# Walchand College of Engineering

*(Government Aided Autonomous Institute)*

Vishrambag, Sangli-416415


***** Platinum Jubilee Year *****

**Syllabus for**

**F. Y. M. Tech. (Data Science)**

**Semester-I and II**

**Effective from AY 2024-25**

# Semester-I

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

## Course Information

| | |
|---|---|
| **Programme** | M. Tech. All Branches |
| **Class, Semester** | First Year M. Tech., Sem I |
| **Course Code** | 7IC501 |
| **Course Name** | Research Methodology |
| **Desired Requisites:** | |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | --- | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

## Course Objectives

| | |
|---|---|
| 1 | To prepare students for undergoing research, identify and formulate the research problems, state the hypothesis, design a research layout, set a research process and methodology. |
| 2 | To enable students to interpret the results, propose theories, suggest possible/alternative solutions, solve, and prove the solution adapted–logically and analytically, conclude the research findings. |
| 3 | To impart knowledge to analyze critically the literature and publish research in reputed conferences/ journals. |
| 4 | To expose students to research ethics, IPR and Patents |

## Course Outcomes (CO) with Bloom's Taxonomy Level

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| CO1 | Demonstrate a research solution in each engineering domain using appropriate Engineering research process and research methodology. | II | Applying |
| CO2 | Device feasible solution to a research problem in the respective engineering domain based on economic, social and legal aspects using appropriate research procedures and practices. | III | Analyzing |
| CO3 | Compose research publications and dissertation reports efficiently. | VI | Creating |
| CO4 | Draft IPR and patent documents, as well as copyright documentation for research work. | VI | Creating |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Engineering Research Process:** Meaning of research problem, Sources of research problem, Criteria and Characteristics of a good research problem, Errors in selecting a research problem, Definition, scope and objectives of research problem. Approaches of investigation of solutions for research problem, data collection, analysis, interpretation. | 6 |
| II | **Research Methodology :** Problem statement formulation, resources identification for solution, Experimental and Analytical modeling, Simulations, Numerical and Statistical methods in engineering research. Hypothesis and its testing by different techniques: T-Test, Z-test etc., | 6 |
| III | **Research Methods:** Uni and Multivariate Analysis: ANOVA, Design of Experiments/Taguchi Method, | 7 |

| | | |
|---|---|---|
| | Regression Analysis. Software tools like spreadsheets.<br>Processing and Analysis of Data: Processing Operations, Types of Analysis-Presentation and Interpretation of Data Editing, Classification and Tabulation-Interpretation. Analyse your results and draw conclusions. | |
| IV | **Research Practices:**<br>Effective literature studies approaches, critical analysis, Plagiarism, Research ethics, Mendeley - Reference Management Software.<br>Research communication- Effective Technical Writing, Writing a research article for Journal/conference paper, Technical report, Dissertation/ Thesis report writing,<br>Software used for report writing such as word, Latex etc. Presentation techniques for paper/report/seminar. Publishing article in Scopus/SCI/Web of science indexed journal or conference. | 7 |
| V | **Intellectual Property Rights (IPR):**<br>Nature of Intellectual Property: Patents, Designs, Trade and Copyright, Ownership of copyright, Term of copyright, Technological research, innovation, patenting, development. International Scenario: International cooperation on Intellectual Property, New developments in IPR, Traditional knowledge, Various Case Studies. | 7 |
| VI | **Patents:**<br>Patent Rights: Scope of Patent Rights, Various Patent databases, Geographical Indications. Procedure for grants of patents, Patenting under Patent Cooperation Treaty (PCT). Licensing and transfer of technology. Administration of Patent System. Introduction to International Scenario: World Intellectual Property Organization (WIPO), Trade-Related Aspects of Intellectual Property Rights (TRIPs), Patenting under PCT. | 6 |

| | |
|---|---|
| | |
| **Textbooks** | |
| 1 | Kothari C. R, "Research Methodology", 5th Edition, New Age International, 2023 |
| 2 | Melville Stuart and Goddard Wayne, "Research Methodology: An Introduction for Science & Engineering Students" Juta and Company Ltd, 4th edition 2023. |
| 3 | Kumar Ranjit, "Research Methodology: A Step-by-Step Guide for beginners", SAGE Publications, , 4th edition 2023. |
| | |
| **References** | |
| 1 | Merges Robert, Menell Peter, Lemley Mark, "Intellectual Property in New Technological Age", ASPEN Publishers, 2018. |
| 2 | Ramappa T., "Intellectual Property Rights Under WTO", S. Chand, 2008 |
| 3 | Mayall, "Industrial Design", McGraw Hill, Oct 2021. |
| 4 | Halbert, "Resisting Intellectual Property", Taylor & Francis Ltd ,2020 |
| 5 | Deepak Chopra and Neena Sondhi, "Research Methodology : Concepts and cases ", Vikas Publishing House, New Delhi |
| | |
| **Useful Links** | |
| 1 | https://onlinecourses.nptel.ac.in/noc21_ge03/preview- Introduction to research |
| 2 | https://onlinecourses.swayam2.ac.in/ntr21_ed23/preview - Academic Research & Report Writing |
| 3 | https://onlinecourses.nptel.ac.in/noc21_ge12/preview - Qualitative Research Methods And Research Writing |
| 4 | https://onlinecourses.nptel.ac.in/noc21_hs44/preview - Effective Writing |
| 5 | https://www.scopus.com/search/form.uri?display=basic#basic |
| 6 | https://webofscienceacademy.clarivate.com/learn |
| 7 | https://www.wipo.int/about-wipo/en/ |
| 8 | https://iprsearch.ipindia.gov.in/publicsearch |

| CO-PO Mapping | | | | | | |
|---|---|---|---|---|---|---|
| | Programme Outcomes (PO) | | | | | |
| | **1** | **2** | 3 | 4 | 5 | 6 |
| **CO1** | 2 | 2 | 1 | | | |
| **CO2** | 3 | 2 | 2 | 3 | 2 | 2 |
| **CO3** | | 3 | | 3 | 1 | |
| **CO4** | | | | 3 | 1 | |
| The strength of mapping is to be written as 1: Low, 2: Medium, 3: High Each CO of the course must map to at least one PO. | | | | | | |

## Assessment

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.
ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.
For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem I |
| **Course Code** | 7DS501 |
| **Course Name** | Mathematics for Data Science |
| **Desired Requisites:** | Basics of Mathematics |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To understand data pre-processing and handling techniques. |
| **2** | To introduce various Statistics and Probability concepts to be used in data science. |
| **3** | To understand and apply Linear Algebra concepts for data modelling. |
| **4** | To understand and use matrices to be used in data science for representation and applications. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Demonstrate understanding of basic mathematical concepts in data science, relating to linear algebra, probability, and calculus. | II | Understanding |
| **CO2** | Employ methods related to these concepts in a variety of data science applications. | III | Applying |
| **CO3** | Apply logical thinking to problem-solving in context. | III | Applying |
| **CO4** | Use appropriate technology to aid problem-solving and data analysis. | III | Applying |
| **CO5** | Analyze data using data pre-processing approaches | IV | Analyzing |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Data Pre-processing:** Data Types, Data collections, Data Cleaning, Data Integration, Data Reduction, Data Discretization, Data Normalization, Data Transformation. | 7 |
| II | **Statistics:** Five-point summary, Boxplot Analysis, Sampling techniques, Basic statistics, sampling distributions, mixture models. | 6 |
| III | **Probability:** Basic probability, conditional probability, total probability, independent events, Bayes' theorem, random variable, Moments, moment functions, distributions, Joint distribution, conditional distribution, transformation of random variables, correlation. | 7 |
| IV | **Basics of Linear Algebra:** Representation of vectors, Linear dependence and independence, vector space and linear transformations, range and null space. | 6 |

| | | |
|---|---|---|
| V | **Matrices-I**<br>Projection transformation, orthogonal decomposition, singular value decomposition, principal component analysis and linear discriminant analysis | 6 |
| VI | **Matrices-II:**<br>Matrices with linear transformations, special matrices, eigenvalues and eigenvectors with applications to data problems, Least square and minimum normed solutions | 7 |

