OUTLINES OF TESTS, SYLLABI AND COURSES OF READINGS

CHOICE-BASED CREDIT SYSTEM

FOR

B.Sc.(Hons.) in ARTIFICIAL INTELLIGENCE AND DATA SCIENCE (SEMESTER SYSTEM)

Second Year (IIIrd & IVth Semester)

[For the Sessions 2023-24 & 2024-25]

MATA GUJRI COLLEGE
SRI FATEHGARH SAHIB
(AN AUTONOMOUS COLLEGE)

AFFILIATED TO PUNJABI UNIVERSITY, PATIALA

B.Sc.(Hons.) ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

Second Year - Third Semester Examinations

Session 2021-22 and 2022-23

Course Code	Course Type	Course Title	Load Allocation			Marks Distribution		Total Marks	Credits
			L	T	P	External	Internal		
BSCHAI-301	CC	Operating System	5	1	0	75	25	100	6
BSCHAI-302	GE-II	GE-II	3	1	0	75	25	100	4
BSCHAI-303	CC	Database Management System	3	1	0	75	25	100	4
BSCHAI-303(P)		Software Lab – V	0	0	4	50	--	50	2
BSCHAI-304	CC	Data Structures	3	1	0	75	25	100	4
BSCHAI-304(P)		Software Lab – VI	0	0	4	50	--	50	2
		Total	**14**	**4**	**8**	**400**	**100**	**500**	**22**

***General Elective-II:**

1	BSCHAI-302 E1	Computational Statistic
3	BSCHAI-302 E2	Optimization Techniques

The breakup of marks for the continuous assessment for theory paper will be as under

I	Two tests will be conducted during the semester. Both the tests will be considered for assessment.	:	50% of the marks allotted for continuous assessment
II	Assignment / Presentations	:	20% of the marks allotted for continuous assessment
III	Class participation & behaviour	:	10% of the marks allotted for continuous assessment
IV	Attendance	:	20% of the marks allotted for continuous assessment

Course Code: BSCHAI-301
COURSE NAME: OPERATING SYSTEM

Maximum Marks: 100 Time: 3 hours
Theory: 75 Marks Pass Marks: 35%
Internal Assessment: 25 Marks

Course Objectives:

- To understand the basic principles of Operating System.
- To learn the policies of scheduling, deadlocks, memory management, synchronization, system calls, and file systems.

Course Learning Outcomes:

Upon Completion of the course, Students will be able to,

- Know the important computer system resources and the role of operating system in their management policies and algorithms.
- Understand the process management policies and scheduling of processes by CPU
- Evaluate the requirement for process synchronization and coordination handled by operating system.
- Analyze the memory management and its allocation policies.
- Identify and evaluate the storage management policies with respect to different storage management technologies.

(A) INSTRUCTIONS FOR THE PAPERSETTER

The question paper will consist of three sections UNIT-I, UNIT-II, and UNIT-III. Each of UNIT-I and UNIT-II will have four questions from the respective Units of the syllabus and each question will carry 12 marks. UNIT-III will have 9 short answer type questions which will cover the entire syllabus uniformly and will carry 3 marks in all.

(B) INSTRUCTIONS FOR THECANDIDATES

Candidates are required to attempt two questions each from UNIT-I and UNIT-II. UNIT-III is Compulsory.

Unit I

Operating System –Definition, Need, Services, Types of operating systems: simple batch system, multi-programming, time sharing system, parallel system, distributed system, real time system. Operating system services, system calls.

Process Management – process definition, process states, process states transition diagram, process scheduling. Basic concepts of thread, Difference between process and thread.

CPU Scheduling – Basic concepts, scheduling criteria, scheduling algorithms – FCFS, SJF, Round Robin, Priority and Multilevel queue scheduling, Multilevel feedback queue scheduling.

Unit II

Deadlocks – Characteristics of deadlocks, methods for handling deadlocks, deadlock prevention, deadlock. Avoidance, deadlock detection and recovery.

Memory Management – Logical versus Physical address space, swapping, contiguous allocation, Paging, Concept of Virtual memory, Implementation by Demand Paging, Page replacement algorithms – FIFO, Optimal, LRU, Concept of thrashing.

File Management – Allocation methods: contiguous allocation, linked allocation and indexed allocation.

Device Management – Disk Scheduling: FCFS, SSTF, SCAN, C-SCAN, LOOK.

Text Books:

1. Abraham Silberschatz, Peter B. Galvin, Operating System Concepts, Addison –Wesley Publishing Co. Engineering, Third Edition 2005,PankajJalote, Narosa Publications. 5th Edition.
2. Andrew S. Tanenbaum, “Modern Operating Systems”, Prentice Hall of India Pvt. Ltd.

Reference Books:

1. William Stallings, “Operating System”, Prentice Hall of India.
2. Pramod Chandra P. Bhatt – “An Introduction to Operating Systems, Concepts and Practice, PHI.
3. Harvey M. Deitel, “Operating Systems”,Pearson Education Pvt. Ltd.

Teaching Plan

Week	Content
1-2	**Operating System –**Definition, Need, Services, Types of operating systems: simple batch system, multi-programming, time sharing system, parallel system, distributed system, real time system. Operating system services, system calls.
3-4	**Process Management** – process definition, process states, process states transition diagram, process scheduling. Basic concepts of thread, Difference between process and thread.
5-6	**CPU Scheduling** – Basic concepts, scheduling criteria, scheduling algorithms – FCFS, SJF, Round Robin, Priority and Multilevel queue scheduling, Multilevel feedback queue scheduling
7-8	**Deadlocks** – Characteristics of deadlocks, methods for handling deadlocks, deadlock prevention, deadlock. Avoidance, deadlock detection and recovery.
9-10	**Memory Management** – Logical versus Physical address space, swapping, contiguous allocation, Paging, Concept of Virtual memory, Implementation by Demand Paging, Page replacement algorithms – FIFO, Optimal, LRU, Concept of thrashing.
11-12	**File Management** – Allocation methods: contiguous allocation, linked allocation and indexed allocation.
13-14	**Device Management** – Disk Scheduling: FCFS, SSTF, SCAN, C-SCAN, LOOK.

Course Code: BSCHAI-302E1
COURSE NAME: COMPUTATIONAL STATISTIC

Maximum Marks: 100 Time: 3 hours
Theory: 75 Marks Pass Marks: 35%
Internal Assessment: 25 Marks

Course Objectives:

- To make students aware of estimations (Point, as well as, Internal) and testing (Simple, as well as, composite hypothesis) procedures.
- This course provides the students the ability to understand the design and conduct experiments, as well as, to analyze and interpret data.

Course Learning Outcomes:

Upon Completion of the course, Students will be able to,

- Learn applying various estimations and testing procedures to deal with real life problems.
- Understand Neyman – Pearson fundamental lemma, UMP test, interval estimation and confidence interval.
- Learn ANOVA for one and two way classification, fixed effect models with equal number of observations per cell.
- Learn how to apply statistical test in real life situations.

