M Sc. Computer Science (Data Analytics)

REGULATIONS, SCHEME AND SYLLABUS

From 2020 Admissions Onward

BOARD OF STUDIES (COMPUTER SCIENCE)

RAJAGIRI COLLEGE OF SOCIAL SCIENCES (AUTONOMOUS)

KALAMASSERY, KOCHI, 683104

KERALA, INDIA

Table of Contents

M.Sc. Computer Science (Data Analytics)

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. Analysing 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 emphases 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.

Eligibility

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.

Evaluation and Grading

There shall be a Semester Examinations at the end of each semester for all credit courses of duration of 3 hours. A question paper 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 (ESE) [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 (ESE) 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 (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 Internal (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	**A+**	Outstanding
4.00 to 4.49	**A**	Excellent
3.50 to 3.99	**B+**	Very good
3.00 to 3.49	**B**	Good (average)
2.50 to 2.99	**C+**	Fair
2.00 to 2.49	**C**	Marginal (pass)
Upto 1.99	**D**	Deficient (fail)

No separate minimum is required for internal evaluation for a pass, but a minimum C grade is required for a pass in an external evaluation. However, a minimum C 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:

Semester Grade Point Average – SGPA (S_j) = $\sum (C_i \times G_i) / \sum C_i$

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

Where 'S_j' is the **j^{th}** semester, 'G_i' is the grade point scored by the student in the **i^{th}** course 'c_i' is the credit of the **i^{th}** course.

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

Cumulative Grade Point Average (CGPA) = $\sum (C_i \times S_i) / \sum C_i$

(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 2.0. However, a student is permitted to move to the next semester irrespective of her/his **SGPA.**

Award Of Degree

The successful completion of all the courses with **'C'** grade within the stipulated period shall be the minimum requirement for the award of the degree.

Credits allotted for Programmes and Courses

Total credit for MCA programme shall be **80**

SCHEME

I Semester

Course No:	Subject	No. of hours per week		Credit
		Lecture	**Lab**	
CSDA101	Operating System	4	-	3
CSDA102	Data Structures Using C	4	-	3
CSDA103	Statistics for Data Analytics	4	-	3
CSDA104	Database Management System	4	-	3
CSDA105	Business Intelligence	4	-	4
CSDA106	Data Structures Lab	-	4	2
CSDA107	DBMS Lab	-	4	2
	Total	20	8	20

II Semester

Course No:	Subject	No. of hours per week		Credit
		Lecture	**Lab**	
CSDA201	Object Oriented Programming using Java	4	-	3
CSDA202	Data Communication and Computer networks	4	-	3
CSDA203	Software Engineering	4	-	3
CSDA204	Artificial Intelligence	4	-	3
CSDA205	Data Mining	4	-	4
CSDA206	Java lab	-	4	2

CSDA207	Data Mining lab	-	4	2
	Total	20	8	20

III Semester

Course No:	Subject	No. of hours per week		Credit
		Lecture	**Lab**	
CSDA301	Data Visualization	4	-	4
CSDA302	Big Data Technologies	4	-	4
CSDA303	Elective I	4	-	3
CSDA304	Elective II	4	-	3
CSDA305	Business Modelling & Applied Analytics Using R	4		4
CSDA306	Python Programming		6	2
	Total	20	6	20

Electives

CSDA303 (1) Data Warehousing
CSDA303 (2) Digital Image Processing
CSDA304 (1) Information Retrieval Techniques
CSDA304 (2) Social Media Mining

IV Semester

Course No:	Subject	No. of hours per week		Credit
		Lecture	**Lab**	
CSDA401	Main Project	One Semester		16
CSDA402	Comprehensive Viva			4
	Total			20
	Total Credits of M.Sc. Computer Science (Data Analytics)			**80**

SEMESTER 1

CSDA101 Operating System

Module 1: File Systems

File Systems, File concept, File support, Access methods, Allocation methods, Directory systems, File protection, free space management

Disk Management-Secondary-Storage Structure, Disk structure, Disk scheduling, Disk management, Swap-space management, Disk reliability.

Module 2: Memory Management

Memory Management, Memory partitioning, Swapping, Paging, Segmentation, Virtual memory, Overlays, Demand paging, Performance of Demand paging, Page replacement algorithms, Allocation algorithms

Module 3: Process Management and Concurrency management

Process and Thread Management, Concept of process and threads, Process states, Process management, Context switching, Interaction between processes and OS, Multithreading, Concurrency Control, Concurrency and Race Conditions, Mutual exclusion requirements,

Module 4: Concurrency Management

Software and hardware solutions for mutual exclusion, Semaphores, Classical IPC problems and solutions Deadlock, Characterization, Avoidance and Prevention, Detection, Recovery

Module 5: Protection

Protection, Goals of protection, Domain of protection, Access matrix, Implementation of access matrix, Revocation of access rights.