| | |
|---|---|
| | |
| **Textbooks** | |
| 1 | Thomas Nield, "Essential Math for Data Science," O'Reilly Media, 2022 |
| 2 | Dirk P. Kroese, Zdravko Botev, Thomas Taimre, Radislav Vaisman, "Data Science and Machine Learning Mathematical and Statistical Methods", CRC Press, 2019 |
| 3 | Sinan Ozdemir, "Principles of Data Science", Packt Publishing, 2016 |
| | |
| **References** | |
| 1 | E. Kreyszig, -"Advanced Engineering Mathematics", John Wiley and Sons, Inc., U.K. (10th Edition) 2015 |
| 2 | M. P. Deisenroth, A. A. Faisal, C. S. Ong; "Mathematics for Machine Learning", Cambridge University Press (1st edition) 2020 |
| 3 | R. A. Johnson, I. Miller, and J. E.Freund, "Miller & Freund's Probability and Statistics for Engineers", Prentice Hall PTR, (8th edition) 2011 |
| 4 | Jeff M. Phillips, "Mathematical Foundations for Data Analysis", Springer International Publishing, 2021 |
| | |
| **Useful Links** | |
| 1 | https://www.coursera.org/specializations/mathematics-for-machine-learning-and-data-science |
| 2 | https://www.codecademy.com/learn/paths/fundamental-math-for-data-science |

**CO-PO Mapping**

| | **Programme Outcomes (PO)** | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | | 3 | | | | |
| **CO2** | 2 | | | | | |
| **CO3** | | | 3 | | | |
| **CO4** | | 2 | 1 | | | 1 |
| **CO5** | | 2 | 2 | | | |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.
ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.
For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

| **Course Information** | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem I |
| **Course Code** | 7DS502 |
| **Course Name** | Data Structures and Algorithms |
| **Desired Requisites:** | C programming |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

| **Course Objectives** | |
|---|---|
| **1** | Exploring basics of data structures and algorithms. |
| **2** | Introduces a variety of data structures such as hash tables, search trees, tries, heaps, graphs |
| **3** | Familiarize sorting and pattern matching algorithms |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Apply variety of data structures that efficiently model problems | III | Applying |
| **CO2** | Illustrate linear and non-linear data structures use in algorithm | III | Applying |
| **CO3** | Study and compare various algorithm techniques | IV | Analyzing |
| **CO4** | Discuss the algorithms used for sorting and pattern matching | V | Evaluating |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Introduction:** Introduction to data structures, Introduction to algorithms, Complexity Analysis, Time and Space complexity of algorithms, asymptotic analysis, big O and other notations. | 6 |
| II | **Linear Structures:** Various structures such as: Linear Lists, Stacks and Queues ,Abstract data types, sequential and linked implementations, equivalence problem, linked lists, doubly linked lists, circular lists | 6 |
| III | **Non-Linear Structures:** Basic terminology, binary trees and its representation, binary tree traversals, operations, expression trees, Binary Search Trees, Heap tree. Adjacency matrix and adjacency list representation of graph, Traversals | 7 |
| IV | **Algorithmic Techniques:** Algorithm design strategies, divide and conquer and performance analysis, Greedy method and its applications; Dynamic programming and its performance analysis, Back-tracking, n-queens problem | 7 |

| | | |
|---|---|---|
| V | **Graph Algorithms:**<br>DFS and BFS, spanning trees, bi-connectivity; Minimum cost spanning trees: Kruskal's, Prim's and Sollin's algorithms; Path finding and shortest path algorithms; Topological sorting; Bipartite graphs. P and NP-classes, NP-hard problems, reduction. | 7 |
| VI | **Pattern Matching and Tries:**<br>Pattern matching algorithms-Brute force, the Boyer –Moore algorithm, the Knuth-Morris-Pratt algorithm, Standard Tries, Compressed Tries, Suffix tries. | 6 |

| **Textbooks** | |
|---|---|
| 1 | Richard F. Gilberg, Behrouz A. Forouzan, "Data Structures, A Pseudocode Approach With C", Cengage Learning, 2nd Edition, 2007 |
| 2 | Cormen T, "Introduction to Algorithms", MIT Press,4th Edition, 2022 |

| **References** | |
|---|---|
| 1 | Brad Miller and David Ranum, Luther College, "Problem Solving with Algorithms and Data Structures Using Python," 2nd edition Franklin, Beedle &Associates, 2017 |
| 2 | Wirth, N., "Algorithms and Data Structures", 4th edition , Prentice-Hall of India, 2013 |

| **Useful Links** | |
|---|---|
| 1 | https://nptel.ac.in/courses/106/102/106102064/ |
| 2 | https://nptel.ac.in/courses/106/106/106106127/ |
| 3 | https://nptel.ac.in/courses/106/103/106103069/ |

**CO-PO Mapping**

| | **Programme Outcomes (PO)** | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | 1 | | | 3 | | |
| **CO2** | | | 3 | | | |
| **CO3** | | | 2 | 2 | | 2 |
| **CO4** | 1 | | 3 | | | |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.
ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.
For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

**Walchand College of Engineering, Sangli**
*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem I |
| **Course Code** | 7DS503 |
| **Course Name** | Principles of Database Systems |
| **Desired Requisites:** | |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 2 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | - | **Credits: 2** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To introduce principles of database management systems |
| **2** | To impart conceptual designs for databases |
| **3** | To describe issues associated with transaction management |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Illustrate the relational database management systems and use of SQL and PL/SQL commands to create and manipulate database objects | II/III | Understanding /Applying |
| **CO2** | Distinguish the basic concepts of relational data model, schemas and instances using various techniques. | IV | Analyzing |
| **CO3** | Classify concurrency control protocols and database recovery methods. | V | Evaluating |
| **CO4** | Design relational database system for concurrency control using several protocols | VI | Creating |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Introduction:** Database applications, purpose, accessing and modifying databases, need for transactions, architecture - users and administrators, data mining, information retrieval. Relational Databases: relational model, database schema, keys, relational query languages, algebra, tuple | 5 |
| II | **Structured Query Language (SQL):** Data definition, basic SQL query structure, set operations, nested sub queries, aggregation, null values, database modification, joins expressions, views. **Database Design:** E-R model, E-R diagram, reduction to relational schema, E-R design issues, database integrity, specifying integrity constraints in SQL: unique columns, foreign key, triggers. | 5 |
| III | **Relational Database Design:** Features of design, Functional Dependency theory, decomposition using functional dependency and normal forms, algorithms for decomposition, normal forms, (optional: multi-valued dependency and 4th normal form). | 4 |

| | | |
|---|---|---|
| IV | **Query Optimization:**<br>Transformation of relational expressions, estimating cost and statistics of expression, choosing evaluation plans, linear and bushy plans, dynamic programming algorithms. | 4 |
| V | **Transactions:**<br>Properties and states, Concurrent execution, Serializability.<br>**Concurrency Control:** Lock-Based Protocols, 2 phase locking protocol, Graph based protocols, Timestamp based protocols, Deadlock handling | 5 |
| VI | **Crash Recovery:**<br>Recovery: Failures and their classification, recovery and atomicity, recovery algorithms, Undo-Redo with write ahead logging, no Undo no Redo and other combinations, buffer management | 4 |

| **Text Books** | |
|---|---|
| 1 | Abraham Silberschatz, Henry F. Korth, and S. Sudarshan, "Database System Concepts", McGraw-Hill Education, 7th Edition, 2019. |
| 2 | Raghu Ramakrishnan, "Database Management Systems", McGraw-Hill Education, 3rd Edition, 2003. |

| **References** | |
|---|---|
| 1 | J.D. Ullman, "Principles of Database Systems", Galgotia Publications, 2nd Edition, 1999 |
| 2 | H Garcia-Molina, JD Ullman and Widom," Database Systems: The Complete Book" ,2nd Ed., Prentice-Hall, 2008. |
| 3 | C.J.Date, A.Kannan, S.Swamynathan, "An Introduction to Database Systems", Pearson Education, 8th Edition, 2006. |

| **Useful Links** | |
|---|---|
| 1 | https://nptel.ac.in/courses/106/105/106105175/ |
| 2 | http://www.nptelvideos.in/2012/11/database-management-system.html |
| 3 | https://www.tutorialspoint.com/mongodb/mongodb_overview.htm |
| 4 | https://www.tutorialspoint.com/mariadb/mariadb_introduction.htm |