(A) INSTRUCTIONS FOR THE PAPERSETTER

The question paper will consist of three sections UNIT-I, UNIT-II, and UNIT-III. Each of UNIT-I and UNIT-II will have four questions from the respective Units of the syllabus and each question will carry 12 marks. UNIT-III will have 9 short answer type questions which will cover the entire syllabus uniformly and will carry 3 marks in all.

(B) INSTRUCTIONS FOR THECANDIDATES

Candidates are required to attempt two questions each from UNIT-I and UNIT-II. UNIT-III is Compulsory.

Unit-I

Point estimation: Estimators and estimates, criteria for good estimator- unbiasedness, consistency, efficiency and sufficiency. Neyman-factorization theorem (only statement and examples)

Methods of Estimation: Maximum likelihood method of estimation and method of moments (properties and examples).

Testing of Hypotheses: The concept of statistical hypotheses, null and alternative hypotheses, simple and composite hypotheses, critical region, level of significance, two types of errors, power of the test. Applications of z-test, t-test, Chi-square and F distributions.

Linear Models: The fixed effect models, the distribution of minimum error sum of squares and the conditional error sum of squares, tests of general linear-hypotheses.

ANOVA: Analysis of one way classified data under the fixed effect model. Analysis of the two way classified data with one observation per cell and multiple but equal observations in cells under the fixed effect model.

Unit-II

Terminology in experimental designs: Basic principles of design- Randomization, replication and local control, Need for design of experiment. Completely Randomized design, randomized block design.

Regression Models: Curve Fitting: Polynomial Regression, Exponential Regression and Geometric Regression.

Index Numbers: Definition, Interpretation and applications of Index Numbers, Problems involved in the construction of Index Numbers, Laspeyre's, Paasche's, Marshal-Edgeworth and Fisher's formulae for Index Numbers, Criterion of Good Index Number.

Time Series: Definition, Components of Time Series, Measurement of Secular Trend by method of Moving Average and fitting of mathematical Curves.

Text Books:

1. Goon, A.M., Gupta, M.K. and Dasgupta, B.: Fundamentals of Statistics, Vol. II, World Press, ED6th (revised and enlarged), 2008.
2. Gupta, S.C. and Kapoor, V.K.: Fundamentals of Applied Statistics, Sultan Chand & Sons Educational Pub. New Delhi2014.
3. Gupta, S.C. and Kapoor, V.K.: Fundamental of Mathematical Statistics, Sultan Chand & Sons Educational Pub. New Delhi, 2019.

Reference Books:

1. Miller, I. and Miller, M.: John E. Freund's Mathematical Statistics, Pearson Education, Inc., Edition 6th, 2002.

Teaching Plan

Week	Content
1-2	**Point estimation:** Estimators and estimates, criteria for good estimator- unbiasedness, consistency, efficiency and sufficiency. Neyman-factorization theorem (only statement and examples) **Methods of Estimation:** Maximum likelihood method of estimation and method of moments (properties and examples).
3-4	**Testing of Hypotheses**: The concept of statistical hypotheses, null and alternative hypotheses, simple and composite hypotheses, critical region, level of significance, two types of errors, power of the test. Applications of z-test, t-test, Chi-square and F distributions.
5-6	**Linear Models:** The fixed effect models, the distribution of minimum error sum of squares and the conditional error sum of squares, tests of general linear-hypotheses. **ANOVA:** Analysis of one way classified data under the fixed effect model. Analysis of the two way classified data with one observation per cell and multiple but equal observations in cells under the fixed effect model.
7-8	**Terminology in experimental designs:** Basic principles of design- Randomization, replication and local control, Need for design of experiment. Completely Randomized design, randomized block design.
9-10	**Regression Models:** Curve Fitting: Polynomial Regression, Exponential Regression and Geometric Regression.
11-12	**Index Numbers:** Definition, Interpretation and applications of Index Numbers, Problems involved in the construction of Index Numbers, Laspeyre's, Paasche's, Marshal-Edgeworth and Fisher's formulae for Index Numbers, Criterion of Good Index Number.
13-14	**Time Series:** Definition, Components of Time Series, Measurement of Secular Trend by method of Moving Average and fitting of mathematical Curves.

Course Code: BSCHAI-302E2
COURSE NAME: OPTIMIZATION TECHNIQUES

Maximum Marks: 100 Time: 3 hours
Theory: 75 Marks Pass Marks: 35%
Internal Assessment: 25 Marks

Course Objectives:

- The main aim of course is to make students able to formulate linear programming models for service and manufacturing system as well as to formulate network models and apply operations research techniques.

Course Learning Outcomes:

Upon Completion of the course, Students will be able to,

- Mathematically formulate an applied word problem involving revenue, costs and constraints as a linear problem.
- Apply simplex algorithm to solve a linear program problem as well as to apply M – Method, Two phase method.
- Produce the dual of a linear problem.
- Solve PERT, CPM network problems.
- Formulate network models for service and manufacturing systems and apply operations research techniques and algorithm to solve network problems.

(A) INSTRUCTIONS FOR THE PAPERSETTER

The question paper will consist of three sections UNIT-I, UNIT-II, and UNIT-III. Each of UNIT-I and UNIT-II will have four questions from the respective Units of the syllabus and each question will carry 12 marks. UNIT-III will have 9 short answer type questions which will cover the entire syllabus uniformly and will carry 3 marks in all.

(B) INSTRUCTIONS FOR THECANDIDATES

Candidates are required to attempt two questions each from UNIT-I and UNIT-II. UNIT-III is Compulsory.

Unit-I

Linear Programming: Essentials of Linear Programming Model, Properties of Linear Programming Model, Formulation of Linear Programming, General Linear Programming Model, Maximization & Minimization Models, Graphical Method for Solving Linear Programming problems, Unbounded LP Problem, Additional Variables Used In Solving LPP, Maximization Case, Minimization Problems, Big M Method, Degeneracy in LP Problems, Unbounded Solutions in LPP, Multiple Solutions in LPP.

CPM/PERT:PERT/CPM Network Components, Rules in Constructing a Network, Scheduling of Activities: Earliest Time and Latest Time, Determination of Float and Slack Times, Critical Path method for project management, Project Evaluation Review Technique – PERT, Gantt chart(time chart). Terminology.

Sequencing: Types of Sequencing Problems, Algorithm for Solving Sequencing Problems, Processing n jobs through 2, 3, m machines. Processing 2 jobs through m machines.

Unit-II

Transportation: General Mathematical model of transportation problem, The transportation algorithm, Method of finding initial solution: North west corner method, Least cost method, Vogel's Approximation method, Test for optimality: MODI method, Variation in transportation problems.

Game Theory: Terminologies of game theory, Two-person-zero-sum-game, Game with pure strategy, Methods of solving game with mixed strategy, Dominance Property, Graphical method for 2xn and mx2 games. Linear Programming approach for games theory,

Inventory Management: Inventory Control Models: Purchase model with instantaneous replenishment with and without shortages, calculate EOQ.

Text Books.

