators and Expressions
- Control the flow of Program using Conditional statements and Loops
- Work with Built-in Functions and User Defined Functions with arguments, Passing arguments By Value and By Reference
- Work with Arrays and Strings
- Understand and Implement Pointers, Pointer to an array, Array of pointers, Pointers and functions
- Defining Structures and Union
- Data File Handling
- Debug C programs effectively

Basic Mathematics and Statistics

- Set theory
Sets and their representations; The empty set; finite and infinite sets; equal and equivalent sets; subsets; power set; universal set; Venn diagrams; complement of a set operation on sets; applications of sets.
- Mathematical Logic
Basic Logical connections; Conjunction; Disjunction; Negation; Negation of Compound Statements; Truth tables. Tautologies; Logical Equivalence; Applications.
- Modern algebra
Binary Operation; Addition Modulo n ; Multiplication modulo n
- Matrices and Determinants
Definition of a matrix; Operations on matrices; Square Matrix and its inverse; determinants; the inverse of a matrix
- Basics Statistics
Measures of central Tendency; Standard Deviation; Discrete series. variance.

Digital Logic and Computer Organization

- **Computer Evolution:** Brief history of Computer, Classification of Computer, Structure of a Computer System, Arithmetic Logic Unit, Control Unit, Bus Structure, Von Neumann Architecture. Bootstrapping.
- **Number systems** - Decimal, Binary, Octal, Hexadecimal conversion from one to another, Basic Arithmetic Operations: Integer Addition and Subtraction, Signed numbers, Binary Arithmetic, 1's and 2's Complement Arithmetic, Fixed and Floating point numbers, Floating point representation.,
- **Digital Logic:** Logic gates, Boolean Algebra, Basic theorem and Properties of Boolean algebra. Basic concepts on Combinational Circuits and Sequential circuits

Basic Principles of Management

- Introduction to principles of management: Planning, organising, staffing, Budgeting, Controlling.
- Scope of IT applications in management, and its benefits
- Socioeconomic environment and information systems, and its impact
- Strategic role of IT in organisations
- Critical success factors as its role in implementing IT applications
- Case studies of successful / failed IT applications

Semester I

CSDA101 Operating Systems with Linux as Case study

Syllabus	
I	<p>File System concepts, Access methods, Allocation methods, Directory systems, File protection.</p> <p>Disk Management - Disk scheduling, Disk management, Disk reliability.</p> <p>Linux:History of Linux: Linux Operating System Layers, The Linux Shell Process: (parent and child processes), Files and Directories (File Structure and directory structure), Linux Basic commands: pwd, cd, mkdir, rm, mv, touch,man,cp,locate, echo, cat, touch, ls, cut, paste and other basic shell management commands</p>
II	<p>Memory Management Memory Management, Memory partitioning, Swapping, Paging, Segmentation, Virtual memory, Demand paging, Page replacement algorithms, Allocation algorithms</p> <p>Linux Commands: df, du, tar,zip, uname, chmod, head, tail,sort, grep, sudo privileges, top, free, vmstat, and other memory related commands. Installation of Linux OS</p>
III	<p>Process Management and Concurrency management Process and Thread Management, Concept of process and threads, Process states, Process management, Context switching, Multithreading, Concurrency Control, Concurrency and Race Conditions</p> <p>Linux: process related commands: fork, exec, ps, kill,nice, foreground process, background process</p>
IV	<p>Concurrency Management Semaphores, Classical IPC problems and solutions. Deadlock, Characterization, Avoidance and Prevention, Detection, Recovery</p> <p>Linux: Shell variables, redirection, filters Shell Scripting,</p>
V	<p>Protection and case STUDY: LINUX Access matrix, Implementation of access matrix, Revocation of access rights.</p> <p>Linux OS - Administering Users and Groups: Administering User Accounts, Working with Group Accounts, Understanding the Root Account, installing packages</p>
REFERENCE BOOKS:	
Silberschatz, Galvin, and Gagne, "Operating System Concepts", Eighth Edition, Wiley Publication, 2011.	

Andrew S. Tanenbaum, "Modern Operating Systems", Second Edition, Pearson Education, 2004.
Gary Nutt, "Operating Systems", Third Edition, Pearson Education, 2004
Harvey M. Deital, "Operating Systems", Third Edition, Pearson Education, 2004.
Milan Milenkovic, "Operating Systems: Concept and Design", 2nd Edition, 2001.
"Linux Command Line and Shell Scripting Bible (English) 2nd Edition", Wiley Publication.
Richard Petersen, "Linux: The Complete Reference", Sixth Edition, 2007

CSDA102 Data Structures using C

Syllabus	
I	<p>Module 1: Introductory Concepts Basics of C language Variables, Data types, Conditional and Loop Structures, Pointers. Introduction to Data structures, Definition, Classification of data structures : primitive and non primitive Operations on data structures. Dynamic memory allocation and pointers, Definition Accessing the address of a variable, Declaring and initializing pointers. Accessing a variable through its pointer. Meaning of static and dynamic memory allocation. Memory allocation functions : malloc, calloc, free and realloc.</p>
II	<p>Module 2: Linear Data structures Stack – Definition, Array representation of stack, Operations on stack: Infix, prefix and postfix notations Conversion of an arithmetic expression from Infix to postfix. A Applications of stacks. Queue - Definition, Array representation of queue, Types of queue: Simple queue, circular queue, double ended queue (deque) priority queue, operations on all types of Queues</p>
III	<p>Module 3: Searching and Sorting techniques Searching and Sorting Search, Basic Search Techniques: Search algorithm searching techniques: sequential search, Binary search – Iterative and Recursive methods. Comparison between sequential and binary search Sort, General Background, Definition, different types: Bubble sort, Selection sort, Merge sort, Insertion sort, Quick sort</p>
IV	<p>Module 4: Non-linear Data Structures -Linked list Definition, Components of linked list, Representation of linked list, Advantages and Disadvantages of linked list. Types of linked list : Singly linked list, Doubly linked list, Circular linked list and circular doubly linked list. Operations on singly linked list creation, insertion, deletion, search and display</p>
V	<p>Module 5: Trees and Graphs.</p>

	<p>Tree - Definition: Tree, Binary tree, Complete binary tree, Binary search tree, Heap Tree terminology : Root, Node, Degree of a node and tree, Terminal nodes, Nonterminal nodes, Siblings, Level, Edge, Path, depth, Parent node, ancestors of a node. Binary tree : Array representation of tree, Creation of binary tree. Traversal of Binary Tree : Preorder, Inorder and postorder.</p> <p>Graphs: Graphs – terminology, Representation, Graph traversals (dfs & bfs)</p>
REFERENCE BOOKS:	
Introduction to Algorithms - Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest	
Fundamentals of data structures – Ellis Horowitz and SartajSahni (Galgotia , 1994)	
Fundamentals of computer algorithms- Ellis Horowitz, SartajSahni, SanguthevarRajeshkharan (Universities Press , 2007)	
Data Structure using C & C++ b, Tannenbaum and Augustine,prentice hall.	
Data Structures – a pseudocode approach with C –Richard F Gilberg, Behrouz A Forouzan, Thomson Learning, 2 Edn., Cengage Learning C2005	
Data Structures and program design – R. L Kruse (Prentice Hall of India),C2001	