**CO-PO Mapping**

| | **Programme Outcomes (PO)** | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | 2 | | 3 | | | |
| **CO2** | | | 2 | | | |
| **CO3** | | | | 2 | | 1 |
| **CO4** | 2 | | 3 | 1 | | |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.
ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.
For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem I |
| **Course Code** | 7DS551 |
| **Course Name** | Data Structures and Algorithm Lab |
| **Desired Requisites:** | C Programming |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Practical** | 2 Hrs/ Week | **LA1** | **LA2** | **Lab ESE** | **Total** |
| **Interaction** | - | 30 | 30 | 40 | 100 |
| | | **Credits: 1** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To summarize concepts of data structures and algorithms |
| **2** | To apply linear and non-linear data structures for problem solving |
| **3** | To analyze performance of various algorithms |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Implement various data structures and algorithms | III | Applying |
| **CO2** | Demonstrate various operations on linear and non-linear data structures | III | Applying |
| **CO3** | Apply different algorithmic technique to solve engineering problem | III | Applying |
| **CO4** | Identify and implement pattern matching algorithms for data science | IV/V | Analyzing/ Evaluating |

**List of Experiments / Lab Activities/Topics**

**List of Lab Activities:**

1. Developing ADT for singly, circular linked list and its applications
2. Developing ADT for stack and queue and their applications
3. Implementation of recursive and non-recursive tree traversals
4. Binary search tree and application
5. Implementation of graph, DFS, BFS
6. Sorting Methods: Insertion sort, shell sort, heap sort, quick sort, merge sort, radix sort etc.
7. Write a program to find solution for knapsack problem using greedy method.
8. Write a program to find minimum cost spanning tree using Prim's Algorithm.
9. Write a program to find minimum cost spanning tree using Kruskal's Algorithm.
10. Write a program to perform Single source shortest path problem for a given graph.
11. Write a program to find solution for job sequencing with deadlines problem.
12. Write a program for all pairs shortest path problem.

| **Textbooks** | |
|---|---|
| 1 | Michael T. Goodrich, Roberto Tamassia , Michael H. Goldwasser , "Data Structures and Algorithms in Python" Wiley Publications, 2nd Edition, 2013 |
| 2 | Cormen T, "Introduction to Algorithms", MIT Press,4th Edition, 2022 |

| **References** | |
|---|---|
| 1 | Yashavant Kanetkar, "Understanding pointers in C", 19th edition, BPB Publication,2022 |
| 2 | Brian W. Kernighan and Dennis M. Ritchie, "The C Programming Language", 2ndEdition, Prentice Hall of India,2015 |

| **Useful Links** | |
|---|---|
| 1 | https://nptel.ac.in/courses/106/102/106102064/ |
| 2 | https://nptel.ac.in/courses/106/106/106106127/ |
| 3 | https://nptel.ac.in/courses/106/103/106103069/ |

| **CO-PO Mapping** | | | | | | |
|---|---|---|---|---|---|---|
| | **Programme Outcomes (PO)** | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | | 2 | | | | |
| **CO2** | | | 3 | 2 | | |
| **CO3** | | 2 | 3 | | | 2 |
| **CO4** | | 3 | | 2 | | |

The strength of mapping is to be written as 1,2,3; where, 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO, and preferably to only one PO.

| Assessment | | | | |
|---|---|---|---|---|
| There are three components of lab assessment, LA1, LA2 and Lab ESE.<br>IMP: Lab ESE is a separate head of passing.(min 40 %), LA1+LA2 should be min 40% | | | | |
| **Assessment** | **Based on** | **Conducted by** | **Typical Schedule** | **Marks** |
| LA1 | Lab activities, attendance, journal | Lab Course Faculty | During Week 1 to Week 4<br>Marks Submission at the end of Week 8 | 30 |
| LA2 | Lab activities, attendance, journal | Lab Course Faculty | During Week 5 to Week 9<br>Marks Submission at the end of Week 9 | 30 |
| Lab ESE | Lab activities, journal/ performance | Lab Course Faculty and External Examiner as applicable | During Week 9 to Week 13<br>Marks Submission at the end of Week 13 | 40 |
| Week 1 indicates starting week of a semester. Lab activities/Lab performance shall include performing experiments, mini-project, presentations, drawings, programming, and other suitable activities, as per the nature and requirement of the lab course. The experimental lab shall have typically 8-10 experiments and related activities if any. | | | | |

**Walchand College of Engineering, Sangli**
*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem I |
| **Course Code** | 7DS552 |
| **Course Name** | Python Programming lab |
| **Desired Requisites:** | Computer Programming |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Practical** | 2 Hrs/Week | **LA1** | **LA2** | **Lab ESE** | **Total** |
| **Interaction** | - | 30 | 30 | 40 | 100 |
| | | **Credits: 1** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To understand why Python is a useful scripting language for developers. |
| **2** | To learn how to design and program Python applications. |
| **3** | To make use of the different libraries of Python. |
| **4** | To implement python code and add visualization using various libraries. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Implement various data structures available in Python programming language and apply them to solve computational problems. | III | Applying |
| **CO2** | Examine and analyze the programming models and make use of the different libraries of Python | IV | Analyzing |
| **CO3** | Build, test and debug the code written in Python. | VI | Creating |
| **CO4** | Produce various kinds of plots using various libraries. | VI | Creating |

| **Module** | **Contents** | |
|---|---|---|
| **I** | **Introduction to Python:** Variables and Data Types, Control Flow: Using conditional statements (if, else, elif) and loops (for, while) to control the execution flow of a program. **Functions:** Defining and calling functions, understanding scope (local and global variables) **Modules and Packages**: Importing and using standard libraries and creating custom modules. Files, System Functions and Parameters, Strings, Tuples. **Data Structures** -Lists and Dictionaries, Lists and Mutability, Functions as Objects. Programming using functions, modules and external packages | |
| **II** | **File handling:** Python File Operations: Reading files, Writing files in python, Programming, using file operations. **Database Programming**: Connecting to a database, Creating Tables, INSERT, UPDATE, DELETE and READ operations, Transaction Control, | |

| | | |
|---|---|---|
| | Disconnecting from a database, and Exception Handling in Databases.<br>**Array handling with Numpy and Pandas:** NumPy: Introduction, Numpy array, Numpy array indexing, Numpy operations. Pandas: Series, Data frames, managing missing data, groupby, merging & concatenation, operations, data input and data output. Introduction to NumPy and Pandas for data manipulation and analysis. | |
| **III** | **Object-Oriented Programming:**<br>Abstract Data Types and Classes, Information Hiding, Class in Python Objects in Python, Polymorphism in Python, Encapsulation in Python<br>Inheritance in Python, Data Abstraction in Python. **Exception Handling**:- Understanding exceptions, Handling exceptions using try, except, finally | |
| **IV** | **Hypothesis testing using python:**<br>Hypothesis testing-Two sample testing , T test, F-test ,One way and Two way ANOVA<br>**Case Studies**: using California Housing Dataset or Iris data set | |
| **V** | **Machine learning using python:**<br>Classification, linear regression. Multiple regression, Concepts of MLE and Logistic regression, ROC and Regression Analysis Model Building, c2 Test.<br>**Case Studies**: Time Series,Simple Linear Regression and Multiple Linear Regression with the California Housing Dataset/Iris data set | |
| **VI** | **Python for Data Visualization:**<br>Working with Graphs: Creating various types of plots (line, bar, scatter, histogram) and customizing them Understanding and implementing graph algorithms, visualizing graphs using libraries -Matplotlib, Seaborn, Plotly and Cufflinks, Geographical Plotting. | |

**List of Experiments / Lab Activities/Topics**

**List of Lab Activities based on the above contents:**

1. Problem solving using core Python functionality like strings, variables, functions.
2. Problem solving using core Python functionality like tuples, dictionary, list, objects
3. Problem solving using Class & object concepts.
4. Problem statement on inheritance in classes
5. Problem based on encapsulation in classes
6. Problem statement on array
7. Problem statement on NumPy libraries with different operations
8. Problem statement on Pandas libraries with different operations
9. Problem statement on NumPy and Pandas use for data manipulation and analysis.
10. Hypothesis testing using all standard test methods using python.
11. Problem statement on data visualization using Matplot Libraries.
12. Problem statement on data visualization using Seaborn Libraries.