1. Hira, D. S. and Gupta, P. K. Operations Research, Sultan Chand and Sons, New Delhi, 2008.
2. Sharma, S.D. Operations Research, KedarNath Ram Nath, India, Edition 4th, 2012.
3. Swarup, K., Gupta, P.K. and Man Mohan. Operations Research, Sultan Chand and Sons, New Delhi, 2nd Edition.

Teaching Plan

Week	Content
1-2	**Linear Programming**: Essentials of Linear Programming Model, Properties of Linear Programming Model, Formulation of Linear Programming, General Linear Programming Model, Maximization & Minimization Models
3-4	Graphical Method for Solving Linear Programming problems, Unbounded LP Problem, Additional Variables Used In Solving LPP, Maximization Case, Minimization Problems, Big M Method, Degeneracy in LP Problems, Unbounded Solutions in LPP, Multiple Solutions in LPP.
5-6	**CPM/PERT:PERT/CPM** Network Components, Rules in Constructing a Network, Scheduling of Activities: Earliest Time and Latest Time, Determination of Float and Slack Times, Critical Path method for project management, Project Evaluation Review Technique – PERT, Gantt chart(time chart). Terminology.
7-8	**Sequencing:** Types of Sequencing Problems, Algorithm for Solving Sequencing Problems, Processing n jobs through 2, 3, m machines. Processing 2 jobs through m machines.
9-10	**Transportation:** General Mathematical model of transportation problem, The transportation algorithm, Method of finding initial solution: North west corner method, Least cost method, Vogel's Approximation method, Test for optimality: MODI method, Variation in transportation problems
11-12	**Game Theory**: Terminologies of game theory, Two-person-zero-sum-game, Game with pure strategy, Methods of solving game with mixed strategy, Dominance Property, Graphical method for 2xn and mx2 games. Linear Programming approach for games theory.
13-14	**Inventory Management:** Inventory Control Models: Purchase model with instantaneous replenishment with and without shortages, calculate EOQ.

Course Code: BSCHAI-303
COURSE NAME: DATABASE MANAGEMENT SYSTEM

Maximum Marks: 100 Time: 3 hours
Theory: 75 Marks Pass Marks: 35%
Internal Assessment: 25 Marks

Course Objectives:

- To present an introduction to database management systems, with an emphasis on how to organize, maintain and retrieve - efficiently, and effectively - information from a DBMS.
- To provide basic introduction to database system technologies.

Course Learning Outcomes:

Upon Completion of the course, Students will be able to,

- Know the fundamental elements of database management systems
- Know the basic concepts of relational data model, entity-relationship model, relational database design, relational algebra and SQL.
- Design ER-models to represent simple database application scenarios
- Convert the ER-model to relational tables, populate relational database and formulate SQL queries on data.
- Improve the database design by normalization.

(A) INSTRUCTIONS FOR THE PAPERSETTER

The question paper will consist of three sections UNIT-I, UNIT-II, and UNIT-III. Each of UNIT-I and UNIT-II will have four questions from the respective Units of the syllabus and each question will carry 12 marks. UNIT-III will have 9 short answer type questions which will cover the entire syllabus uniformly and will carry 3 marks in all.

(B) INSTRUCTIONS FOR THE CANDIDATES

Candidates are required to attempt two questions each from UNIT-I and UNIT-II. UNIT-III is Compulsory.

Unit-I

Traditional file processing system: characteristics, limitations. Database: definition.

DBMS: definition, characteristics, advantages, database schema, instance, DBA and its responsibilities, three level architecture of DBMS, mapping between different levels, Data independence.

Database languages: DML, DCL, DDL Keys: primary Key, candidate Key, Super Key, Foreign Key, Composite Key

E-R Model :Definition, Entity and Relationship, cardinality of a relationship, E-R Diagram Notations, Modelling using E-R Diagrams, weak entity sets, strong entity sets, converting E-R diagrams to tables. Aggregation, Generalization, Specialization.

Data Base Models: Hierarchical, Network, Relational Models & their difference.

Unit-II

Normalization: Definition, Need, Functional Dependency, Full Functional Dependency, Partial Dependency, Transitive dependency, Multivalued Dependency, Types of Normal Forms:(1NF,2NF,3NF,BCNF).

Database concurrency: definition, Problems of concurrency: Lost Update, Dirty read, Incorrect Summary. Concurrency control techniques: Locking, Deadlock.

Introduction to Database Recovery: Transaction Log: Deferred Update

Introduction to Database Security: Privileges, Granting and Revoking Privileges and roles

Text Books:

1. Fundamentals of Database Systems, Third Edition, by Elmasri/Navathe
2. Korth and Silberschatz Abraham, Database

Reference Books:

1. An introduction to Database Systems by C.J.Date.
2. An introduction to Database Systems by Bipin C.Desai.
3. SQL, PL/SQL, The programming language of oracle, Ivan Bayross. BPB Publication.

Teaching Plan

Week	Content
1-2	**Traditional file processing system:** characteristics, limitations. Database: definition. **DBMS:** definition, characteristics, advantages, database schema, instance, DBA and its responsibilities, three level architecture of DBMS, mapping between different levels, Data independence.
3-4	**Database languages:** DML, DCL, DDL Keys: primary Key, candidate Key, Super Key, Foreign Key, Composite Key
5-6	**E-R Model** :Definition, Entity and Relationship, cardinality of a relationship, E-R Diagram Notations, Modelling using E-R Diagrams, weak entity sets, strong entity sets, converting E-R diagrams to tables. Aggregation, Generalization, Specialization
7-8	**Data Base Models**: Hierarchical, Network, Relational Models & their difference.
9-10	**Normalization:** Definition, Need, Functional Dependency, Full Functional Dependency, Partial Dependency, Transitive dependency, Multivalued Dependency, Types of Normal Forms:(1NF,2NF,3NF,BCNF).
11-12	**Database concurrency**: definition, Problems of concurrency: Lost Update, Dirty read, Incorrect Summary. Concurrency control techniques: Locking, Deadlock.
13-14	**Introduction to Database Recovery:** Transaction Log: Deferred Update **Introduction to Database Security:** Privileges, Granting and Revoking Privileges and roles

Course Code: BSCHAI-303(P)
COURSE NAME: SOFTWARE LAB-V

Maximum Marks: 50 Time: 3 hours
External Examination: 50 Marks Pass Marks: 35%
Practical Sessions to be conducted: 40-50 Hrs

Course Objectives:

- Understand basic concepts of how a database stores information via tables
- Understanding of SQL syntax used with SQL
- Learn how to retrieve and manipulate data from one or more tables
- Know how to filter data based upon multiple conditions
- Updating and inserting data into existing tables
- Learning how the relationships between tables will affect the SQL.
- The advantages of store procedures with storing data using variables and functions.