CSDA103 Statistics for Data Analytics

Syllabus	
I	<p>Module 1:-Basic Statistics</p> <p>Measures of central tendency: - mean, median, mode; Measures of dispersion: Range, Mean deviation, Quartile deviation and Standard deviation; Moments, Skewness and Kurtosis, Correlation - Linear correlation, Karl Pearson's coefficient of Correlation, Rank correlation, Linear regression.</p>
II	<p>Module 2:- Probability Theory</p> <p>Sample space, Events, Different approaches to probability, Addition and multiplication theorems on probability, Independent events, Conditional probability, Bayes Theorem</p>
III	<p>Module 3:- Random variables and Distribution</p> <p>Random variables, Probability density functions and distribution functions, Marginal density functions, Joint density functions, mathematical expectations, moments and moment generating functions. Discrete probability distributions - Binomial, Poisson distribution, Continuous probability distributions- uniform distribution and normal distribution.</p>
IV	<p>Module 4:- Sampling and Estimation</p>

	<p>Theory of Sampling: - Population and sample, Types of sampling Theory of Estimation: - Introduction, point estimation, methods of point estimation-Maximum Likelihood estimation and method of moments, Central Limit Theorem (Statement only).</p>
V	<p>Module 5:-Testing of hypothesis</p> <p>Null and alternative hypothesis, types of errors, level of significance, critical region, Large sample tests – Testing of hypothesis concerning mean of a population and equality of means of two populations Small sample tests – t Test for single mean, difference of means. Paired t-test, Chi-square test (Concept of test statistic $n\sigma^2$), F test - test for equality of two population variances. ANOVA – one-way & two-way classification</p>
REFERENCE BOOKS:	
Fundamentals of statistics: S.C.Gupta, 6th Revised and enlarged edition April 2004, Himalaya Publications.	
Introduction to Probability and Statistics, Medenhall, Thomson Learning , 12 Edn.	
Fundamentals of Mathematical Statistics- S.C.Gupta ,V.K.Kapoor. Sultan Chand Publications.	
Introduction to Mathematical Statistics -Robert V. Hogg &Allen T. Craig. Pearson education.	

CSDA104 Database Management Systems with SQL/PL-SQL

Syllabus	
I	<p>Module 1: Introductory concepts of DBMS</p> <p>Introduction and applications of DBMS, Purpose of data base, Data, Independence, Database System architecture- levels, Mappings, Database, users and DBA Relational Model : Structure of relational databases, Domains, Relations, Entity-Relationship model Basic concepts, Design process, constraints, Keys, Design issues, E-R diagrams, weak entity sets, extended E-R features – generalization, specialization, aggregation, reduction to E-R database schema</p>
II	<p>Module 2: Relational Database design</p> <p>Functional Dependency – definition, trivial and non-trivial FD, closure of FD set, closure of attributes, irreducible set of FD, Normalization – 1NF, 2NF, 3NF, Decomposition using FD-dependency preservation, BCNF, Multivalued dependency, 4NF, Join dependency and 5NF</p>
III	<p>Module 3: SQL Concepts</p> <p>Basics of SQL, DDL,DML,DCL, structure – creation, alteration, defining constraints – Primary key, foreign key, unique, not null, check, IN operator, Functions - aggregate functions, Built-in functions – numeric, date, string functions, set operations, sub-</p>

	queries, correlated sub-queries, Use of group by, having, order by, join and its types, Exist, Any, All, view and its types. transaction control commands – Commit, Rollback, Savepoint
IV	Module 4: PL/SQL Introduction to PL/SQL, PL/SQL Identifiers, Control Structures, Composite Data Types, Explicit Cursors, Stored Procedures and Functions, Triggers, Compound, DDL, and Event Database Triggers
V	Module 5: Transaction Management Transaction concepts, properties of transactions, serializability of transactions, testing for serializability, System recovery, Two-Phase Commit protocol, Recovery and Atomicity, Log-based recovery, concurrent executions of transactions and related problems, Locking mechanism, solution to concurrency related problems, deadlock, two-phase locking protocol, Isolation, Intent locking
REFERENCE BOOKS:	
Database Management Systems – Raghu Ramakrishnan and Johannes Gehrke, Third Edition, McGraw Hill, 2003	
Database Systems: Design, Implementation and Management, Peter Rob, Thomson Learning, 7Edn.	
Concept of Database Management, Pratt, Thomson Learning, 5Edn.	
Database System Concepts – Silberchatz, Korth and Sudarsan, Fifth Edition, McGraw Hill, 2006	
The Complete Reference SQL – James R Groff and Paul N Weinberg	

CSDA105 Business Intelligence

Syllabus	
I	Business Intelligence - Introduction, Definition, History, and Evolution, Business Intelligence Segments, Difference between Information and Intelligence, Defining Business Intelligence Value Chain, Factors of Business Intelligence System, Real time Business Intelligence, Business Intelligence Applications, Roles of Business Intelligence in Modern Business, Challenges of Business Intelligence.
II	Business Intelligence Types: Introduction, Creating Business Intelligence Environment, Business Intelligence Landscape, Types of Business Intelligence, Business Intelligence Platforms, Dynamic roles in Business Intelligence, Roles of Business Intelligence in Modern Business, Information Workers, Architecting Data and Enterprise Data Models, Total Data Quality Management (TDQM).
III	Business Intelligence Life-cycle: Introduction, Enterprise Performance Life Cycle (EPLC) Framework Elements, Life Cycle Phases, Human Factors in BI Implementation, BI Strategy, Objectives and Deliverables, Transformation Roadmap, Building a transformation roadmap, BI Development Stages and Steps, Parallel Development Tracks, BI Framework, Advanced/Emerging BI Technologies, Future of Business Intelligence.