Case Study

Linux OS –File System, basic commands Processes, Access permissions, redirection, filters.

References:

- 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

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.

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

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
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
Module 5: Trees and Graphs.
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.
Graphs: Graphs – terminology, Representation, Graph traversals (dfs & bfs)

References:

- 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

CSDA103 Statistics for Data Analytics

Module 1:-Basic Statistics

Measures of central tendency: - mean, median, mode; Measures of dispersion: Range, Mean deviation, Quartile deviation and Standard deviation; Moments, Skewness and Kurtosis, Linear correlation, Karl Pearson's coefficient of Correlation, Rank correlation and linear regression.

Module 2:- Probability Theory

Sample space, Events, Different approaches to probability, Addition and multiplication theorems on probability, Independent events, Conditional probability, Bayes Theorem

Module 3:- Random variables and Distribution

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.

Module 4:- Sampling and Estimation

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).

Module 5:-Testing of hypothesis

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 ns2/σ2), F test - test for equality of two population variances.

References

- Fundamentals of statistics: S.C.Gupta, 6thRevised 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 System

Module 1: Introductory concepts of DBMS
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
Module 2: Relational Database design
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
Module 3: SQL Concepts
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-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
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
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 ,Implementaion 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, Second

Edition, Tata McGraw Hill, 2003

CSDA105 Business Intelligence

Module 1:

Business Intelligence an Introduction: Introduction, Definition, 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.

Creating Business Intelligence Environment, Business Intelligence Landscape, Types of Business Intelligence, Business Intelligence Platform, Dynamic roles in Business Intelligence, Roles of Business Intelligence in Modern Business- Challenges of BI

Module 2:

Business Intelligence Types: Introduction, Multiplicity of Business Intelligence Tools, Types of Business Intelligence Tools, Modern Business Intelligence, the Enterprise Business Intelligence, Information Workers

Architecting the Data: Introduction, Types of Data, Enterprise Data Model, Enterprise Subject Area Model, Enterprise Conceptual Model, Enterprise Conceptual Entity Model, Granularity of the Data, Data Reporting and Query Tools, Data Partitioning, Metadata, Total Data Quality Management (TDQM).

Module 3:

Introduction to Data Mining: Definition of Data Mining, Architecture of Data Mining, Kinds of Data which can be mined, Functionalities of Data Mining, Classification on Data Mining system, Various risks in Data Mining, Advantages and disadvantages of Data Mining, Ethical issues in Data Mining, Analysis of Ethical issues

Introduction to Data Warehousing: Introduction, Advantages and Disadvantages of Data Warehousing, Data Warehouse, Data Mart, Aspects of Data Mart, Online Analytical Processing, Characteristics of OLAP, OLAP Tools, OLAP Data Modeling, OLAP Tools and the Internet, Difference between OLAP and OLTP, Multidimensional Data Model

Module 4:

Types of Business Models, B2B Business Intelligence Model, Electronic Data Interchange & E-Commerce Models, Advantages of E-Commerce for B2B Businesses, Systems for Improving B2B E-Commerce, B2C Business Intelligence Model, Need of B2C model in Data warehousing, Different types of B2B intelligence Models

Knowledge Management: Introduction, Characteristics of Knowledge Management, Knowledge assets, Generic Knowledge Management Process, Knowledge Management Technologies, Essentials of Knowledge Management Process

Module 5:

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

Business Intelligence Life Cycle: Introduction, Business Intelligence Lifecycle, Enterprise Performance Life Cycle (EPLC)Framework Elements, Life Cycle Phases, BI Strategy, Objectives and Deliverables, Transformation Roadmap, Building a transformation roadmap, BI Development Stages and Steps, Parallel Development

Tracks, BI Framework

References:

- 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

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

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

CSDA107 DBMS Lab

1. Oracle Installation.
2. Table Design- Using foreign key and Normalization
3. Practice SQL Data Definition Language (DDL) commands
    a. Table creation and alteration (include integrity constraints such as primary key, Referential integrity constraints, check, unique and null constraints both column and table level.
    b. Other database objects such as view, index, cluster, sequence, synonym etc.
4. Practice SQL Data Manipulation Language (DML) commands
    a. Row insertion, deletion and updating
    b. Retrieval of data
        i. Simple select query
        ii. Select with where options (include all relational and logical operators)
    c. Functions: Numeric, Date, Character, Conversion and Group functions with having clause.
    d. Set operators
    e. Sorting data
    f. Sub query (returning single row, multiple rows, more than one column, correlated sub query)
    g. Joining tables(single join, self join, outer join)
5. Practice Transaction Control Language (TCL) commands (Grant, revoke, commit and save point options)
6. Usage of triggers, functions and procedures
7. Cursors