Course Learning Outcomes:

Upon Completion of the course, Students will be able to,

- Understand, appreciate and effectively explain the underlying concepts of database technologies
- Design and implement a database schema for a given problem-domain
- Normalize a database
- Populate and query a database using SQL DML/DDL commands.
- Program PL/SQL including stored procedures, stored functions, cursors, packages

The student should execute various SQL Queries

Introduction to SQL and installation of SQL Server / Oracle, Data Types, Creating Tables, Retrieval of Rows using Select Statement, Conditional Retrieval of Rows, Alter and Drop Statements.

Working with Null Values, Matching a Pattern from a Table, Ordering the Result of a Query, Aggregate Functions, Grouping the Result of a Query, Update and Delete Statements.

Set Operators, Nested Queries, Joins, Sequences.

Views, Indexes, Database Security and Privileges: Grant and Revoke Commands, Commit and Rollback Commands.

PL/SQL Architecture, Assignments and Expressions, Writing PL/SQL Code, Referencing Non-SQL parameters.

Stored Procedures and Exception Handling.

Triggers and Cursor Management in PL/SQL.

The breakup of marks for the practical will be as under

I.	Lab Record (Internal Assessment)	10Marks
II.	Viva Voce (External Evaluation)	20Marks
III.	Program Development and Execution(ExternalEvaluation)	20Marks

Course Code: BSCHAI-304
COURSE NAME: DATA STRUCTURES

Maximum Marks: 100 Time: 3 hours
Theory: 75 Marks Pass Marks: 35%
Internal Assessment: 25 Marks

Course Objectives:

- To introduce the fundamental concept of data structures and to emphasize the importance of data structures in developing and implementing efficient algorithms.
- To introduce fundamentals of different types of data structures and also the ways to implement them.
- To develop application using data structures.
- To understand concepts about searching and sorting techniques

Course Learning Outcomes:

Upon Completion of the course, Students will be able to,

- Know how arrays, linked structures, stacks, queues and trees are represented in memory and used by algorithms.
- Know common applications for arrays, linked structures, stacks, queues and trees.
- Compare and contrast the benefits of dynamic and static data structures implementations.

(A) INSTRUCTIONS FOR THE PAPER.SETTER

The question paper will consist of three sections UNIT-I, UNIT-II, and UNIT-III. Each of UNIT-I and UNIT-II will have four questions from the respective Units of the syllabus and each question will carry 12 marks. UNIT-III will have 9 short answer type questions which will cover the entire syllabus uniformly and will carry 3 marks in all.

(B) INSTRUCTIONS FOR THECANDIDATES

Candidates are required to attempt two questions each from UNIT-I and UNIT-II. UNIT-III is Compulsory.

Unit – I

Data Structure: Introduction to data structure and algorithm, complexity of an algorithm. Algorithm analysis: Time space trade off, Big O notation, Algorithmic notations & Complexity.
Arrays: Introduction, one dimensional and multidimensional array, memory representation of arrays, Operations on arrays: Insertion, Deletion, searching, sorting.
Stacks: Introduction, Operation on stacks, Implementation of stacks, Application of stacks: evaluation of arithmetic expressions, Parenthesis matching, String Reversal,
Polish & Reverse Polish Notation.
Queues: Introduction, operation on queues, circular queue, memory representation of queues, Dequeues, Priority queues, application of queues.

Unit – II

Linked List: Introduction to operation on linked list, circular linked list, doubly linked list, header linked list, implementation of linked list, application of linked lists.
Trees: Introduction to Trees, Binary Tree; Binary Search Tree, Heaps: Insertion and Deletion.
Searching: Linear search, Binary Search.
Graphs: Introduction, Memory Representation, Graph Traversal(DFS and BFS)
Sorting: Bubble Sort, Selection Sort, Insertion Sort, Merge Sort, Shell Sort, Radix Sort, Quick Sort .

Text Books:

1. SeymourLipschultz, “Data Structures using C", McGraw-Hill,
2. Tanenbaum, Y. Lanhgsam and A.J. Augenstein, "Data Structures Using C", Prentice Hall of India,
Loomis, “Data and File Structures”.

References:

1. SeymourLipschultz, “Theory and Problems of Data Structures", McGraw-Hill.
2. E. Horowitz and S. Sahni, “Data Structures with Pascal", Galgotia, 3rd Edition,
3. RobertSedgewick, “Algorithms in C”, Pearson Education.
4. M. J. Folk, B. Zoellick, G Riccardi, “File Structures”, Pearson Education

Teaching Plan

Week	Content
1-2	**Data Structure:** Introduction to data structure and algorithm, complexity of an algorithm. Algorithm analysis: Time space trade off, Big O notation, Algorithmic notations & Complexity.
3-4	**Arrays:** Introduction, one dimensional and multidimensional array, memory representation of arrays, Operations on arrays: Insertion, Deletion, searching, sorting
5-6	**Stacks:** Introduction, Operation on stacks, Implementation of stacks, Application of stacks: evaluation of arithmetic expressions, Parenthesis matching, String Reversal, Polish & Reverse Polish Notation.
7-8	**Stacks:** Introduction, Polish & Reverse Polish Notation, Operation on stacks, Application of stacks: evaluation of arithmetic expressions, Parenthesis matching, String Reversal.
9-10	**Queues:** Introduction, operation on queues, circular queue, memory representation of queues, Dequeues, Priority queues, application of queues.
11-12	**Linked List:** Introduction to operation on linked list, circular linked list, doubly linked list, header linked list, implementation of linked list, application of linked lists. **Trees:** Introduction to Trees, Binary Tree; Binary Search Tree, Heaps: Insertion and Deletion.
13-14	**Searching:** Linear search, Binary Search. **Graphs:** Introduction, Memory Representation, Graph Traversal(DFS and BFS) **Sorting:** Bubble Sort, Selection Sort, Insertion Sort, Merge Sort, Shell Sort, Radix Sort, Quick Sort .

Course Code: BSCHAI-304(P)
COURSE NAME: SOFTWARE LAB-VI

Maximum Marks: 50 Time: 3 hours
External Examination: 50 Marks Pass Marks: 35%
Practical Sessions to be conducted: 40-50 Hrs

Course Objectives:

- The course is designed to develop skills to design and analyze simple linear and non linear data structures.
- It strengthen the ability to the students to identify and apply the suitable data structure for the given real world problem.
- It enables them to gain knowledge in practical applications of data structures.

Course Learning Outcomes:

Upon Completion of the course, Students will be able to,

- Analyze and Compare various linear and non-linear data structures
- Code, debug and demonstrate the working nature of different types of data structures and their applications
- Implement, analyze and evaluate the searching and sorting algorithms
- Choose the appropriate data structure for solving real world problems

The student should implement various data structures learnt under the subject BSCHAI-304 as prescribed by subject teacher.

Support the program with functions for each of the above operations.