IV	Knowledge Management: Introduction, Characteristics of Knowledge Management, Knowledge assets, Generic Knowledge Management Process, Knowledge Management Technologies, Essentials of Knowledge Management Process, Data Extraction: Introduction, Data Extraction, Role of ETL process, Importance of source identification, Various data extraction techniques, Logical extraction methods, Physical extraction methods, Change data capture.
V	Business Intelligence Issues and Challenges: Introduction, Critical Challenges for Business Intelligence success, Cross-Organizational Partnership, Business Sponsors, Dedicated Business Representation, Availability of Skilled Team Members, Business Intelligence Application Development methodology, Planning the BI Projects, Business Analysis and Data Standardization, effect of Dirty Data on Business profitability, Importance of Meta-Data, Silver Bullet Syndrome, Customer Pain Points, Creating Cost Effective Enterprise friendly BI solution

REFERENCE BOOKS:

Business Intelligence Guidebook: From Data Integration to Analytics by Rick Sherman

Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications by Larissa T. Moss and Shaku Atre

The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling by Ralph Kimball and Margy Ross

Successful Business Intelligence, Second Edition: Unlock the Value of BI & Big Data by Cindi Howson

Business Intelligence for Dummies by Swain Scheps

Successful Business Intelligence by Cindi Howson

Relentlessly Practical Tools for Data Warehousing and Business Intelligence by Ralph Kimball

Business Intelligence: Practices, Technologies, and Management, Rajiv Sabherwal, Irma Becerra-Fernandez

Predictive Business Analytics: Forward Looking Capabilities to Improve Business Performance, Lawrence Maisel, Gary Cokins

CSDA106 Data Structures Lab**Syllabus**

- | | |
|---|---|
| I | <ol style="list-style-type: none"> 1. Program to represent Searching procedures (Linear search and Binary search) 2. Program to represent sorting procedures (Selection, Bubble , Insertion) 3. Polynomial addition using array 4. Polynomial multiplication using array 5. Program to represent sparse matrix manipulation using arrays. 6. Program to allocate two dimensional arrays dynamically. 7. Program to demonstrate the use of realloc(). 8. Represent Graph using array 9. Stack using array 10. Reverse a string using stack 11. Implement Queue using array 12. Circular Queue using array |
|---|---|

	13. Double ended queue using array
II	<ol style="list-style-type: none"> 1. Program to represent Singly Linked List. 2. Program to represent Doubly Linked List. 3. Program to represent Circular Linked List. 4. Polynomial addition using Linked List. 5. Polynomial multiplication using linked list. 6. Implement a linked stack 7. Program to represent Queue using linked list 8. Represent a graph using linked list. 9. Program for Conversion of infix to postfix. 10. Program for Evaluation of Expressions. 11. Program for binary search tree using recursion. 12. Program to represent Binary search Tree Traversals without recursion
REFERENCE BOOKS:	
Fundamentals of Data Structures in C by Horowitz, Sahni and Anderson-Freed.	
Data Structures Through C in Depth by S.K Srivastava, Deepali Srivastava.	
Data Structures Using C Aaron M. Tenenbaum	
Data Structures Using C, Reema Thareja	

CSDA107 DBMS Lab

Syllabus	
I	INTRODUCTION TO SQL Data Definition, Constraints, and Schema Changes, Data Types Create Schema Create Table Drop Table Alter Table Drop a Column (An Attribute)
II	BASIC QUERIES IN SQL Aliases, * and Distinct, Empty Where-Clause Unspecified Where-Clause Use of Distinct Set Operations Nesting of Queries
III	THE EXISTS FUNCTION Explicit Sets Nulls in SQL Queries Aggregate Functions Grouping The Having-Clause Substring Comparison
IV	ARITHMETIC OPERATIONS Order by Specifying Updates in SQL- Insert, Delete, Update
V	SQL TRIGGERS

	Views in SQL Procedures, Functions, Cursors DB Connectivity to any Front End platform
REFERENCE BOOKS:	
Database Management Systems – Raghu Ramakrishnan and Johannes Gehrke, Third Edition, McGraw Hill, 2003	
Database Systems: Design , Implementation and Management, Peter Rob, Thomson Learning, 7Edn.	
Concept of Database Management, Pratt, Thomson Learning, 5Edn.	
Database System Concepts – Silberchatz, Korth and Sudarsan, Fifth Edition, McGraw Hill, 2006	
The Complete Reference SQL – James R Groff and Paul N Weinberg	

Semester II

CSDA201 Java Programming

Syllabus	
I	Basics of Java: Java - What, Where and Why?, History and Features of Java, Internals of Java Program, Difference between JDK,JRE and JVM, Internal Details of JVM, Variable and Data Type, Unicode System, Naming Convention. OOPS Concepts: Advantage of OOPs, Object and Class, Method Overloading, Constructor, static variable, method and block, this keyword, Inheritance (IS-A), Aggregation and Composition(HAS-A), Method Overriding, Covariant Return Type, super keyword, Instance Initializer block, final keyword, Runtime Polymorphism, static and Dynamic binding, Abstract class and Interface, Downcasting with instanceof operator ,Package and Access Modifiers, Encapsulation, Object class, Object Cloning, Java Array, Call By Value and Call By Reference
II	Core java Features: String Handling, Exception Handling, Nested classes, Packages and Interfaces. Multithreaded Programming – synchronization, Input/Output – Files – Directory ,Utility Classes, Generics, Generic Class, Generic methods.
III	Serialization: Serialization & Deserialization, Serialization with IS-A and Has-A, Transient keyword. Networking: Socket Programming, URL class, Displaying data of a web page, InetAddress class, DatagramSocket and DatagramPacket, Two way communication
IV	JDBC: - Overview, JDBC implementation, Connection class, Statements, Catching Database Results, handling database Queries. Error Checking and the SQLExceptionClass , The SQLWarning Class, JDBC Driver Types, ResultSetMetaData, Using a Prepared Statement, Parameterized Statements, Stored Procedures, Transaction Management. Collection:

	Collection Framework, ArrayList class, LinkedList class, ListIterator interface, HashSet class
V	Introducing AWT: Working with Windows Graphics and Text. Using AWT Controls, Layout Managers, adapter classes and Menus. Swing: Basics of Swing, JButton class, JRadioButton class, JTextArea class, JComboBox class, JTable class, JColorChooser class, JProgressBar class, JSlider class, Displaying Image, JMenu for Notepad, Open Dialog Box

REFERENCE BOOKS:

JAVA The Complete Reference- Patrick Naughton and Herbert Schidt.- fifth Edition
Tata McGraw Hill.

The Complete reference J2SE - Jim Keogh – Tata McGraw Hills

Programming and Problem Solving With Java, Slack, Thomson Learning, 1Edn.

Java Programming Advanced Topics, Wigglesworth, Thomson Learning, 3Edn.

Java Programming, John P. Flynt , Thomson Learning, 2Edn.

Ken Arnold and James Gosling, The Java Programming language, Addison Wesley,
2nd Edition, 1998

Patrick Naughton and Herbert Schidt. The Complete Reference, JAVA fifth Edition
Tata McGraw Hill.