SEMESTER 2

CSDA201 Object Oriented Programming using Java

Module 1
Introduction to Object Oriented Concepts
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 instance of operator ,Package and Access Modifiers, Encapsulation, Object class, Object Cloning, Java Array, Call By Value and Call By Reference
Module 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.
Module 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
Module 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

Module 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
Java applets- Life cycle of an applet – Adding images to an applet – Adding sound to an applet. Passing parameters to an applet. Event Handling.

References

- 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 Professional,2003
- Java Servlets IInd edition Karl Moss Tata McGraw Hils
- Professional JSP – Wrox
- Thinking java – Bruce Eckel – Pearson Education Association
- JavaScript: A Beginner's Guide, Second Edition By John Pollock, McGraw-Hill Professional – Publisher

CSDA202 Data Communication and Computer networks

Module 1

Introduction: Data Communications, Computer Networks, Network Layering- Principles of 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

Module 2

Physical Layer: Transmission and Switching Transmission Modes, Transmission media-Guided, unguided media. Multiplexing, Switching-Circuit Switching, packet switching

Module 3

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, HDLC
Media Access Control: Random Access-ALOHA, CSMA, CSMA/CD, CSMA/CD, Controlled Access, Channelization-FDMA, TDMA, CDMA.

Module 4

Wired LANS: Ethernet Protocol- IEEE 802. Standard Ethernet- Characteristics,

Addressing, Access method. Network Layer: Services, Routing Algorithms: Distance Vector, Link State, Path Vector, and Unicast Routing Algorithms. IP Protocol, IP address, subnetting

Module 5

Multicasting Basics: Addresses, Delivery at Data Link Layer, Multicast Forwarding, Two Approaches to Multicasting.

IP Addressing, Classes, Subnetting.

References

- 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 Software Engineering

Module I: Software process

Software engineering definition, Software problems, important qualities of a software product, software engineering principles. Process Models – The Waterfall Model, Prototyping, incremental model, Spiral Model, V-Model. Agile development

Module II: Requirement Analysis, Design

Understanding Requirements, Requirements Modeling: Scenarios, Software requirements specification, SRS, Role & Skills of system Analyst, Design Concepts, Software Architecture, User Interface Design

Module III: Coding, Testing and Maintenance

Coding – programming principles and guidelines,Coding Standards, refactoring, verification, complexity metrics. Testing – Levels of testing, testing for conventional and object oriented applications, Maintenance – Need for maintenance, Management of maintenance, challenges of maintenance phase.

Module IV: Quality Management

Quality concepts, Software Metrics- LOC based, Function point Metric, Quality Metrics, Review techniques, software quality assurance, Software configuration management, Change Management

Module V: Software Project Management

Project Management Concepts, Estimation for Software Projects, Project Scheduling, Risk Management

References

- Software Engineering, a Practitioner's Approach- Roger S Pressman 7th Edition, Tata Mc-GrawHilll 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-GrawHilll 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
- Information Technology and Project Management, Schwalbe, Thomson Learning 4Edn.

CSDA204 Artificial Intelligence

Module 1:

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.

Module 2:

Introduction to structure and Strategies for State Space search, Graph theory, Strategies for state space search, Using the State Space to Represent Reasoning with the Predicate calculus (Sate space description of a logical system, AND/OR Graph).

Heuristic Search : introduction, Hill-Climbing and Dynamic Programming, The Best-first Search Algorithm, Admissibility, Monotonicity and informedness, Using Heuristics in Games.

Module 3:

Building Control Algorithm for Statespace search – Introduction, Production Systems, The blackboard architecture for Problem solving.

Knowledge Representation – Issues, History of AI representational schemes, Conceptual Graphs, Alternatives to explicit Representation, Agent based and distributed problem solving.

Module 4:

Strong Method Problem Solving – Introduction, Overview of Expert System Technology, Rule Based Expert system, Model -Based, Case-Based and Hybrid Systems (Introduction to Model based reasoning, Introduction to Case Based Reasoning, Hybrid design), Introduction to Planning.

Reasoning in Uncertain Situation – introduction, logic based Adductive Inference.