The breakup of marks for the practical will be as under

I.	Lab Record (Internal Assessment)	10Marks
II.	VivaVoce (External Evaluation)	20Marks
III.	Program Development and Execution (External Evaluation)	20Marks

B.Sc.(Hons.) ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

Second Year - Fourth Semester Examinations

Session 2023-24 and 2024-25

Course Code	Course Type	Course Title	Load Allocation			Marks Distribution		Total Marks	Credits
			L	T	P	External	Internal		
BSCHAI-401	CC	Mathematics for Data Science	5	1	0	75	25	100	6
BSCHAI-402	DSE	DSE – I	4	0	4	75	25	100	6
BSCHAI-403	CC	Data Warehousing & Data Mining	3	1	0	75	25	100	4
BSCHAI-403(P)		Software Lab – VII	0	0	4	50	--	50	2
BSCHAI-404	CC	Data Analysis using Python	3	1	0	75	25	100	4
BSCHAI-404(P)		Software Lab – VIII	0	0	4	50	--	50	2
DA-4001		Drug Abuse (Q)							
		Total	**15**	**3**	**12**	**400**	**100**	**500**	**24**

*** Discipline Specific Elective-I:**

1	BSCHAI-402 DSE1	Cloud Computing
2	BSCHAI-402 DSE2	Information Security

The breakup of marks for the continuous assessment for theory paper will be as under

I	Two tests will be conducted during the semester. Both the tests will be considered for assessment.	:	50% of the marks allotted for continuous assessment
II	Assignment / Presentations	:	20% of the marks allotted for continuous assessment
III	Attendance, Class participation & behaviour	:	10% of the marks allotted for continuous assessment
IV	Attendance	:	20% of the marks allotted for continuous assessment

Course Code: BSCHAI-401
COURSE NAME: MATHEMATICS FOR DATA SCIENCE

Maximum Marks: 100 Time: 3 hours
Theory: 75 Marks Pass Marks: 35%
Internal Assessment: 25 Marks

Course Objectives:

- To recognize the algebraic structure: group, ring, vector space and subspace
- To understand the fundamental concept of Linear Transformation
- To enable the identification of square matrix as operator
- To understand the concept of Boolean Algebra and Lattice.

Course Learning Outcomes:

Upon Completion of the course, Students will be able to:

- Understand the concept of Recurrence Relation and generating functions.
- Learn the concept of Boolean algebra.
- Understand matrices and their application to system of linear equations.
- Familiarize with characteristic roots and characteristic vectors.
- Find the inverse of a matrix by Cayley-Hamilton theorem.
- Understand the concept of Linear Transformation

(A) INSTRUCTIONS FOR THE PAPERSETTER

The question paper will consist of three sections UNIT-I, UNIT-II, and UNIT-III. Each of UNIT-I and UNIT-II will have four questions from the respective Units of the syllabus and each question will carry 12 marks. UNIT-III will have 9 short answer type questions which will cover the entire syllabus uniformly and will carry 3 marks in all.

(B) INSTRUCTIONS FOR THE CANDIDATES

Candidates are required to attempt two questions each from UNIT-I and UNIT-II. UNIT-III is Compulsory.

Unit-I

Boolean Algebra: Lattices and Algebraic structure, Duality, Distribution and Complemented Lattices, Boolean Lattices and Boolean Algebras, Boolean functions and Expressions, Switching Circuits.

Recurrence relations and Generating Functions, solution of Recurrence relations using Generating Functions; Permutations and Combinations; Pigeon Hole Principle.

Matrices & Determinants: Matrix, Definition, types, Addition, Subtraction, Multiplication of Matrices, Singular and Non-Singular Matrices, Rank of a Matrix, Solution of simultaneous Equations, Cayley Hamilton Theorem, Eigen Values & Eigen Vectors, Diagonalization of a Matrix.

Unit-II

Matrix Algebra and Linear Algebra: Introduction of groups, rings and Vector Spaces. Linear Independence and Dependence of Vectors, Linear Combination. Basis and Dimension of Vector space, Sub-Space, Intersection, Union of sub Spaces.

Linear Transformation: Matrices as Linear Mapping, Kernel and Image. Statement of Rank Nullity Theorem, Singular and Non- Singular Linear Mappings.

Text Books:

1. Seymour Lipschutz, Marc Lipson, Linear Algebra, Schaum Series
2. Seymour Lipschutz , Marc Lipson , H. Patil, Discrete Structures, Schaum Series
3. Elliott Mendelson , Frank Ayres , Calculus, Schaum Series

Reference Books:

1. Coding the Matrix: Linear Algebra Through Applications to Computer Science by Philip N. Klein, Lightning Source Inc.
2. Linear Algebra Done Right by Sheldon Axler, Springer.
3. Linear Algebra by Kenneth M. Hoffman, Ray Kunze, Pearson.
4. Linear Algebra and its Applications by David C. Lay and Steven R. Lay, Pearson.

Teaching Plan

Week	Content
1-2	**Boolean Algebra:** Lattices and Algebraic structure, Duality, Distribution and Complemented Lattices, Boolean Lattices and Boolean Algebras, Boolean functions and Expressions, Switching Circuits.
3-4	Recurrence relations and Generating Functions, solution of Recurrence relations using Generating Functions; Permutations and Combinations; Pigeon Hole Principle.
5-6	**Matrices & Determinants**: Matrix, Definition, types, Addition, Subtraction, Multiplication of Matrices, Singular and Non-Singular Matrices, Rank of a Matrix
7-8	Solution of simultaneous Equations, Cayley Hamilton Theorem, Eigen Values & Eigen Vectors, Diagonalization of a Matrix.
9-10	**Matrix Algebra and Linear Algebra:** Introduction of groups, rings and Vector Spaces. Linear Independence and Dependence of Vectors, Linear Combination.
11-12	Basis and Dimension of Vector space, Sub-Space, Intersection, Union of sub Spaces.
13-14	**Linear Transformation:** Matrices as Linear Mapping, Kernel and Image. Statement of Rank Nullity Theorem, Singular and Non- Singular Linear Mappings.

Course Code: BSCHAI-402DSE1
COURSE NAME: CLOUD COMPUTING

Maximum Marks: 100 Time: 3 hours
Theory: 75 Marks Pass Marks: 35%
Internal Assessment: 25 Marks

Course Objectives:

- To provide students with the fundamentals and essentials of Cloud Computing.
- To provide students a sound foundation of the Cloud Computing so that they are able to start using and adopting Cloud Computing services and tools in their real life scenarios.
- To enable students exploring some important cloud computing driven commercial systems and applications.
- To expose the students to frontier areas of Cloud Computing and information systems, while providing sufficient foundations to enable further study and research.

Course Learning Outcomes:

Upon Completion of the course, Students will be able to,

- Explain the core concepts of the cloud computing paradigm: how and why this paradigm shift came about, the characteristics, advantages and challenges brought about by the various models and services in cloud computing.
- Apply the fundamental concepts in datacenters to understand the tradeoffs in power, efficiency and cost.
- Identify resource management fundamentals, i.e. resource abstraction, sharing and sandboxing and outline their role in managing infrastructure in cloud computing.
- Analyze various cloud programming models and apply them to solve problems on the cloud.