Maydene Fisher, Jon Ellis, Jonathan Bruce; JDBC API Tutorial and Reference, Third
Edition, Publisher: Addison-Wesley

Thinking java – Bruce Eckel – Pearson Education Association

CSDA202 Data Communications and Computer Networks**Syllabus**

I	Introduction: Data Communications, Computer Networks, Network Layering-OSI reference Model, TCP-IP Protocol Suite. Physical Layer:Data and Signals, Periodic Analog Signals, Digital Signals, Transmission Impairment, Data rate Limits. Digital-to-Digital Conversion, Analog-to-Digital Conversion, Digital-to-Analog Conversion, Analog-to-Digital Conversion
II	Physical Layer: Transmission and Switching Transmission Modes, Transmission media- Guided, unguided media. Multiplexing, Switching-Circuit Switching, packet switching
III	Data Link Layer: Nodes and Links, Link-Layer Addressing, error Detection and Correction- Block coding, Cyclic Codes, Checksum, Forward Error Correction, Simple, Stop-and-wait, Go-back-N, Selective Repeat Media Access Control: Random Access-ALOHA, CSMA, CSMA/CD, CSMA/CD, Controlled Access, Channelization-FDMA, TDMA, CDMA
IV	Network Layer: Services, Routing Algorithms: Distance Vector, Link State, Path Vector, and Unicast Routing Algorithms.
V	Multicasting Basics: Addresses, Delivery at Data Link Layer, Multicast Forwarding, Two Approaches to Multicasting. IP Addressing, Classes, Subnetting.

REFERENCE BOOKS:

Forouzan, "Data Communications and Networking", 5th Edition, McGraw Hill, 2013.

Andrews. Tanenbaum, "Computer Networks", 5th edition . Prentice-Hall.

William Stallings, "Data and Computer Communication", 8th edition

CSDA203 Advanced Software Engineering**Syllabus**

I	Introduction to Software Engineering , Process Models, Understanding Requirements Agile methodology - Agile – Primer, Manifesto, Characteristics, Daily Stand-up, Definition of Done, Release Planning, Iteration Planning, Product Backlog
II	Requirements Modelling - Analysis, UML Models, Data Modelling, Class-Based Modelling, Webapps Design Concepts - Design Model, Software Architecture- Styles- Design, Component Level Design- Class based Components, User-Interface Design- Interface Analysis, Interface design, WebApp Design
III	Software Quality Assurance , Software Testing Strategies, Testing Applications- Conventional-Object-oriented- Web,
IV	Project Management Concepts - Process Metrics, Estimation, Scheduling, Risk Management, Maintenance and re-engineering
V	DevOps - JUnit - git - github - Docker - Containers - Continuous Integration - Selenium - HTTP loadtestingtool-Designpatterns.

REFERENCE BOOKS:

Software Engineering, a Practitioner's Approach- Roger S Pressman 7th Edition, Tata Mc-Graw Hill Publishing Co. Ltd.

Software Engineering – Ian Somerville 9th Edition, Pearson Education

An Integrated Approach to Software Engineering- Pankaj Jalote 3rd edition, Narosa Publishing House

Fundamentals of Software Engineering- Ghezzi, Jazayer's and Mandriolli 2nd Edition, PHI

Software Engineering principles & Practice- Waman S Jawadekar 2nd Edition, Tata Mc-Graw Hill Publishing Co. Ltd.

Software Project Management: Pankaj Jalote, Pearson Education

Software Project Management –A Unified Framework: Walker Royce, Pearson Education.

Software Project Management –S A Kelkar .Prentice Hall India

SeleniumSimplified, secondedition.

CSDA204 Artificial Intelligence

Syllabus

I	Introduction - Overview of AI applications. Introduction to representation and search. The Propositional calculus, Predicate Calculus, Using Inference Rules to produce. Predicate Calculus expressions, Application – A Logic based financial advisor.
II	Introduction to structure and Strategies for State Space search, Graph theory, Strategies for state space search, Heuristic Search: introduction, Hill-Climbing and Dynamic Programming, The Best-first Search Algorithm, Admissibility, Monotonicity and informedness, Using Heuristics in Games-Minimax algorithm, Alpha beta pruning.
III	Building Control Algorithm for Statespace search – Introduction, Production Systems, Knowledge Representation – Issues, History of AI representational schemes, Conceptual Graphs, Alternatives to explicit Representation, Agent based and distributed problem solving.
IV	Strong Method Problem Solving – Introduction, Overview of Expert System Technology, Rule Based Expert system, Introduction to Planning. Reasoning in Uncertain Situation – introduction, Bayes Rule Inference, Belief Network, Utility Based System, Decision Network
V	Machine Learning: Symbol Based – Introduction, Frame -work. The ID3 Decision tree Induction algorithm, Unsupervised learning, Reinforcement Learning, Machine Learning : Connectionist – Introduction, foundations, Perceptron learning. Machine learning: Social and emergent: Models, The Genetic Algorithm.

TEXT/REFERENCE BOOKS:

George F Luger, Artificial Intelligence – Structures and Strategies for Complex problem solving, 5thEdn, pearson.

E. Rich, K. Knight, S B Nair, Artificial intelligence, 3rdEdn, McGraw Hill.

S. Russel and p. Norvig, Artificial intelligence – A Modern Approach, 3rdEdn, Pearson

D W Patterson, introduction to Artificial Intelligence and Expert Systems, PHI, 1990

Nilsson N.J., Artificial Intelligence - A New Synthesis, Harcourt Asia Pvt. Ltd.

CSDA205 Data Mining

Syllabus	
I	<p>Introduction Data Warehousing, Multidimensional Data Model, OLAP Operations, Introduction to KDD process, Data mining, Data mining -On What kinds of Data, Data mining Functionalities, Classification of Data Mining Systems.</p> <p>Data Preprocessing Data Cleaning, Data Integration and Transformation, Data Reduction, Data discretization and concept hierarchy generation</p>
II	<p>Exploring Data and Visualization Techniques General Concepts, Techniques, Visualizing Higher Dimensional Data, Tools</p> <p>Association Analysis Basic Concepts, Efficient and Scalable Frequent Item set Mining Methods:Apriori Algorithm, generating association Rules from Frequent Item sets, Improving the Efficiency of Apriori. Mining Frequent item-sets without Candidate Generation, Evaluation of Association Patterns, Visualization. A Case Study on Association using Orange Tool</p>
III	<p>Classification Introduction to Classification and Prediction, Classification by Decision Tree Induction: Decision Tree induction, Attribute Selection Measures, Tree Pruning, Bayesian Classification: Bayes' theorem, Naïve Bayesian Classification, Rule Based Algorithms: Using If - Then rules of Classification, Rule Extraction from a Decision Tree, Rule Induction Using a Sequential Covering algorithm, K- Nearest Neighbour Classifiers, Support Vector Machine. Evaluating the performance of a Classifier, Methods for comparing classifiers, Visualization. A Case Study on Classification using Orange Tool</p>
IV	<p>Prediction Linear Regression, Nonlinear Regression, Other Regression-Based Methods</p> <p>Cluster Analysis I: Basic Concepts and Algorithms Cluster Analysis, Requirements of Cluster Analysis' Types of Data in Cluster Analysis, Categorization of Major Clustering Methods, Partitioning Methods: k-Means and k- Medoids, From K-Medoids to CLARANS A Case Study on Clustering using Orange Tool.</p>
V	<p>Cluster Analysis II: Hierarchical Method: Agglomerative and Divisive Hierarchical Clustering. Density-based Clustering - DBSCAN, Grid based clustering-STING Evaluation of Clustering Method</p>
TEXT/REFERENCE BOOKS:	
R	Pang-Ning Tan, Michael Steinbach, Vipin Kumar, 'Introduction to Data Mining'
R	Data Mining Concepts and Techniques - Jiawei Han and MichelineKamber, Second Edition, Elsevier, 2006