Introduction to PROLOG , Syntax for predicate Calculus programming, ADTs, A production system example.

Module 5:

Machine Learning: Symbol Based – Introduction, Frame –work. The ID3 Decision tree Induction algorithm. Inductive bias and Learnability, Knowledge and Learning, Unsupervised learning, Reinforcement Learning,

Machine Learning : Connectionist – Introduction, foundations, Perceptron learning.

Machine learning: Social and emergent: Models, The Genetic Algorithm, Artificial Life and Social based Learning.

References

- 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

Module I 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.

Data Pre-processing

Data Cleaning, Data Integration and Transformation, Data Reduction, Data discretization and concept hierarchy generation

Module II Exploring Data and Visualization Techniques

General Concepts, Techniques, Visualizing Higher Dimensional Data, Tools

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

Module III 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

Module IV Prediction

Linear Regression, Nonlinear Regression, Other Regression-Based Methods

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.

Module V

Cluster Analysis II: Hierarchical Method: Agglomerative and Divisive Hierarchical Clustering.

Comparison of data mining methods. Applicability of data mining methods for different scenarios. Considerations for mining unstructured data.

References

- Pang-Ning Tan, Michael Steinbach, Vipin Kumar, 'Introduction to Data Mining'
- Data Mining Concepts and Techniques – Jiawei Han and MichelineKamber, Second Edition, Elsevier, 2006
- G. K. Gupta, "Introduction to Data Mining with Case Studies", Easter Economy Edition, Prentice Hall of India, 2006.
- Making sense of Data: A practical guide to exploratory Data Analysis and Data Mining-Glenn J Myatt

CSDA206 Java lab

- Program to illustrate class, objects and constructors
- Program to implement overloading, overriding, polymorphism etc.
- Program to implement the usage of packages
- Program to create user defined and predefined exception
- Program for handling file operation
- Directory manipulation in java
- 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.
- Program to implement JDBC in GUI and Console Application
- Applet program for passing parameters
- Applet program for loading an image and running an audio file
- Program for event-driven paradigm in Java
- Event driven program for Graphical Drawing Application
- Program that uses Menu driven Application

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
    - 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

5. Project

SEMESTER 3

CSDA301 Data Visualization

Module 1

Computational Statistics and Data Visualization, Data Visualization and Theory, Presentation and Exploratory Graphics, Graphics and Computing, Statistical Historiography

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

Module 2

Complete Plots, Sensible Defaults, Customization-Setting Parameters, Arranging Plots, Annotation, Extensibility-Building Blocks, Combining Graphical Elements, 3-D Plots, Speed, Output Formats, Data Handling

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.

Module 3

High-dimensional Data Visualization
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 Multidimensional Scaling
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
Module 4 -Tableau.

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 Calculations- Operators, functions, Calculations, LOD Expressions.

Module 5 : 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

References

- 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

CSDA302 Big Data Technologies

Module 1: INTRODUCTION TO BIG DATA
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.
Module 2: INTRODUCTION TO BIG DATA AND HADOOP
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.
Module 3: HDFS(Hadoop Distributed File System)
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.
Module 4: Map Reduce
Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features.
Module 5: Hadoop Eco System
Pig : Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators.
Hive : Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, Querying Data and User Defined Functions.
Hbase : HBasics, Concepts, Clients, Example, Hbase Versus RDBMS.
Big SQL : Introduction

References:

- Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007.
- AnandRajaraman 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] SoumendraMohanty," Data Warehousing design,development and Best practices",TATA McGraw-Hill, 4th reprint 2007.

CSDA303 (2) Digital Image Processing

Module 1
Fundamentals of Image Processing: Introduction – Elements of visual perception, Stepsin Image Processing Systems, image Acquisition – Sampling and Quantization – PixelRelationships – Colour Fundamentals and Models, File Formats. Introduction to theMathematical tools.
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.
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.
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 – ErrorFree Compression – Lossy Compression – Compression Standards – JPEG/MPEG.
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
- Processing Using MATLAB”, First Edition, Pearson Education, 2004.

CSDA304 (1) Information Retrieval Techniques

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

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

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

Module 4: MULTIMEDIA INFORMATION RETRIEVAL
Data Models – Query Languages – Spatial Access Models – Generic Approach – One Dimensional Time Series – Two Dimensional Color Images – Feature Extraction

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*,”Mining the Massive”,Cambridge*

University Press, 2008.