**Best Practices for lab:**

- Writing clean and readable code
- Testing and debugging
- Documentation and comments
- Version control with Git

| Textbooks | |
|---|---|
| 1 | R. Nageswara Rao, “Core Python Programming‖”, Dreamtech Press, 3rd Edition, 2021 |
| 2 | Chun, J Wesley, “Core Python Programming‖” Pearson, 2nd Edition, 2007 Reprint 2021 |
| 3 | Eric Matthes - "Python Crash Course", "Automate the Boring Stuff with Python" 2nd Edition,2019 |
| 4 | Paul J. Deitel, Harvey Deitel”Python for Programmers” fourth Edition By Pearson, 2022 |
| | |
| **References** | |
| 1 | Barry, Paul, Head *First Python*, O Rielly,2nd Edition, 2010 |
| 2 | Lutz, Mark, *Learning Python*, O Rielly, 4th Edition, 2009 |
| 3 | B. Uma Maheswari, R. Sujatha - Introduction to Data Science: Practical Approach with R and Python - wiley – 1st edition, October 2021 |
| | |
| **Useful Links** | |
| 1 | https://onlinecourses.nptel.ac.in/noc19_mg47/preview |
| 2 | https://docs.python.org/3/tutorial/ |
| 3 | https://www.learnpython.org/ |
| 4 | https://www.hackerrank.com/ |

| CO-PO Mapping | | | | | | |
|---|---|---|---|---|---|---|
| | Programme Outcomes (PO) | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | 2 | 2 | | | | |
| **CO2** | | 2 | 3 | | | 2 |
| **CO3** | | | 3 | | 1 | |
| **CO4** | 2 | | 3 | | | |
| The strength of mapping is to be written as 1: Low, 2: Medium, 3: High<br>Each CO of the course must map to at least one PO. | | | | | | |

**Assessment**

There are three components of lab assessment, LA1, LA2 and Lab ESE.
IMP: Lab ESE is a separate head of passing.(min 40 %), LA1+LA2 should be min 40%

| Assessment | Based on | Conducted by | Typical Schedule | Marks |
|---|---|---|---|---|
| LA1 | Lab activities, attendance, journal | Lab Course Faculty | During Week 1 to Week 4<br>Marks Submission at the end of Week 4 | 30 |
| LA2 | Lab activities, attendance, journal | Lab Course Faculty | During Week 5 to Week 9<br>Marks Submission at the end of Week 9 | 30 |
| Lab ESE | Lab activities, journal/ performance | Lab Course Faculty and External Examiner as applicable | During Week 9 to Week 13<br>Marks Submission at the end of Week 13 | 40 |

Week 1 indicates starting week of a semester. Lab activities/Lab performance shall include performing experiments, mini-project, presentations, drawings, programming, and other suitable activities, as per the nature and requirement of the lab course. The experimental lab shall have typically 8-12 experiments and related activities if any.

**Walchand College of Engineering, Sangli**
*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem I |
| **Course Code** | 7DS553 |
| **Course Name** | Logical Programming for Data Science |
| **Desired Requisites:** | Computer Programming |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Practical** | 2 Hrs/Week | **LA1** | **LA2** | **Lab ESE** | **Total** |
| **Interaction** | 1Hrs | 30 | 30 | 40 | 100 |
| | | **Credits: 2** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | Learn the foundational concepts and installation processes for Prolog programming. |
| **2** | Understand and implement logical operators, rules, clauses, and list operations in Prolog. |
| **3** | Develop ANN using Python |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Demonstrate the ability to perform various list operations and implement logical operators in Prolog | III/IV | Applying/ Analyzing |
| **CO2** | Install Prolog and create programs utilizing facts, rules, and clauses | III/VI | Applying/ creating |
| **CO3** | Recommend AI techniques to solve real life problems | V | Evaluating |
| **CO4** | Develop neural network models and use it effectively | VI | Creating |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Introduction :** Introduction to Prolog Programming, Facts, Rules, Clauses, and Lists in Prolog | **2** |
| II | **Recursion in Prolog:** Unification, Backtracking ,Logical Operators in Prolog, Recursion in Prolog | **2** |
| III | **List Processing in Prolog:** Prolog Program for Various Relations, List Operations in Prolog, Union and Intersection. | **2** |
| IV | **Problem solving in AI:** Crypt arithmetic, Monkey Banana problem solving using Prolog, Water Jug problem. | **2** |

| | | |
|---|---|---|
| V | **Heuristic searching in AI:**<br>Concept of Heuristic Search in AI: Informed searching Techniques , A* , Best first searching algorithm. | **3** |
| VI | **ANN using Python:**<br>ANN, Perceptron learning, Multi-layer Feed forward network, Hopfield model for pattern storage task. | **2** |

**List of Experiments / Lab Activities/Topics**

**List of Lab Activities:**

1. Write a simple fact and rules program e.g. Family Relation.
2. Write predicates one converts centigrade temperatures to Fahrenheit, the other checks if a temperature is below freezing.
3. Write a program in PROLOG to implement factorial (N, F) where F represents the factorial of a number N..
4. Write a program to solve water jug problems using Prolog.
5. Write a program to solve the Monkey Banana problem.
6. Write a Prolog program to implement conc (L1, L2, L3) where L2 is the list to be appended with L1 to get the resulted list L3
7. Write a Prolog program to implement reverse (L, R) where List L is original and List R is reversed list.
8. Write a program in PROLOG to implement palindrome (L) which checks whether a list L is a palindrome or not.
9. Heuristic searching Techniques.

   ---------- with support of Virtual Lab —----------
10. Perceptron Learning
11. Multilayer feed forward Neural Network.

**Textbooks**


| | |
|---|---|
| 1 | Stuart Russell and Peter Norvig, ,"Artificial Intelligence: A Modern Approach", 4 th Edition Pearson Education, 2022. |
| 2 | Ivan Bratko., "Prolog Programming for Artificial Intelligence, Addison-Wesley",4th edition 2011. |


**References**


| | |
|---|---|
| 1 | B. Uma Maheswari, R. Sujatha – "Introduction to Data Science: Practical Approach with R and Python "-1 st edition ,Wiley - October 2021 |
| 2 | https://www.tutorialspoint.com/artificial_intelligence_with_python/artificial_intelligence_with_python_natural_language_processing.htm |

| Useful Links | |
|---|---|
| 1 | https://onlinecourses.swayam2.ac.in/nou23_cs14/preview |
| 2 | https://www.javatpoint.com/prolog |
| 3 | https://swish.swi-prolog.org/ |

**CO-PO Mapping**

| | Programme Outcomes (PO) | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | 1 | 3 | | | | |
| **CO2** | 1 | 3 | | | | |
| **CO3** | | | 3 | | | |
| **CO4** | | 1 | 3 | | | 2 |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

There are three components of lab assessment, LA1, LA2 and Lab ESE.
IMP: Lab ESE is a separate head of passing.(min 40 %), LA1+LA2 should be min 40%

| Assessment | Based on | Conducted by | Typical Schedule | Marks |
|---|---|---|---|---|
| LA1 | Lab activities, attendance, journal | Lab Course Faculty | During Week 1 to Week 4 Marks Submission at the end of Week 4 | 30 |
| LA2 | Lab activities, attendance, journal | Lab Course Faculty | During Week 5 to Week 9 Marks Submission at the end of Week 9 | 30 |
| Lab ESE | Lab activities, journal/ performance | Lab Course Faculty and External Examiner as applicable | During Week 9 to Week 13 Marks Submission at the end of Week 13 | 40 |

Week 1 indicates starting week of a semester. Lab activities/Lab performance shall include performing experiments, mini-project, presentations, drawings, programming, and other suitable activities, as per the nature and requirement of the lab course. The experimental lab shall have typically 8-12 experiments and related activities if any.