R	G. K. Gupta, "Introduction to Data Mining with Case Studies", Eastern Economy Edition, Prentice Hall of India, 2006.
R	Making sense of Data: A practical guide to exploratory Data Analysis and Data Mining-Glenn J Myatt
COURSE PRE-REQUISITES:	
MCA101, MCA 104	

CSDA206 Java Programming Lab

Syllabus	
I	<ul style="list-style-type: none"> • Program to illustrate class, objects and constructors • Program to implement overloading, overriding, polymorphism etc.
II	<ul style="list-style-type: none"> • Program to implement the usage of packages • Program to create user defined and predefined exception • Program for handling file operation • Directory manipulation in java
III	<ul style="list-style-type: none"> • Implement the concept of multithreading and synchronization • Program to implement Generic class and generic methods • Socket programming to implement communications • Broadcasting program using UDP protocol • Program for downloading web pages from the internet using URL.
IV	<ul style="list-style-type: none"> • Program to implement JDBC in GUI and Console Application
V	<ul style="list-style-type: none"> • Program for event-driven paradigm in Java • Event driven program for Graphical Drawing Application • Program that uses Menu driven Application
TEXT/REFERENCE BOOKS:	
JAVA The Complete Reference- Patrick Naughton and Herbert Schidt.- fifth Edition Tata McGraw Hill.	
The Complete reference J2SE - Jim Keogh – Tata McGraw Hills	
Programming and Problem Solving With Java, Slack, Thomson Learning, 1Edn.	
Java Programming Advanced Topics, Wigglesworth, Thomson Learning, 3Edn.	
Java Programming, John P. Flynt , Thomson Learning, 2Edn.	
Ken Arnold and James Gosling, The Java Programming language, Addison Wesley, 2nd Edition, 1998	
Patrick Naughton and Herbert Schidt. The Complete Reference, JAVA fifth Edition Tata McGraw Hill.	

Maydene Fisher, Jon Ellis, Jonathan Bruce; JDBC API Tutorial and Reference, Third Edition, Publisher: Addison-Wesley

Thinking java – Bruce Eckel – Pearson Education Association

CSDA207 Data Mining lab

1. Demonstration of Pre-processing techniques
2. Demonstration of Association Rule Mining – Analysis and Evaluation of Model Performance
 - Apriori Algorithm
 - FP-Growth Algorithm
3. Demonstration of Classification and Prediction Techniques- Analysis and Evaluation of Model Performance
 - Decision Tree
 - Random Forest
 - Naïve Bayesian Classifier
 - K-Nearest Neighbour Classification
 - Support Vector Machines
 - Linear Regression
4. Demonstration of Clustering Techniques- Analysis and Evaluation of Model Performance
 - K-Means Algorithm
 - K-Medoids Algorithm
 - Hierarchical Clustering Algorithms
 - Evaluating the Clusters
5. Project

Semester III

CSDA301 Data Visualization

Syllabus	
I	<p>Computational Statistics and Data Visualization, Data Visualization and Theory, Presentation and Exploratory Graphics, Graphics and Computing, Statistical Historiography</p> <p>Good Graphics –Introduction, Content, Context and Construction, Presentation Graphics and Exploratory Graphics, Presentation (What to Whom, How and Why), Choice of Graphical Form, Graphical Display Options, Higher-dimensional Displays and Special Structures, Scatterplot Matrices (Sploms), Parallel Coordinates, Mosaic Plots, Small Multiples and Trellis Displays, Time Series and Maps</p>
II	<p>Complete Plots, Sensible Defaults, Customization-Setting Parameters, Arranging Plots, Annotation, Extensibility-Building Blocks, Combining Graphical Elements, 3-D Plots, Speed, Output Formats, Data Handling</p> <p>Data and Graphs, Graph Layout Techniques- Force-directed Techniques, Multidimensional Scaling, The Pulling Under Constraints Model, Bipartite Graphs Graph Drawing, Hierarchical Trees, Spanning Trees, Networks, Directed Graphs, Treemaps.</p>
III	<p>High-dimensional Data Visualization</p> <p>Introduction, Mosaic Plots, Associations in High-dimensional Data, Response Models, Models, Trellis Displays, Definition, Trellis Display vs. Mosaic Plots, Visualization of Models, Parallel Coordinate Plots, Geometrical Aspects vs. Data Analysis Aspects, Limits</p> <p>Multidimensional Scaling</p> <p>Proximity Data, Metric MDS, Non-metric MDS, Example: Shakespeare Keywords, Procrustes Analysis, Unidimensional Scaling, INDSCAL, Correspondence Analysis and Reciprocal Averaging, Large Data Sets and Other Numerical Approaches</p>
IV	<p>Tableau.</p> <p>Introduction- Environmental setup, Design Flow, File Types, Data Types. Data Sources- Custom Data View, Extracting Data, Field operations, Metadata, Data Joining and Blending, Worksheets- Adding, renaming, reordering Worksheet, Pages Workbook</p> <p>Calculations- Operators, functions, Calculations, LOD Expressions.</p>
V	<p>Sort and Filters- Sorting, Quick filtering, Context filtering, Condition filtering, Filter operations, Charts, Advanced tableau, Tableau – Bar Chart, Line Chart, Multiple Measure Line Chart, Pie Chart, Crosstab, Scatter Plot, Bubble Chart, Bullet Graph, Box Plot. Dashboard, Forecasting</p>
REFERENCE BOOKS:	

Handbook of Data Visualization by Chun-houh Chen, Wolfgang Härdle, Antony Unwin
The Functional Art by Alberto Cairo
The Visual Display of Quantitative Information by Edward R. Tufte
Learning tableau by Joshua N. Milligan
Tableau Dashboard Cookbook by Jen Stirrup
Handbook of Data Visualization by Chun-houh Chen, Wolfgang Härdle, Antony Unwin