CSDA304 (2) Social Media Mining

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

Module 2:
Data Mining Essentials- Data, Data Preprocessing, Data Mining Algorithms, Supervised Learning , Unsupervised Learning

Module 3:
Communities and Interactions- Community Analysis, Community Evolution, Community Evaluation Information Diffusion in Social Media- Herd Behavior, Information Cascades , Diffusion of Epidemics

Module 4:
Influence and Homophily- MeasuringAssortativity , Influence, Homophily , Distinguishing Influence and Homophily
Recommendation in Social Media- Challenges , Classical Recommendation Algorithms, Recommendation Using Social , Evaluating Recommendations

Module 5:
Behavior Analytics- Individual Behavior, Individual Behavior Analysis, Individual Behavior Modelling, Individual Behavior Prediction, Collective Behavior

References

- *Social Media Mining- An Introduction*, Reza Zafarani, Mohammad Ali Abbasi. Huan. Cambridge University Press, 2014
- *Mining of Massive Datasets*, Jure Leskovec,AnandRajaraman, Jeffrey D. Ullman,

CSDA305 Business Modelling & Applied Analytics Using R

Module 1: Introduction to R

Introduction to R and Familiarization of R Studio, Basic components in R Studio. R Syntax and programming - Variables & Operators, Vectors, List, Matrices & Arrays, Factors, Data Frames & Functions Reading data using R - Basic read write operations.
Exploratory functions to cover Summary & Structure of data, Measures of central

tendency and measures of dispersion.

Module 2: Data Handling and Visualization

Functions used for cleaning data - handling messy data and missing data – Basic charts and their purpose - pie, bar and histogram.

Boxplot, Scatterplot. Understanding ggplot2 package, Functions in ggplot2 Quickplot

Module 3: Supervised Learning & Unsupervised Learning

Supervised modelling technique. Family of Regressions SLR, BLR, MLR Modelling, Decision Tree- Random Forest. Unsupervised modelling techniques Clustering Concept – K Means Clustering, Association Rules- ARM Concept – Apriori.

Module 4: Applied Analytics - HR & Operation

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

Operations Analytics: Understanding role of analytics in Operations Analytics – Introduction- Distribution channel development - using predictive analytics in setting up distribution centers.

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 books

1 Hands-On Programming with R by Grolemund and Garrett
2 Beginning R: The Statistical Programming Language by Mark Gardener
3 R for Everyone: Advanced Analytics and Graphics by Jared P. Lander
4 Applied Predictive Analytics: Principles and Techniques for The Professional Data Analyst by Dean Abbott
5 Predictive Marketing: Easy Ways Every Marketer Can Use Customer Analytics and Big Data by Omer Artun and Dominique Levin
6 HR Analytics: Understanding Theories and Applications by Dipak Kumar Bhattacharyya.

CSDA306 Python Programming

Lab Cycle

Introduction

1. Python syntax, functions, packages and libraries-
2. Types-Expressions
3. Variables-String Operations.
4. Python Data Structures: lists & Tuple –Sets -Dictionaries.
5. Programming Fundamentals: Conditions and Branching- Loops-Functions-Objects and Classes

Working with Data and Libraries

1. Importing Datasets: Understanding the Dataset
2. Importing and Exporting Data in Python
3. Introduction to python libraries: Numpy- Scikit- Pandas-Matplotlib.-
4. Data Visualization in Python

Cleaning and Preparing the Data

1. Data cleansing and pre-processing: Identify and Handle Missing Values
2. Data Formatting
3. Data Normalization Sets
4. Binning- Indicator variables. S
5. Summarizing the Data Frame
6. Basic of Grouping- ANOVA- Correlation

Supervised learning models

1. Regression Models: Linear Regression (SLR & MLR)
2. Logistic Regression
3. Decision Tree
4. K Nearest Neighbor- Random Forest
5. Gradient Boosting algorithms: XGboost
6. Support Vector Machine

Unsupervised learning models

1. Clustering Techniques: K means clustering
2. Apriori algorithm.
3. Model Evaluation: Over-fitting, Under-fitting
4. Model Selection-Ridge Regression- Grid Search-Model Refinement.

References:

- Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Python ,2nd edition, Wes McKinney, O'Reilly Media (2017)

Semester 4

CSDA401 Main Project

The Entire Semester is dedicated to Course CSDA601 Main Project. Each student should implement a project in Data Analytics Domain. The project should be preferably done as an internship pertaining to Data Analytics domain in a software firm. The implementation of a Research Project in the Data Analytics domain can also be considered as the Main Project. Evaluation is based on Interim Presentation, Extensive Report and Final demonstration of the Project.

CSDA402 Course Viva

A comprehensive Viva based on subjects learned during the course, by an external Examiner