# Semester-II

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M. Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem II |
| **Course Code** | 7DS521 |
| **Course Name** | Data Mining and Warehousing |
| **Desired Requisites:** | Statistics and Probability |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To educate students to the various concepts, algorithms and techniques in data mining |
| **2** | To understand data warehousing concepts and applications. |
| **3** | To apply data mining methods for complex data types and new application areas. |
| **4** | To apply data mining concepts in real word applications. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Understand data mining concepts, methods, and applications. | II | Understanding |
| **CO2** | Choose appropriate data pre-processing tasks such as data cleaning, normalization, transformation, feature selection, and dimensionality reduction. | III | Applying |
| **CO3** | Use various data mining techniques, including classification, clustering, association rule mining, and anomaly detection. | III | Applying |
| **CO4** | Identify real-world applications of data mining in various domains. | IV | Analyzing |
| **CO5** | Estimate the performance of different data mining models using appropriate metrics. | V | Evaluate |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Introduction to Data Mining:** Motivation and significance of data mining, data mining functionalities, interestingness measures, classification of data mining system, major issues in data mining, Data pre-processing, concept hierarchy generalization. | 6 |
| II | **Data warehouse and Online analytical processing (OLAP) technology:** Data warehouse, data warehouse architecture, data warehouse implementation, data warehouse basics, Schemas, schema models, multidimensional data models, OLAP types, on-line analytical processing. | 6 |
| III | **Mining frequent patterns:** Basic concepts, frequent item set mining algorithms, Mining various kinds of association rules, multilevel and multidimensional association rules, correlations, association rule mining versus correlation analysis, constraint based association mining. | 7 |
| IV | **Classification and prediction:** Definition, decision tree induction, Bayesian classification, rule based classification, classification by back propagation and support vector machines, associative classification, and prediction, accuracy and error measures. | 7 |

| | | |
|---|---|---|
| V | **Cluster Analysis**:<br>Definition, Clustering Algorithms - partitioning, hierarchical, density based, grid based and model based, Clustering high dimensional data, constraint based cluster analysis, density based and distance based outliers. | 7 |
| VI | **Data mining on complex data and applications:**<br>Algorithms for mining of spatial data, multimedia data, text data, Outlier Analysis, data mining applications, social impacts of data mining, trends in data mining. | 6 |

| **Textbooks** | |
|---|---|
| 1 | Han Jiawei and Kamber Micheline "Data Mining – Concepts and Techniques" The Morgan Kaufmann Series in Data Management Systems, 3rd Edition, Elsevier, 2012. |
| 2 | Dunham M. H, "Data Mining: Introductory and Advanced topics", Pearson, 2nd Edition, 2006 |
| 3 | Chattamvelli Rajan, "Data Mining Methods: Concepts & Applications", Narosa Publishing House, 2nd Edition, 2010 |
| 4 | Mitra Sushmita, Acharya Tinku, "Data Mining Multimedia, Soft Computing and Biometrics", WILEY Publication, 3rd Edition, 2016 |
| **References** | |
| 1 | Marakas, George M. "Modern data warehousing, mining, and visualization: core concepts. " Prentice Hall, 2003. |
| 2 | Pujari, Arun K. Data mining techniques. Universities press, 2001. |
| 3 | Gupta, Gopal K. Introduction to data mining with case studies. PHI Learning Pvt. Ltd., 2014. |
| **Useful Links** | |
| 1 | https://www.kdnuggets.com/websites/index.html |
| 2 | https://feedly.com/i/top/data-mining-blogs |
| 3 | https://onlinecourses.nptel.ac.in/noc21_cs06/preview |

**CO-PO Mapping**

| | **Programme Outcomes (PO)** | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | | 3 | | | | |
| **CO2** | 1 | 1 | 2 | | | |
| **CO3** | | 2 | 3 | | | 1 |
| **CO4** | 1 | | 2 | | 2 | 3 |
| **CO5** | | 1 | | 1 | | 1 |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.
ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.
For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

**Walchand College of Engineering, Sangli**
*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech, Sem II |
| **Course Code** | 7DS522 |
| **Course Name** | Data Handling and Visualization |
| **Desired Requisites:** | Programming Fundamentals |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 2 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | - | **Credits: 2** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To use R for analytical programming. |
| **2** | To visualize data in R. |
| **3** | To discuss problem solving approaches using appropriate machine learning techniques. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Describe critical R programming concepts in detail. | II | Understanding |
| **CO2** | Analyze data and generate reports based on the data. | IV | Analyzing |
| **CO3** | Construct bar charts, histograms, pie charts, scatter plots, line graphs,box plots, and maps using R and related packages. | V | Creating |
| **CO4** | Produce high-quality reports and presentations using LaTeX | V | Creating |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Introduction:**<br>Introduction to Data Science, Overview of the Data Science process, Introduction to Data Science technologies, Introduction to Machine Learning, Regressions, Classification, Clustering, Recommendation systems | 5 |
| II | **Working with Data :**<br>Variables , Vectors, Matrices, lists & Data frames , Logical vectored operators Image data type, Image representation, categorical data using Factors in R. | 4 |
| III | **Visualization of data using R:**<br>Using graphs to visualize data, Basic plotting in R, Manipulating the plotting window, Advanced plotting using lattice library in R. Image visualization in using Image processing | 4 |

| | | |
|---|---|---|
| | tools. | |
| IV | **Models in Machine Learning:**<br>Regression Models, Classification Models, Unsupervised Learning Models, Recommendation Models. Models considered: – Linear regression: lm() – logistic regression: glm() – Poisson regression: glm() – Survival analysis: Surv(), coxph() – Linear mixed models: lme() | 5 |
| V | **Introduction to LaTeX and Document Structure**<br>Overview of LaTeX, Basic Document, Structure, Environments and Lists, Including Graphics and Tables. Formatting contents, | 4 |
| VI | **Advanced LaTeX Features**<br>Mathematical Typesetting, Beamer, Presentations., Introduction to creating slides, adding frames, dividing the slide into multiple columns, adding different blocks, etc. | 4 |

| **Text Books** | |
|---|---|
| 1 | Dr. Mark Gardner, "Beginning R:statistical Programming Languages", Wrox (Amazon), 2nd edition Mar 2013 |
| 2 | Griffithas, Higham," Learning LATEX", Society for Industrial and Applied Mathematics, 2nd Edition, 2016. |

| **References** | |
|---|---|
| 1 | Basic Data Analysis Tutorial, by Jacob Whitehill, Department of Computer Science, University of the Western Cape, [UWCDataAnalysisTutorial.pdf] |
| 2 | NPTEL,edx,COURSERA (MOOC courses) |

| **Useful Links** | |
|---|---|
| 1 | https://www.coursera.org/learn/what-is-datascience?specialization=introduction- data science |
| 2 | https://onlinecourses.nptel.ac.in/noc21_cs23/preview |
| 3 | https://www.overleaf.com/learn/latex/Free_online_introduction_to_LaTeX_(part_1) |

| **CO-PO Mapping** | | | | | | |
|---|---|---|---|---|---|---|
| | **Programme Outcomes (PO)** | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | | 2 | | | | |
| **CO2** | 1 | 1 | 3 | | | |
| **CO3** | | 2 | 2 | | | 2 |
| **CO4** | | | 2 | 2 | | |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.
ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.For passing a theory course, Min. 40% marks in MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem II |
| **Course Code** | 7DS523 |
| **Course Name** | Multidimensional Data Analysis |
| **Desired Requisites:** | Basics of Discrete data and Statistics |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To discuss commonly used data and its representations for various applications |
| **2** | To explain implantation strategies of data science algorithms |
| **3** | To describe fundamentals of data analytics |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Define data handling strategies and its representation by identifying characteristics of data | I | Remembering |
| **CO2** | Convert data into appropriate forms by following transformation procedures | II | Understanding |
| **CO3** | Apply dimensionality reduction on datasets by practicing relevant algorithms | III | Applying |
| **CO6** | Separate data patterns into various classes by integrating learning techniques | IV | Analyzing |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Introduction:** Overview of multimedia, multidimensional and multimodal data Types of multidimensional data- spatial, temporal, spatiotemporal Data structures and formats- matrices, tensors, data cubes ,Requirement of multidimensional data analysis | 6 |
| II | **Data Pre-processing and Transformation:** Categorical and continuous data, Data sampling and aliasing, Handling missing values, outliers, and noise in multidimensional datasets, Data normalization and standardization. | 6 |
| III | **Dimensionality Reduction:** Concept of features, Feature selection and extraction ,Dimensionality reduction techniques- Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and t-Distributed Stochastic Neighbours Embedding (t-SNE) | 7 |
| IV | **Statistical Methods for Multidimensional Data:** Multivariate descriptive statistics, Hypothesis testing in multidimensional settings, Multivariate analysis of variance (MANOVA), Canonical correlation analysis (CCA), Multidimensional scaling (MDS) | 7 |