CSDA302 Big Data Analytics

Syllabus	
I	<p>Module 1: INTRODUCTION TO BIG DATA</p> <p>Introduction to BigData Platform – Traits of Big data -Challenges of Conventional Systems - Web Data – Evolution Of Analytic Scalability - Analytic Processes and Tools - Analysis vs Reporting - Modern Data Analytic Tools - Statistical Concepts: Sampling Distributions – ReSampling - Statistical Inference - Prediction Error.</p>
II	<p>Module 2: INTRODUCTION TO BIG DATA AND HADOOP</p> <p>Types of Digital Data, Introduction to Big Data, Big Data Analytics, History of Hadoop, Apache Hadoop, Analyzing Data with Unix tools, Analyzing Data with Hadoop, Hadoop Streaming, Hadoop Echo System, IBM Big Data Strategy, Introduction to Infosphere BigInsights and Big Sheets.</p>
III	<p>Module 3: HDFS(Hadoop Distributed File System)</p> <p>The Design of HDFS, HDFS Concepts, Command Line Interface, Hadoop file system interfaces, Data flow, Data Ingest with Flume and Scoop and Hadoop archives, Hadoop I/O: Compression, Serialization, Avro and File-Based Data structures.</p>
IV	<p>Module 4: Map Reduce</p> <p>Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features.</p>
V	<p>Module 5: Hadoop Eco System</p> <p>Pig : Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators.</p> <p>Hive : Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, Querying Data and User Defined Functions.</p> <p>Hbase : HBasics, Concepts, Clients, Example, Hbase Versus RDBMS.</p>

Big SQL : Introduction
REFERENCE BOOKS:
Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007.
Anand Rajaraman and Jeffrey David Ullman, "Mining of Massive Datasets", Cambridge University Press, 2012.
Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics", John Wiley & sons, 2012.
Glenn J. Myatt, "Making Sense of Data", John Wiley & Sons, 2007
Pete Warden, "Big Data Glossary", O'Reilly, 2011.

CSDA303 (1) Data Warehousing

Module 1:
Introduction to Data Warehouse: Basic elements of the Data Warehouse: Source system-Data staging Area-Presentation Server-Dimensional Model-Business process-Data Mart-Data warehouse. Data Warehouse Design: The case for dimensional modeling – Putting Dimensional modeling together: the data warehouse bus architecture – Basic dimensional modeling techniques.
Module 2:
Data Warehouse Architecture: The value of architecture – An architectural framework and approach – Technical architecture overview – Back room data stores – Back room services. Back Room Services. Data Staging: Data staging overview – Plan effectively – Dimension Table staging – Fact Table loads and warehouse operations – Data quality and cleansing – issues.
Module 3:
Metadata: Metadata, metadata interchange initiative, metadata repository, metadata management, implementation examples, metadata trends, reporting and query tools and applications- tool categories, the need for applications. OLAP: Operational Data Store-OLAP: ROLAP, MOLAP and HOLAP. Need for OLAP, multidimensional data model, OLAP guidelines, multidimensional versus multi relational OLAP, categorization of OLAP tools.
Module 4:
Building a data warehouse: Business considerations, Design considerations, technical considerations, implementation considerations, integrated solutions, benefits of data warehousing, Relational data base technology for data warehouse, database architectures for parallel processing, parallel RDBMS features, alternative technologies
Module 5:
DBMS schemas for decision support :Data layout for best access, multidimensional data model, star schema, STARjoin and STARindex, bitmapped indexing, column local storage, complex data types, Data extraction, clean up and transformation tools-tool requirements, vendor approaches, access to legacy data, vendor solutions, transformation engines

References:

- [1] Kimball Ralph,Reeves,Ross,Thronthwaite ,”The Data warehouse lifecycle toolkit”, Wiley India, 2nd Edition, 2006.

- [2] Berson Alex, Stephen J Smith, "Data Warehousing, Data Mining and OLAP", TATA McGraw-Hill, 13th reprint 2008.
- [3] Soumendra Mohanty, "Data Warehousing design, development and Best practices", TATA McGraw-Hill, 4th reprint 2007.

CSDA303 (2) Digital Image Processing

Syllabus	
I	Module 1 Fundamentals of Image Processing: Introduction – Elements of visual perception, Steps in Image Processing Systems, image Acquisition – Sampling and Quantization – Pixel Relationships – Colour Fundamentals and Models, File Formats. Introduction to the Mathematical tools.
II	Module 2 Image Enhancement and Restoration : Spatial Domain Gray level Transformations Histogram Processing Spatial Filtering – Smoothing and Sharpening. Frequency Domain: Filtering in Frequency Domain – DFT, FFT, DCT, Smoothing and Sharpening filters – Homomorphic Filtering., Noise models, Constrained and Unconstrained restoration models.
III	Module 3 Image Segmentation and Feature Analysis: Detection of Discontinuities – Edge Operators – Edge Linking and Boundary Detection – Thresholding – Region Based Segmentation – Motion Segmentation, Feature Analysis and Extraction.
IV	Module 4: Multi Resolution Analysis and Compressions: Multi Resolution Analysis: Image Pyramids – Multi resolution expansion – Wavelet Transforms, Fast Wavelet transforms, Wavelet Packets. Image Compression: Fundamentals – Models – Elements of Information Theory – Error Free Compression – Lossy Compression – Compression Standards – JPEG/MPEG.
V	Module 5: Applications of Image Processing: Representation and Description, Image Recognition-Image Understanding – Image Classification – Video Motion Analysis – Image Fusion – Steganography – Colour Image Processing.
References:	
	Rafael C.Gonzalez and Richard E.Woods, "Digital Image Processing", Third Edition,

	Pearson Education, 2008.
	Milan Sonka, Vaclav Hlavac and Roger Boyle, "Image Processing, Analysis and
	Machine Vision", Third Edition, Third Edition, Brooks Cole, 2008.
	Anil K.Jain, "Fundamentals of Digital Image Processing", Prentice-Hall India, 2007.
	Madhuri A. Joshi, 'Digital Image Processing: An Algorithmic Approach", Prentice-
	Hall India, 2006.
	Rafael C.Gonzalez , Richard E.Woods and Steven L. Eddins, "Digital Image

CSDA304 Business Modelling & Applied Analytics Using R

Syllabus	
I	<p>Module 1: Introduction to R</p> <p>Introduction to R and Familiarization of R Studio, Basic components in R Studio.</p> <p>R Syntax and programming - Variables & Operators, Vectors, List, Matrices & Arrays, Factors, Data Frames & Functions Reading data using R - Basic read write operations.</p> <p>Exploratory functions to cover Summary & Structure of data, Measures of central tendency and measures of dispersion.</p>
II	<p>Module 2: Data Handling and Visualization</p> <p>Functions used for cleaning data - handling messy data and missing data -</p> <p>Basic charts and their purpose - pie, bar and histogram.</p> <p>Boxplot, Scatterplot. Understanding ggplot2 package, Functions in ggplot2</p> <p>Quickplot</p>
III	<p>Module 3: Supervised Learning & Unsupervised Learning</p> <p>Supervised modelling technique. Family of Regressions SLR, BLR, MLR Modelling, Decision Tree- Random Forest. Unsupervised modelling techniques</p> <p>Clustering Concept - K Means Clustering, Association Rules- ARM Concept - Apriori.</p>
IV	<p>Module 4: Applied Analytics - HR & Operation</p> <p>HR Analytics: Understanding role of analytics in HR Function, Understanding KPI's that needs to be modelled. Modelling Attrition - Understanding how modelling attrition helps an organization. Model Building, Model Diagnostics and evaluation. CTC prediction model- Modelling CTC prediction and evaluating social networks</p> <p>Operations Analytics: Understanding role of analytics in</p>