(A) INSTRUCTIONS FOR THE PAPERSETTER

The question paper will consist of three sections UNIT-I, UNIT-II, and UNIT-III. Each of UNIT-I and UNIT-II will have four questions from the respective Units of the syllabus and each question will carry 12 marks. UNIT-III will have 9 short answer type questions which will cover the entire syllabus uniformly and will carry 3 marks in all.

(B) INSTRUCTIONS FOR THECANDIDATES

Candidates are required to attempt two questions each from UNIT-I and UNIT-II. UNIT-III is Compulsory.

Unit – I

Introduction : Definition of cloud, characteristics of cloud, historical developments & challenges ahead, the vision of cloud computing, Driving factors towards cloud, Comparing grid with utility computing, cloud computing and other computing systems, types of workload patterns for the cloud, IT as a service, Applications of cloud computing.

Cloud computing concepts: Introduction to virtualization techniques, Characteristics of virtualization, Pros and Cons of virtualization Technology, Hypervisors, Types of hypervisors, Multitenancy, Application programming interfaces (API), Elasticity and scalability.

Unit – II

Cloud service models: Cloud service models, Infrastructure as a service (IaaS) architecture- details and example, Platform as a service (PaaS) architecture- details and example, Software as a service (SaaS) architecture-- details and example, Comparison of cloud service delivery models.

Cloud deployment models: Introduction to cloud deployment models, Public clouds, Private clouds, Hybrid clouds, Community clouds, Migration paths for cloud, Selection criteria for cloud deployment

Security in cloud computing: Understanding security risks, Principal security dangers to cloud computing, Internal security breaches, User account and service hijacking, measures to reduce cloud security breaches Case Studies: Comparison of existing Cloud platforms /Web Services

Text Books:

1.Cloud Computing Explained: Implementation Handbook for Enterprises, John Rhoton, Publication

Reference Books:

1. Cloud Computing Bible, Barrie Sosinsky, Wiley-India, 2010
2. Cloud Computing: Principles and Paradigms, Editors: RajkumarBuyya, James Broberg, Andrzej M. Goscinski, Wile
3. Cloud Security: A Comprehensive Guide to Secure Cloud Computing, Ronald L. Krutz, Russell Dean Vines, Wiley-India, 2010
4. Gautam Shroff, Enterprise Cloud Computing Technology Architecture Applications , Adobe Reader ebooks available from eBooks.com,2010
5. Toby Velte, Anthony Velte, Robert Elsenpeter, Cloud Computing, A Practical Approach ,McGraw Hills, 2010.

Teaching Plan

Week	Content
1-2	**Overview of Computing Paradigm**: Recent trends in Computing: Grid Computing, Cluster Computing, Distributed Computing, Utility Computing, Cloud Computing.
3-4	**Introduction to Cloud Computing**: History of Cloud Computing, Cloud service providers, Benefits and limitations of Cloud Computing.
5-6	**Cloud Computing Architecture**: Comparison with traditional computing architecture (client/server), Services provided at various levels, Service Models- Infrastructure as a Service(IaaS), Platform as a Service(PaaS), Software as a Service(SaaS).
7-8	How Cloud Computing Works, Deployment Models- Public cloud, Private cloud, Hybrid cloud, Community cloud, Case study of NIST architecture.
9-10	**Case Studies:** Case study of Service model using Google App Engine, Microsoft Azure, Amazon EC2 , Eucalyptus.
11-12	**Service Management in Cloud Computing:** Service Level Agreements (SLAs), Billing & Accounting, Comparing Scaling Hardware: Traditional vs. Cloud, Economics of scaling.
13-14	**Cloud Security:** Infrastructure Security- Network level security, Host level security, Application level security, Data security and Storage- Data privacy and security Issues, Jurisdictional issues raised by Data location, Authentication in cloud computing.

Course Code: BSCHAI-402DSE2
COURSE NAME: INFORMATION SECURITY

Maximum Marks: 100 Time: 3 hours
Theory: 75 Marks Pass Marks: 35%
Internal Assessment: 25 Marks

Course Objectives:

- In this course students learn basics of information security. in both management aspect and technical aspect.
- Students understand of various types of security incidents and attacks, and learn methods to prevent detect and react incidents and attacks.
- Students will also learn basics of application of cryptography which are one of the key technologies to implement security function.

Course Learning Outcomes:

Upon Completion of the course, Students will be able to,

- Evaluate vulnerability of an information system and establish a plan for risk management.
- Evaluate the authentication and encryption needs of an information system.
- Know how to secure a database and network.

(A) INSTRUCTIONS FOR THE PAPERSETTER

The question paper will consist of three sections UNIT-I, UNIT-II, and UNIT-III. Each of UNIT-I and UNIT-II will have four questions from the respective Units of the syllabus and each question will carry 12 marks. UNIT-III will have 9 short answer type questions which will cover the entire syllabus uniformly and will carry 3 marks in all.

(B) INSTRUCTIONS FOR THECANDIDATES

Candidates are required to attempt two questions each from UNIT-I and UNIT-II. UNIT-III is Compulsory.

Unit-I

Introduction: Security, Attacks, Computer Criminals, Security Services, Security Mechanisms.

Cryptography: Substitution ciphers, Transpositions Cipher, Confusion, diffusion, Symmetric, Asymmetric Encryption. DES Modes of DES, Uses of Encryption, Hash function, key exchange, Digital Signatures, Digital Certificates.

Program Security: Secure programs, Non malicious Program errors, Malicious codes virus, Trap doors, Salami attacks, Covert channels, Control against program

Unit–II

Threats: Protection in OS: Memory and Address Protection, Access control, File Protection, User Authentication.

Database Security: Requirements, Reliability, Integrity, Sensitive data, Inference, Multilevel Security.

Security in Networks: Threats in Networks, Security Controls, firewalls, Intrusion detection systems, Secure e-mails

Administrating Security: Security Planning, Risk Analysis, Organisational Security Policy, Physical Security. Ethical issues in Security: Protecting Programs and data. Information and law.

Text Books:

1) Computer Security: Principles and Practice, William Stallings; Lawrie Brown
2) Introduction to Information Security and Cyber Laws Paperback-by Surya Prakash Tripathi, Ritendra Goel, Praveen Kumar Shukla
3) Principles of Information Security.Paperback-by Whitman

Reference Books:

1. Introduction to Computer Security, 2004 Matt Bishop, Addison-Wesley, ISBN 0-321-24744-
2. Buchmann J. A., Introduction to Cryptography, Springer Verlag (2001).
3. Stallings William, Cryptography and Network Security, Pearson Education (2006).
4. Schneier Bruce, Applied Cryptography, John Wiley and Sons (1996)

Teaching Plan

Week	Content
1-2	**Introduction:** Security, Attacks, Computer Criminals, Security Services, Security Mechanisms.
3-4	**Cryptography:** Substitution ciphers, Transpositions Cipher, Confusion, diffusion, Symmetric, Asymmetric Encryption. DES Modes of DES, Uses of Encryption, Hash function, key exchange, Digital Signatures, Digital Certificates.
5-6	**Program Security:** Secure programs, Non malicious Program errors, Malicious codes virus, Trap doors, Salami attacks, Covert channels, Control against program.
7-8	**Threats:** Protection in OS: Memory and Address Protection, Access control, File Protection, User Authentication.
9-10	**Database Security:** Requirements, Reliability, Integrity, Sensitive data, Inference, Multilevel Security.
11-12	**Security in Networks:** Threats in Networks, Security Controls, firewalls, Intrusion detection systems, Secure e-mails.
13-14	**Administrating Security:** Security Planning, Risk Analysis, Organisational Security Policy, Physical Security. Ethical issues in Security: Protecting Programs and data. Information and law.