| | | |
|---|---|---|
| V | **Basics of Machine Learning for Multidimensional Data:**<br>Machine learning in data analytics, Basic techniques and algorithms in Machine Learning ,Classification and Clustering Performance evaluation metrics for assessing Machine Learning models | 7 |
| VI | **Applications and Case Studies:**<br>Applications of multidimensional data analysis in finance, healthcare, social sciences etc. Case Studies based on-Spatial data analysis, Temporal data analysis, Spatiotemporal data analysis, Discussions on complexity and related issues | 6 |

| **Textbooks** | |
|---|---|
| 1 | Peter Smith, Emma Johnson, "Multidimensional Data Analytics: Concepts, Techniques, and Applications", Springer, 2023 |
| 2 | Robert Wilson, Alice Davis, "Multidimensional Data Analytics: Techniques and Tools", CRC Press, 2020 |
| **References** | |
| 1 | David Miller, Sarah White, "Foundations of Multidimensional Data Analytics", Cambridge University Press, 2021 |
| 2 | Laura Harris, Charles Wilson. "Applications of Multidimensional Data Analytics: Case Studies and Examples", Routledge, 2021 |
| **Useful Links** | |
| 1 | Introduction to multivariate statistical modelling: https://nptel.ac.in/courses/110105060 |
| 2 | Data Analytics with Python: https://onlinecourses.nptel.ac.in/noc21_cs45/preview |

**CO-PO Mapping**

| | **Programme Outcomes (PO)** | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | | 3 | | | | |
| **CO2** | | 3 | | | | |
| **CO3** | 1 | | 2 | | | 2 |
| **CO4** | 1 | | 3 | | | |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.
ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6. For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M. Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem II |
| **Course Code** | 7DS571 |
| **Course Name** | Data Mining and Warehousing Lab |
| **Desired Requisites:** | Statistics and Probability, Programming |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Practical** | 2 Hrs/ Week | **LA1** | **LA2** | **Lab ESE** | **Total** |
| **Interaction** | - | 30 | 30 | 40 | 100 |
| | | **Credits: 1** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To demonstrate basic concepts of data processing, data warehousing and data mining. |
| **2** | To introduce data mining algorithms and tools for analysing data. |
| **3** | To inculcate fundamental concepts that provides foundation for data mining. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Distinguish tools in data mining for data preprocessing, classification, association rule, clustering etc. | II | Understanding |
| **CO2** | Apply data preprocessing, exploration and visualization techniques. | III | Applying |
| **CO3** | Implement data mining algorithms using data mining tools. | III | Applying |
| **CO4** | Analyze data mining results obtained using various data mining algorithms. | IV | Analyzing |
| **CO5** | Evaluate performance of data mining algorithms. | V | Evaluating |
| **CO6** | Formulate data mining solution for real-world problems including multidisciplinary problems. | VI | Creating |

**List of Experiments / Lab Activities/Topics**

**List of Lab Activities:**

Experiment 1: Perform data smoothing, data transformation.

Experiment 2: Perform data Normalization (Min-max and Z-score).

Experiment 3: Extract Five-point summary for dataset.

Experiment 4: Plotting various types of graphs from dataset.

Experiment 5: Perform Data Preparation and Exploration, Apply Visualization Techniques.

Experiment 6: Performance Metrics and Assessment Metrics for Prediction and Classification.

Experiment 7: Generate Association Rules using the Apriori algorithm.

Experiment 8: Build a Decision Tree using ID3 algorithm.

Experiment 9: Demonstrate classification process on a given dataset using Naïve Bayesian Classifier.

Experiment 10: Supervised Learning Methods, linear and Logistic Regression.

Experiment 11: Cluster the given dataset by using the k-Means algorithm and visualize the cluster mean values and standard deviation of dataset attributes.

Experiment 12: Perform various data mining tasks using WEKA and KNIME OSS

Experiment 13: Using some sample data sets implement and test data mining techniques

Experiment 14: Mini-Project based any data mining application.

| | |
|---|---|
| | |
| **Textbooks** | |
| 1 | Jiawei Han and Micheline Kamber, "Data Mining – Concepts and Techniques", 4rd Edition, The Morgan Kaufmann Series in Data Management Systems, 2022. |
| 2 | Ian Witten, Eibe Frank and Mark Hall, "Data Mining: Practical Machine Learning Tools and Techniques", 3rd Edition, 2011 |
| 3 | Chattamvelli Rajan, "Data Mining Meth99ods: Concepts & Applications", Narosa Publishing House, 2nd Edition, 2016 |
| | |
| **References** | |
| 1 | Chris Pal, Ian Witten, Eibe Frank, and Mark Hall, "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann Series in Data Management Systems, 4th Edition, 2013 |
| 2 | Bostjan Kaluza, "Instant Weka How-to", Packt Publishing Limited, June 2013 |
| | |
| **Useful Links** | |
| 1 | https://nptel.ac.in/courses/110/107/110107092/ |
| 2 | https://nptel.ac.in/courses/110/107/110107095/ |
| 3 | https://www.kdnuggets.com/websites/index.html |

| CO-PO Mapping | | | | | | |
|---|---|---|---|---|---|---|
| | Programme Outcomes (PO) | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | | 3 | | | | |
| **CO2** | | 1 | | | | |
| **CO3** | | 3 | 1 | | 1 | 3 |
| **CO4** | | | 2 | | | |
| **CO5** | 1 | | 2 | | | 1 |
| **CO6** | 1 | | 3 | | 3 | 3 |
| The strength of mapping is to be written as 1: Low, 2: Medium, 3: High<br>Each CO of the course must map to at least one PO. | | | | | | |

| Assessment | | | | |
|---|---|---|---|---|
| There are three components of lab assessment, LA1, LA2 and Lab ESE.<br>IMP: Lab ESE is a separate head of passing.(min 40 %), LA1+LA2 should be min 40% | | | | |
| **Assessment** | **Based on** | **Conducted by** | **Typical Schedule** | **Marks** |
| LA1 | Lab activities, attendance, journal | Lab Course Faculty | During Week 1 to Week 4<br>Marks Submission at the end of Week 4 | 30 |
| LA2 | Lab activities, attendance, journal | Lab Course Faculty | During Week 5 to Week 9<br>Marks Submission at the end of Week 9 | 30 |
| Lab ESE | Lab activities, journal/ performance | Lab Course Faculty and External Examiner as applicable | During Week 9 to Week 13<br>Marks Submission at the end of Week 13 | 40 |
| Week 1 indicates starting week of a semester. Lab activities/Lab performance shall include performing experiments, mini-project, presentations, drawings, programming, and other suitable activities, as per the nature and requirement of the lab course. The experimental lab shall have typically 8-10 experiments and related activities if any. | | | | |

**Walchand College of Engineering, Sangli**
*(Government Aided Autonomous Institute)*

**AY 2024-25**

## Course Information

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech, Sem II |
| **Course Code** | 7DS572 |
| **Course Name** | Data Handling and Visualization lab |
| **Desired Requisites:** | Programming Fundamentals |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 2 Hrs/Week | **LA1** | **LA2** | **Lab ESE** | **Total** |
| **Tutorial** | - | 30 | 30 | 40 | 100 |
| | | **Credits: 1** | | | |

## Course Objectives

| | |
|---|---|
| **1** | Develop proficiency in creating and using various functions in R. |
| **2** | Gain skills in generating and manipulating different types of data structures in R. |
| **3** | Learn to visualize data using various plotting techniques in R. |
| **4** | Acquire the ability to produce professional reports and presentations using LaTeX. |