	Operations Analytics - Introduction- Distribution channel development - using predictive analytics in setting up distribution centers.
V	Module 5: Applied Analytics - Finance & Marketing Finance Analytics: Understanding role of analytics in finance. Customer profiling using clustering techniques Applied Credit risk modelling using classification and regression techniques Marketing Analytics: Understanding analytics in marketing. Usage of predictive modelling in Sales forecasting, Customer segmentation, Customer feedback analysis. Retail analytics, Market Basket Analysis
Reference	
	Hands-On Programming with R by Golemund and Garrett
	Beginning R: The Statistical Programming Language by Mark Gardener
	R for Everyone: Advanced Analytics and Graphics by Jared P. Lander
	Applied Predictive Analytics: Principles and Techniques for The Professional Data Analyst by Dean Abbott
	Predictive Marketing: Easy Ways Every Marketer Can Use Customer Analytics and Big Data by Omer Artun and Dominique Levin
	HR Analytics: Understanding Theories and Applications by Dipak Kumar Bhattacharyya.

CSDA305 Web Technology Lab

Syllabus	
I	HTML Basics, CSS, Javascript, Ajax, Json, XML, PHP Basics
II	Laravel- Installation Application Structure Configuration Routing Middleware Namespaces Controllers
III	Request Cookie Response Views Blade Templates Redirections Working with Database Errors & Logging Forms Localization Sessions Cookies

	Validation File Uploading Sending Email Ajax Error Handling Event Handling
IV	Facades Contracts CSRF Protection Authentication Authorization Artisan Console Hashing Artisan Commands
V	Development of web-based application with Database connectivity

CSDA306 Data Analytics using Python

Syllabus	
I	Data Types and Data Structures Introduction to Python: - using the Python interpreter, Overview of programming in Python, Expressions and Variables-String Operations. Python Data Structures: lists & Tuple -Sets - Dictionaries. Programming Fundamentals: Conditions and Branching- Loops-Functions: formal arguments, variable-length arguments.
II	Classes, files and modules Introduction to Classes and Objects: -classes, class attributes, instances, instance attributes, binding and method invocation, inheritance, polymorphism, Built-in functions for classes and instances. Files and input/output, reading and writing files, methods of file objects, using standard library functions, dates and times Exceptions, detecting and handling exceptions.
III	Database and web programming Python database application programmer's interface (DB- API), connection and cursor objects, Type objects and constructors, python database adapters. Creating simple web clients, introduction to CGI, CGI module, building CGI applications.
IV	Introduction to Data Science using Python Python libraries: Numpy- Scikit- Pandas-Matplotlib. - Data Visualization. Importing Datasets: Importing and Exporting Data in Python- Basic Insights from Datasets. Data cleansing and pre-processing: Identify and Handle Missing Values. Summarizing the Data

	Frame: Descriptive Statistics- Basic of Grouping- ANOVA- Correlation
V	Model Development and Evaluation Regression Models: Linear Regression (SLR & MLR) - Logistic Regression-Decision Tree- Random Forest. Clustering Techniques: K means clustering - Apriori algorithm. Model Evaluation: Over-fitting, Under-fitting.
REFERENCE BOOKS:	
Core Python Programming by Wesley J. Chun, 2nd Edition , Pearson Education	
An Introduction to Python by Guido Van Russom, Fred L.Drake, Network Theory Limited.	
Beginning Python: From Novice To Professional By Magnus Lie Hetland, Second Edition Apress	
Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython ,2nd edition, Wes McKinney, O'Reilly Media (2017)	
Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems ,Aurélien Géron, O'Reilly Media (2017)	
Data Science from Scratch: First Principles with Python, Joel Grus, O'Reilly Media (2015)	

Semester IV

CSDA401 Deep Learning

Syllabus	
I	Introduction to Deep Learning Why Deep Learning? What is a neural network? -Basics: Biological Neuron, Idea of computational units, McCulloch-Pitts unit and Thresholding logic, Linear Perceptron, Perceptron Learning Algorithm, Linear separability. Convergence theorem for Perceptron Learning Algorithm. Backpropagation, Multi-layer Perceptrons
II	Introduction to Tensorflow, simple ML examples. Basic operations, constants, variables, Control dependencies, Data pipeline, TensorBoard, Linear and Logistic Regression, Tensorflow's Optimizers, tf.data-Birth rate - life expectancy, MNIST dataset
III	Loss Functions and Optimization, Image features, Optimization, stochastic gradient descent, Convolutional Neural Networks, Convnet in TensorFlow- image classification Solving a problem with CNNs on Tensorflow.
IV	Recurrent Neural Networks, Language modelling Image captioning, Soft attention Back propagation through time, Long Short-Term Memory, LSTMs, Bidirectional RNNs, Solving a problem with RNNs on Tensorflow
V	Practical: <ul style="list-style-type: none"> • Introduction to TensorFlow • TensorFlow Basic Syntax • TensorFlow Graphs • TensorFlow - A Neural Network • TensorFlow Regression Example • TensorFlow Classification Example • CNN Implementation • CNN MNIST Code • Introduction to RNN Section • Manual Creation of RNN
REFERENCE BOOKS:	
Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016. http://www.deeplearningbook.org .	
K. P. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012.	
C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.	
Neural Networks and Deep Learning by Michael Nielsen, Online	
Hands-On Machine Learning with Scikit-Learn and TensorFlow, by Aurélien Géron	

CSDA402 (1) Information Retrieval Techniques

Syllabus	
I	Module 1: INTRODUCTION Basic Concepts - Retrieval Process - Modeling - Classic Information Retrieval - Set Theoretic, Algebraic and Probabilistic Models - Structured Text Retrieval Models - Retrieval Evaluation - Word Sense Disambiguation
II	Module 2: QUERYING Languages - Key Word based Querying - Pattern Matching - Structural Queries - Query Operations - User Relevance Feedback - Local and Global Analysis - Text and Multimedia languages
III	Module 3: TEXT OPERATIONS AND USER INTERFACE Document Preprocessing - Clustering - Text Compression - Indexing and Searching - inverted files - Boolean Queries - Sequential searching - Pattern matching - User Interface and Visualization - Human Computer Interaction - Access Process - Starting Points - Query Specification - Context - User relevance Judgment - Interface for Search
IV	Module 4: MULTIMEDIA INFORMATION RETRIEVAL Data Models - Query Languages - Spatial Access Models - Generic Approach - One Dimensional Time Series - Two Dimensional Color Images - Feature Extraction
V	Module 5: APPLICATIONS Searching the Web - Challenges - Characterizing the Web - Search Engines - Browsing - Meta-searchers - Online IR systems - Online Public Access Catalogs - Digital Libraries - Architectural Issues - Document Models, Representations and Access - Prototypes and Standards. Case study - Google search engine
References	
	Ricardo Baeza-Yate, Berthier Ribeiro-Neto, "Modern Information Retrieval: The Concepts and Technology behind Search", Pearson Education, 2011.
	G.G. Chowdhury, "Introduction to Modern Information Retrieval", Neal-Schuman Publishers; 2nd edition, 2003.
	Daniel Jurafsky and James H. Martin, "Speech and Language Processing", Pearson Education, 2000
	David A. Grossman, Ophir Frieder, "Information Retrieval: Algorithms, and Heuristics", Academic Press, 2000
	C. Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval", Cambridge University Press, 2008.