Course Code: BSCHAI-403
COURSE NAME: DATA WAREHOUSING AND DATA MINING

Maximum Marks: 100 Time: 3 hours
Theory: 75 Marks Pass Marks: 35%
Internal Assessment: 25 Marks

Course Objectives:

- To understand data warehouse concepts, architecture, business analysis and tools
- To understand data pre-processing and data visualization techniques
- To study algorithms for finding hidden and interesting patterns in data
- To understand and apply various classification and clustering techniques using tools.

Course Learning Outcomes:

Upon completion of the course, the students should be able to:

- Design a Data warehouse system and perform business analysis with OLAP tools.
- Apply suitable pre-processing and visualization techniques for data analysis
- Apply frequent pattern and association rule mining techniques for data analysis
- Apply appropriate classification techniques for data analysis

(A) INSTRUCTIONS FOR THE PAPERSETTER

The question paper will consist of three sections UNIT-I, UNIT-II, and UNIT-III. Each of UNIT-I and UNIT-II will have four questions from the respective Units of the syllabus and each question will carry 12 marks. UNIT-III will have 9 short answer type questions which will cover the entire syllabus uniformly and will carry 3 marks in all.

(B) INSTRUCTIONS FOR THECANDIDATES

Candidates are required to attempt two questions each from UNIT-I and UNIT-II. UNIT-III is Compulsory.

Unit-I

Data Warehousing & modelling: Basic Concepts: Data Warehousing: A multitier Architecture, Data warehouse models: Enterprise warehouse, Data mart and virtual warehouse, Extraction, Transformation and loading, Data Cube: A multidimensional data model, Stars, Snowflakes and Fact constellations: Schemas for multidimensional Data models, Dimensions: The role of concept Hierarchies, Measures: Their Categorization and computation, Typical OLAP Operations.

Data warehouse implementation: Efficient Data Cube computation: An overview, Indexing OLAP Data: Bitmap index and join index, Efficient processing of OLAP Queries, OLAP server Architecture ROLAP versus MOLAP Versus HOLAP.

Data Mining: Introduction, Challenges, Data Mining Tasks, Data: Types of Data, Data Quality, Data Pre-processing, Measures of Similarity and Dissimilarity.

Unit-II

Association Analysis: Problem Definition, Frequent Item set Generation, Rule generation. Alternative Methods for Generating Frequent Item sets, FP-Growth Algorithm, Evaluation of Association Patterns.

Classification: Decision Trees Induction, Method for Comparing Classifiers, Rule Based Classifiers, Nearest Neighbor Classifiers, Bayesian Classifiers.

Clustering Analysis: Overview, K-Means, Agglomerative Hierarchical Clustering, DBSCAN, Cluster Evaluation, Density-Based Clustering, Graph- Based Clustering, Scalable Clustering Algorithms.

Text Books:

1. Alex Berson and Stephen J.Smith, “Data Warehousing, Data Mining and OLAP”, Tata McGraw – Hill Edition, Thirteenth Reprint 2008.

2. Jiawei Han and MichelineKamber, “Data Mining Concepts and Techniques”, Third Edition, Elsevier, 2012.

References:

1. Pang-Ning Tan, Michael Steinbach and Vipin Kumar, "Introduction to Data Mining", Person Education, 2007.

2. K.P. Soman, ShyamDiwakar and V. Aja, "Insight into Data Mining Theory and Practice", Eastern Economy Edition, Prentice Hall of India, 2006.

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

4. Daniel T.Larose, "Data Mining Methods and Models", Wiley-Interscience, 2006.

5. Modern Data Warehousing, Mining & Visualization Core Concepts, George M Marakas, First Edition, Pearson Education

Teaching Plan

Week	Content
1-2	**Data Warehousing & modelling:** Basic Concepts: Data Warehousing: A multitier Architecture, Data warehouse models: Enterprise warehouse, Data mart and virtual warehouse, Extraction, Transformation and loading
3-4	**Data Cube:** A multidimensional data model, Stars, Snowflakes and Fact constellations: Schemas for multidimensional Data models, Dimensions: The role of concept Hierarchies, Measures: Their Categorization and computation, Typical OLAP Operations.
5-6	**Data warehouse implementation**: Efficient Data Cube computation: An overview, Indexing OLAP Data: Bitmap index and join index, Efficient processing of OLAP Queries, OLAP server Architecture ROLAP versus MOLAP Versus HOLAP.
7-8	**Data Mining**: Introduction, Challenges, Data Mining Tasks, Data: Types of Data, Data Quality, Data Pre-processing, Measures of Similarity and Dissimilarity.
9-10	**Association Analysis:** Problem Definition, Frequent Item set Generation, Rule generation. Alternative Methods for Generating Frequent Item sets, FP-Growth Algorithm, Evaluation of Association Patterns.
11-12	**Classification:** Decision Trees Induction, Method for Comparing Classifiers, Rule Based Classifiers, Nearest Neighbor Classifiers, Bayesian Classifiers.
13-14	**Clustering Analysis:** Overview, K-Means, Agglomerative Hierarchical Clustering, DBSCAN, Cluster Evaluation, Density-Based Clustering, Graph- Based Clustering, Scalable Clustering Algorithms.

Course Code: BSCHAI-403(P)
COURSE NAME: SOFTWARE LAB-VII

Maximum Marks: 50 Time: 3 hours
External Examination: 50 Marks Pass Marks: 35%
Practical Sessions to be conducted: 40-50 Hrs

Course Objectives:

- The main objective of this lab is to impart the knowledge on how to implement classical models and algorithms in data warehousing and data mining and to characterize the kinds of patterns that can be discovered by association rule mining, classification and clustering. At the end to compare and contrast different conceptions of data mining.

Course Learning Outcomes:

Upon Completion of the course, Students will be able to,

- To evaluate the different models of OLAP and data preprocessing.
- To enlist various algorithms used in information analysis of Data Mining Techniques.
- To demonstrate the knowledge retrieved through solving problems

It is expected that student should implement concept of Data Mining and Warehousing. The open source Data Mining Tools like Rapid Miner, Weka etc. can be used to implement the concept of Data Mining and Warehousing. Some examples are as follows (Subject Teacher may add more):

1. Implementation of OLAP operations
2. Implementation of Varying Arrays
3. Implementation of Nested Tables
4. Demonstration of any ETL tool
5. Write a program of Apriori algorithm using any programming language.
6. Create data-set in .arff file format.
7. Demonstration of preprocessing on WEKA data-set.
8. Demonstration of Association rule process on data-set contact lenses.arff /supermarket (or any other data set) using apriori algorithm.
9. Demonstration of classification rule process on WEKA data-set using j48 algorithm.
10. Demonstration of classification rule process on WEKA data-set using Naive Bayes algorithm.