## Course Outcomes (CO) with Bloom's Taxonomy Level

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Understand and implement functions, including recursive ones, in R. | II | Understanding |
| **CO2** | Effectively construct and visualize data through line graphs, pie charts, bar plots, histograms, and scatter plots in R. | IV | Analyzing |
| **CO3** | Collect, manipulate, and analyze matrices and data frames using R. | IV/V | Analyzing /Creating |
| **CO4** | Produce high-quality reports and presentations using LaTeX. | V | Creating |

| List of Experiments / Lab Activities/Topics |
|---|
| **List of Lab Activities:** 8 to 10 Assignments based on following:<br>1. Create functions in R.<br>2. Implement recursive functions in R.<br>3. Generate and manipulate matrices in R.<br>4. Utilize factors in R.<br>5. Work with data frames in R.<br>6. Create line graphs, pie charts, and bar plots in R.<br>7. Generate histograms and scatter plots in R.<br>8. Implement regression analysis in R.<br>9. Apply k-means clustering in R.<br>10. Prepare a report using LaTeX.<br>11. Create a presentation using Beamer in LaTeX...<br>**Best Practices for lab:**<br>o Writing clean and readable code<br>o Testing and debugging<br>o Documentation and comments<br>o Version control with Git |

| Textbooks | |
|---|---|
| 1 | Dr. Mark Gardner, "Beginning R:statistical Programming Languages," Wrox (Amazon),Mar 2013 |
| 2 | Griffithas, Higham," Learning LATEX" , Society for Industrial and Applied Mathematics , 2nd Edition ,2016. |
| | |
| **References** | |
| 1 | Basic Data Analysis Tutorial, by Jacob Whitehill, Department of Computer Science, University of the Western Cape, [UWCDataAnalysisTutorial.pdf] |
| 2 | NPTEL,edx,COURSERA (MOOC courses) |
| | |
| **Useful Links** | |
| 1 | https://onlinecourses.nptel.ac.in/noc24_mg113/preview |
| 2 | https://www.overleaf.com/learn/latex/Free_online_introduction_to_LaTeX_(part_1) |

| CO-PO Mapping | | | | | | |
|---|---|---|---|---|---|---|
| | Programme Outcomes (PO) | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | 2 | | | | | |
| **CO2** | 1 | 2 | | | | |
| **CO3** | | 2 | 2 | | | 2 |
| **CO4** | | | 3 | | | |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

| Assessment | | | | |
|---|---|---|---|---|
| There are three components of lab assessment, LA1, LA2 and Lab ESE.<br>IMP: Lab ESE is a separate head of passing.(min 40 %), LA1+LA2 should be min 40% | | | | |
| **Assessment** | **Based on** | **Conducted by** | **Typical Schedule** | **Marks** |
| LA1 | Lab activities, attendance, journal | Lab Course Faculty | During Week 1 to Week 4 Marks Submission at the end of Week 8 | 30 |
| LA2 | Lab activities, attendance, journal | Lab Course Faculty | During Week 5 to Week 9 Marks Submission at the end of Week 9 | 30 |
| Lab ESE | Lab activities, journal/ performance | Lab Course Faculty and External Examiner as applicable | During Week 10 to Week 13 Marks Submission at the end of Week 13 | 40 |
| Week 1 indicates the starting week of a semester. Lab activities/Lab performance shall include performing experiments, mini-project, presentations, drawings, programming, and other suitable activities, as per the nature and requirement of the lab course. The experimental lab shall have typically 8-10 experiments and related activities if any. | | | | |

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem II |
| **Course Code** | 7DS573 |
| Course Name | Multidimensional Data Analysis Lab |
| **Desired Requisites:** | Python Programming |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Practical** | 2 Hrs/ Week | **LA1** | **LA2** | **Lab ESE** | **Total** |
| **Interaction** | - | 30 | 30 | 40 | 100 |
| | | **Credits: 1** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To gain hands-on experience in manipulating and visualizing multidimensional datasets, |
| **2** | To apply practical data mining and analysis techniques using relevant software tools. |
| **3** | To develop proficiency in creating and interpreting multidimensional models for real-world applications. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Convert data into appropriate forms by outlining data handling methodologies | II | Understanding |
| **CO2** | Apply dimensionality reduction on datasets by practicing relevant algorithms | III | Applying |
| **CO3** | Separate data patterns into various classes by integrating learning techniques | IV | Analyzing |
| **CO4** | Verify data classification and its assessment by discussing with domain experts | V | Evaluating |

**List of Experiments / Lab Activities/Topics**

**List of Lab Activities:**

1. Collect s and store datasets containing spatial, temporal, and spatiotemporal data types
2. Manipulate matrices, tensors, and data cubes using software tools (e.g. R, Python or equivalent)
3. Set up the environment for data analysis tasks using selected software (e.g., R Studio, Jupyter Notebook Python or equivalent) Execute basic operations to handle multidimensional data and interpret results
4. Apply techniques for de-noising data, handling missing values and outlier detection on real-world datasets. Evaluate the impact of data pre-processing on subsequent analysis outcomes
5. Implement normalization methods such as Min-Max scaling and Z-score normalization. Compare the effects of different normalization techniques on data distributions
6. Employ feature selection techniques (e.g., correlation analysis, forward/backward selection or similar)
7. Utilize feature extraction methods (e.g., PCA to reduce dimensionality and visualize results
8. Utilize feature extraction methods (e.g., LDA) to reduce dimensionality and visualize results
9. Implement supervised learning algorithms (e.g., linear regression, SVM) and evaluate model performance
10. Apply unsupervised learning techniques (e.g., clustering, dimensionality reduction) to analyze multidimensional data

| | |
|---|---|
| **Textbooks** | |
| 1 | Robert Wilson, Alice Davis, “Multidimensional Data Analytics: Techniques and Tools”, CRC Press, 2020 |
| 2 | Jake VanderPlas, "Python Data Science Handbook: Essential Tools for Working with Data", O'Reilly Media, 2016 |
| 3 | John Maindonald, W. John Braun, "Data Analysis and Graphics Using R: An Example-Based Approach", Cambridge University Press, 2020 |
| **References** | |
| 1 | Michael Brown, Jennifer Lee, "Advanced Methods in Multidimensional Data Analysis", Wiley, 2022 |
| 2 | Dan Toomey, Jonathan Whitmore, "Mastering Jupyter: Building Data Analytics and Machine Learning Platforms", Packt Publishing, 2018 |
| **Useful Links** | |
| 1 | Data Analytics with Python: **https://onlinecourses.nptel.ac.in/noc21_cs45/preview** |
| 2 | Essentials of Data Science With R Software **: https://onlinecourses.nptel.ac.in/noc21_ma35/preview** |

**CO-PO Mapping**

| | Programme Outcomes (PO) | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | | 3 | | | | |
| **CO2** | 2 | | 1 | | | 2 |
| **CO3** | | | 3 | | | |
| **CO4** | 2 | | 2 | | | |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

| Assessment | | | | |
|---|---|---|---|---|
| There are three components of lab assessment, LA1, LA2 and Lab ESE.<br>IMP: Lab ESE is a separate head of passing.(min 40 %), LA1+LA2 should be min 40% | | | | |
| **Assessment** | **Based on** | **Conducted by** | **Typical Schedule** | **Marks** |
| LA1 | Lab activities, attendance, journal | Lab Course Faculty | During Week 1 to Week 4<br>Marks Submission at the end of Week 8 | 30 |
| LA2 | Lab activities, attendance, journal | Lab Course Faculty | During Week 5 to Week 9 Marks Submission at the end of Week 9 | 30 |
| Lab ESE | Lab activities, journal/ performance | Lab Course Faculty and External Examiner as applicable | During Week 10 to Week 13<br>Marks Submission at the end of Week 13 | 40 |
| Week 1 indicates the starting week of a semester. Lab activities/Lab performance shall include performing experiments, mini-project, presentations, drawings, programming, and other suitable activities, as per the nature and requirement of the lab course. The experimental lab shall have typically 8-10 experiments and related activities if any. | | | | |