AnandRajaraman and Jeffery D.ullman, " <i>Mining the Massive</i> ", Cambridge University Press, 2008.

CSDA402 (2) Social Media Mining

Syllbus	
I	Module 1: Introduction-New Challenges for Mining, Graph basics- Graph Representation , Types of Graphs, Connectivity in Graphs, Special Graphs, graph algorithms, Network measures-centrality, transitivity and reciprocity, balance and status, similarity, Network Models -Properties of Real-World Networks, Random Graphs, Small-World Model , Preferential Attachment Model
II	Module 2: Data Mining Essentials- Data, Data Preprocessing, Data Mining Algorithms, Supervised Learning , Unsupervised Learning
III	Module 3: Communities and Interactions- Community Analysis, Community Evolution, Community Evaluation Information Diffusion in Social Media- Herd Behavior, Information Cascades , Diffusion of Epidemics
IV	Module 4: Influence and Homophily- Measuring Assortativity , Influence, Homophily , Distinguishing Influence and Homophily Recommendation in Social Media- Challenges , Classical Recommendation Algorithms, Recommendation Using Social , Evaluating Recommendations
V	Module 5: Behavior Analytics- Individual Behavior, Individual Behavior Analysis, Individual Behavior Modelling, Individual Behavior Prediction, Collective Behavior
References	
	<i>Social Media Mining- An Introduction</i> , Reza Zafarani, Mohammad Ali Abbasi. Huan. Cambridge University Press, 2014
	<i>Mining of Massive Datasets</i> , Jure Leskovec, AnandRajaraman, Jeffrey D. Ullman,

CSDA403 Comprehensive Viva Voce

Course Code	MCA403	Course Name	Comprehensive Viva Voce
Course Type	Viva Voce	Contact Hours	Nil
Credit	2	Domain	Professional Core
Will be conducted at the end of Semester. A comprehensive Viva based on subjects learned during the course, by Internal Examiner for internal Evaluation and by an external Examiner			

MCA404 Main Project

Course Description
<p>The Internship/main project Course allows M.Sc Students to gain practical experience in the workplace before receiving their Graduation Degrees. The student identifies companies willing to hire him/her on a full-time basis for an 8-week period (minimum required) during their last semester.</p> <p>Responsibilities of an Intern</p> <ul style="list-style-type: none"> • Work closely with teams at the workplace to facilitate the rapid development of high-quality applications related to data Analytics which may include: <ul style="list-style-type: none"> ○ Understand and Plan the business problem / research question to analytics question ○ Work cross functionally to gather and pre-process the data ○ Analyze and solve the problem ○ Visualising the results ○ Properly presenting the inferences of the study ○ Documentation leading to a research article
Internship/ Main project Guidelines:
<ul style="list-style-type: none"> • Step 1: Request Letter/ Email from Internship Coordinator of the college should go to industry/ research institute to allot various slots of 8 weeks as internship periods for the students. Students request letter/profile/ interest areas may be submitted to industries for their willingness for providing the training. • Step 2: Industry/ research institute will confirm the training slots and the number of seats allocated for internships via Confirmation Letter/ Email to the Internship Coordinator • Step 3: Students on joining Training at the concerned Industry / Organization, submit the Joining Report/Letters / Email. • Step 4: Students undergo training at the concerned Industry / Organization. In-between Faculty Member(s) evaluate(s) the performance of students once/twice by visiting the Industry/Organization/ through Online Interactions and Evaluation Report of the students is submitted in department office with the consent of Industry persons

- Step 5: Students will submit training report after completion of internship along with the Attendance Log to the Internship Coordinator.
- Step 6: Training Certificate to be obtained from industry and a copy to be submitted to the Office of the Coordinator.
- Step 7: Assessment of the Internship Outcomes through a Comprehensive Viva and extensive evaluation of the Internship Report.

INTERNSHIP/ MAIN PROJECT REPORT GUIDELINES

Every student is required to write an Internship/ project report upon completion of their internship/ project and required to submit **two copies** (student copy + department copy in pdf) of the report to Internship Coordinator (along with certificate given by the company) for final evaluation and awarding of Credit Scores. Before submitting the report to the Internship Coordinator, the student required to go through multiple rounds of revision in collaboration with the department internship mentor/coordinator/supervisor.

The Internship/ project Report serves multiple purposes:

- Help the student develop written communication skills.
- Serve as an archival record of the internship experience.
- Give the student an opportunity to reflect on the professional aspects of the internship experience and the skills that were learned.
- Allow the student to describe the science content of the internship.
- Have the student to reflect on the initial goals of the internship and how they were (or were not) achieved during the internship.

Text Format in the report:

- Cambria 12 or similar, with 1.5 line spacing.
- Margins 1.5" left and 1" all other side.

Binding & report length:

- Soft binding & report length of minimum 20 pages with one side printing with a designed Cover Page

General information:

- Student is eligible for internship evaluation if only if he/she completed 8 weeks of internship training. (Minimum of 40 Working days)

EACH INTERNSHIP/ PROJECT REPORT WILL FOLLOW THE FORMAT (not limited to):

Title Page

College certificate Page

Internship/ project certificate provided by the internship institution

Acknowledgement

Executive summary/Abstract (200 Words)

Chapter 1: Introduction

A paragraph each on:

- The company
- Business/ Research Problem
- Objectives of the Study
- Plan of Action
- Chapter Summary

Chapter 2: Literature Review (Minimum of 2000 Words)

Chapter 3: Research/ Business Problem (500 Words)
Chapter 4: Working with Data (1000 Words)
Chapter 5: Methodology (2000 Words)
Chapter 6: Results and Inferences (1000 Words)
Chapter 7: Conclusion of the Study/ internship (500 Words)

Bibliography

Include references to books, articles, reports referred to in the report.

Note: A handbook with the formats of Certificate and Details will be given at the start of Semester IV.

*****End of the Document*****



RCSS
RAJAGIRI COLLEGE OF
SOCIAL SCIENCES
(AUTONOMOUS)