The breakup of marks for the practical will be as under

I.	LabRecord (InternalAssessment)	10Marks
II.	VivaVoce (ExternalEvaluation)	20Marks
III.	Program Development and Execution (ExternalEvaluation)	20Marks

Course Code: BSCHAI-404
COURSE NAME: DATA ANALYSIS USING PYTHON

Maximum Marks: 100 Time: 3 hours
Theory: 75 Marks Pass Marks: 35%
Internal Assessment: 25 Marks

Course Objectives:

- Python programming is relatively quick to learn and has a great set of tools for importing, transforming, exploring, extracting insights from, making predictions with, and exporting the data. This course introduces the major Python tools used for preparing the data for analysis, the tools available for understanding the data, and using the data for insights and predictions.
- Students will learn to turn data into actionable information.

Course Learning Outcomes:

Upon Completion of the course, Students will be able to,

- Learn how to work with NumPy data types
- Know about pandas Series
- Know how to import and clean data
- Work through a complete data analysis to understand how the tools interact with each other.
- Understand how to use data visualization

(A) INSTRUCTIONS FOR THE PAPERSETTER

The question paper will consist of three sections UNIT-I, UNIT-II, and UNIT-III. Each of UNIT-I and UNIT-II will have four questions from the respective Units of the syllabus and each question will carry 12 marks. UNIT-III will have 9 short answer type questions which will cover the entire syllabus uniformly and will carry 3 marks in all.

(B) INSTRUCTIONS FOR THECANDIDATES

Candidates are required to attempt two questions each from UNIT-I and UNIT-II. UNIT-III is Compulsory.

Unit-I

Introduction to Data Analysis, Kinds of Data, Essential Python Libraries: NumPy, Pandas, Matplotlib, IPython, Jupyter, SciPy, Scikit-learn, Statsmodels

Python Language Basics

NumPy Basics: Arrays and Vectorized Computation, The NumPyndarray: A Multidimensional Array Object, Universal Functions: Fast Element-Wise Array Functions, Array-Oriented Programming with Arrays, Broadcasting, ndarray Object Internals

Getting Started With Pandas: Introduction to pandas Data Structures

Essential Functionality: Reindexing, Dropping Entries from an Axis, Indexing, Selection, Filtering, Integer Indexes, Arithmetic and Data Alignment, Function Application and Mapping, Sorting and Ranking, Axis Indexes with Duplicate Labels

Summarizing and Computing Descriptive Statistics: Data Loading, Storage, and File Formats, Reading and Writing Data in Text Format: Binary Data Formats

Unit-II

Data Cleaning and Preparation: Handling Missing Data, Data Transformation, String Manipulation, Data Wrangling, Join, Combine, and Reshape, Hierarchical Indexing.

Data Aggregation and Group Operations: GroupByMechanics, Data Aggregation

General split-apply-combine: Suppressing the Group Keys, Quantile and Bucket Analysis, Example: Filling Missing Values with Group-Specific Values, Example: Random Sampling and Permutation, Example: Group Weighted Average and Correlation.

Advanced pandas: Categorical Data, Advanced GroupBy Use, Techniques for Method Chaining

Introduction to statsmodels: Estimating Linear Models.

Introduction to scikit-learn

Plotting and Visualization: A Brief matplotlib API Primer, Plotting with pandas and seaborn.

Text Book:

1. Wes McKinney, Python for Data Analysis, Shroff Publications and Distributors

Reference Book:

1. Michael Milton, A Brain Friendly Guide: Head First Data Analysis, Shroff Publications and Distributors.

Teaching Plan

Week	Content
1-2	**Introduction to Data Analysis, Kinds of Data, Essential Python Libraries:** NumPy, Pandas, Matplotlib, IPython, Jupyter, SciPy, Scikit-learn, Statsmodels **Introduction to IPython, and Jupyter Notebooks:** The Python Interpreter **Python Language Basics**
3-4	**NumPy Basics:** Arrays and Vectorized Computation, The NumPyndarray: A Multidimensional Array Object, Universal Functions: Fast Element-Wise Array Functions,Array-Oriented Programming with Arrays. **Advanced NumPy:** ndarray Object Internals, Advanced Array Manipulation, Broadcasting, Advanced ufunc Usage, Structures and Record Arrays, More About Sorting
5-6	**Getting Started With Pandas:** Introduction to pandas Data Structures **Essential Functionality:** Reindexing, Dropping Entries from an Axis, Indexing, Selection, Filtering, Integer Indexes, Arithmetic and Data Alignment, Function Application and Mapping, Sorting and Ranking, Axis Indexes with Duplicate Labels
7-8	**Summarizing and Computing Descriptive Statistics:** Data Loading, Storage, and File Formats, Reading and Writing Data in Text Format: Binary Data Formats **Data Cleaning and Preparation:** Handling Missing Data, Data Transformation, String Manipulation, Data Wrangling, Join, Combine, and Reshape, Hierarchical Indexing, Combining and Merging Datasets, Reshaping and Pivoting
9-10	**Data Aggregation and Group Operations:** GroupByMechanics, Data Aggregation **General split-apply-combine:** Suppressing the Group Keys, Quantile and Bucket Analysis, Example: Filling Missing Values with Group-Specific Values, Example: Random Sampling and Permutation, Example: Group Weighted Average and Correlation
11-12	**Advanced pandas:** Categorical Data, Advanced GroupBy Use, Techniques for Method Chaining **Introduction to Modelling Libraries in Python:** Interfacing Between pandas and Model Code, Creating Model Descriptions with Patsy **Introduction to stats models:** Estimating Linear Models, Estimating Time Series Processes
13-14	**Introduction to scikit-learn** **Plotting and Visualization:** A Brief matplotlib API Primer, Plotting with pandas and seaborn

Course Code: BSCHAI-404(P)
COURSE NAME: SOFTWARE LAB-VIII

Maximum Marks: 50 Time: 3 hours
External Examination: 50 Marks Pass Marks: 35%
Practical Sessions to be conducted: 40-50 Hrs

Course Objectives:

- This Lab course introduces the major Python tools used for preparing the data for analysis, the tools available for understanding the data, and using the data for insights and predictions.

Course Learning Outcomes:

Upon Completion of the course, Students will be able to,

- Learn how to work with NumPydatatypes
- Be proficient in pandas Series
- Know how to import and clean data
- Work through a complete data analysis to understand how the tools interact with each other.
- Understand how to use data visualization

The breakup of marks for the practical will be as under

I.	Lab Record (Internal Assessment)	10Marks
II.	Viva Voce (External Evaluation)	20Marks
III.	Program Development and Execution (External Evaluation)	20Marks