**Walchand College of Engineering, Sangli**
*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M. Tech. ( Data Science) |
| **Class, Semester** | First Year M. Tech., Sem II |
| **Course Code** | 7DS574 |
| **Course Name** | Seminar |
| **Desired Requisites:** | |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Practical** | 2 Hrs/ Week | **LA1** | **LA2** | **Lab ESE** | **Total** |
| **Interaction** | - | 30 | 30 | 40 | 100 |
| | | **Credits:1** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To Review and increase students' understanding of the specific topics. |
| **2** | To induce Learning management of values. |
| **3** | To teach how research papers are written and read such papers critically and efficiently and to summarize and review them to gain an understanding of a new field, in the absence of a textbook. |
| **4** | To teach how to judge the value of different contributions and identify promising new directions in specified area. |
| **5** | To write and present seminar content effectively. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Apply the existing knowledge to solve real life problems | III | Applying |
| **CO2** | Examine the selected topic/system using various methods. | IV | Analyzing |
| **CO3** | Justify the outcome of the work has solved the specified problem or not. | V | Evaluating |
| **CO4** | Build and present the seminar report in an effective way. | VI | Creating |

**Lab Activities**

**Contents:**

The pre-dissertation work will start in semester II and should preferably be a problem with research potential and should involve scientific research review, design, generation/collection and analysis of data, determining solution and must preferably bring out the individual contribution. Seminar should be based preferably on the area in which the candidate is interested to undertake the dissertation work. The candidate has to be in regular contact with their guide and the topic of seminar/dissertation must be mutually decided. The examination shall consist of the preparation of report consisting literature review, detailed problem statement, case studies, etc, according to type of work carried out. The work has to be presented in front of the examiners panel formed by Dept. for evaluation.

| Useful Links | |
|---|---|
| 1 | https://onlinecourses.nptel.ac.in/noc19_ge21/preview |

| CO-PO Mapping | | | | | | |
|---|---|---|---|---|---|---|
| | Programme Outcomes (PO) | | | | | |
| | **1** | **2** | **3** | **4** | **5** | **6** |
| **CO1** | 3 | | 1 | | | |
| **CO2** | 1 | 3 | | | | |
| **CO3** | | | 3 | | 1 | 1 |
| **CO4** | | | | 3 | 1 | |
| The strength of mapping is to be written as 1,2,3; Where, 1:Low, 2:Medium, 3:High<br>Each CO of the course must map to at least one PO. | | | | | | |

**Assessment**

There are three components of lab assessment, LA1, LA2 and Lab ESE.
IMP: Lab ESE is a separate head of passing. LA1, LA2 together is treated as In-Semester Evaluation.

| Assessment | Based on | Conducted by | Typical Schedule | Marks |
|---|---|---|---|---|
| LA1 | Lab activities, attendance, journal | Lab Course Faculty | During Week 1 to Week 4<br>Marks Submission at the end of Week 4 | 30 |
| LA2 | Lab activities, attendance, journal | Lab Course Faculty | During Week 5 to Week 10<br>Marks Submission at the end of Week 10 | 30 |
| Lab ESE | Lab activities, attendance, journal | Lab Course Faculty | During Week 11 to Week 13<br>Marks Submission at the end of Week 13 | 40 |

Week 1 indicates starting week of a semester. Lab activities/Lab performance shall include performing experiments, mini-project, presentations, drawings, programming, and other suitable activities, as per the nature and requirement of the lab course. The experimental lab shall have typically 8-10 experiments and related activities if any.

# Track wise- Electives

| Elective course name | Level | T1-Mathematical Data Analysis | T2-Data Modelling | T3-Data Science Applications |
|---|---|---|---|---|
| Statistical Inference | 1 | YES | NO | YES |
| Time Series Data Analysis | 1 | YES | YES | YES |
| Multi-Criteria Decision Making | 1 | YES | YES | YES |
| Data Modeling and Simulation | 1 | YES | YES | YES |
| Data-driven Analytics | 2 | YES | YES | NO |
| AIML in Data Science | 2 | YES | YES | YES |
| Numerical Optimization in Data Science | 2 | YES | YES | YES |
| Graph Theory in Data Science | 2 | YES | YES | NO |
| Pattern Recognition | 3 | YES | YES | YES |
| Financial Data Science | 3 | NO | YES | YES |
| Social Data Analysis | 3 | NO | YES | YES |
| Data Science in Businesses | 3 | YES | YES | YES |
| Game theory | 3 | YES | YES | YES |
| | | 11 | 12 | 11 |

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

| **Course Information** | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech. Sem I/II |
| **Course Code** | 7DS511 |
| **Course Name** | Statistical Inference |
| **Desired Requisites:** | |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

| **Course Objectives** | |
|---|---|
| **1** | Understand the Fundamentals concepts of statistical inference |
| **2** | Develop Skills in Estimation and Hypothesis Testing |
| **3** | Apply Statistical Models to Real-World Data |
| **4** | Cultivate Critical Thinking and Decision-Making Abilities |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Develop proficiency in applying Statistical Techniques | II | Applying |
| **CO2** | Interpret and Communicate Statistical Findings using various parameters | II | Applying |
| **CO3** | Discover Competence in using Statistical Software | III | Analyzing |
| **CO4** | Perceive Critical Thinking and Analytical Skills | IV/V | Analyzing Evaluating |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Principle of Data Reduction:** Sufficiency principle, Factorization criterion, minimal sufficiency, Completeness and bounded completeness, Likelihood principle, Equivariance principle. | 6 |
| II | **Theory of Estimation:** Basic concepts of estimation, Methods of Point estimation, Methods of Interval Estimation, Methods of of Least Squares Estimation; Method of moments- Properties of Moment Estimators, Drawbacks of Moment Estimators, Method of maximum likelihood- Properties of Maximum Likelihood Estimators | 6 |
| III | **Theory of Estimation:** Un-biasedness, Minimum variance estimation, Cramer – Rao bound and its generalization, Rao Blackwell theorem, Existence of minimum-variance unbiased estimator (MVUE) or uniformly minimum-variance unbiased estimator (UMVUE). Interval Estimation, Some results for normal population cases. | 7 |
| IV | **Testing of Hypothesis:** Tests of Hypotheses, Null and alternative hypothesis, Type I and II errors error probability and power function, Method of finding tests. | 6 |

| | | |
|---|---|---|
| V | **Testing of Hypothesis:**<br>Neyman – Pearson lemma, Uniformly most powerful tests, Likelihood ratio principle, Likelihood ratio test, Sequential probability ratio test, Some results based on normal population. | 7 |
| VI | **Analysis of Variance (ANOVA):**<br>Analysis of Variance: Basic Concepts Source of Variance ,One-Way Classification Model, simple linear regression analysis with normal distribution ,Test Procedure, Sums of Squares ,Preparation of ANOVA Table | 6 |

| **Textbooks** | |
|---|---|
| 1 | Miller, I. and Miller, M.,"Freund's Mathematical Statistics with Applications", Prentice Hall PTR, 7th edition,2006 |
| 2 | Lehman, E.L., "Testing of Statistical Hypothesis", Wiley Eastern Ltd, 3rd edition 2008 |
| 3 | G. Casella, R. L. Berger, "Statistical Inference", Duxbury Press ,2nd edition 2002 |

| **References** | |
|---|---|
| 1 | Lehman, E.L., "Point Estimation", John Wiley & sons , 1984 |
| 2 | Rohatgi, V.K., "Statistical Inference", Dover Publications, Dover Ed, 2003 |

| **Useful Links** | |
|---|---|
| 1 | https://archive.nptel.ac.in/courses/111/105/111105043/ |
| 2 | https://www.youtube.com/@statisticsfordatascience_i793 |

| **CO-PO Mapping** | | | | | | |
|---|---|---|---|---|---|---|
| | **Programme Outcomes (PO)** | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | 2 | | | | | |
| **CO2** | | 2 | | | | |
| **CO3** | | | 2 | | | |
| **CO4** | | | | 3 | | 1 |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.
ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6. For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

| **Course Information** | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem I/II |
| **Course Code** | 7DS512 |
| **Course Name** | Time Series Data Analysis |
| **Desired Requisites:** | |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

| **Course Objectives** | |
|---|---|
| **1** | Develop a thorough understanding of time series concepts and components. |
| **2** | Master techniques for modeling and forecasting time series data using various methods. |
| **3** | Gain proficiency in identifying and handling trends, seasonality, and noise in time series datasets. |
| **4** | Acquire the ability to apply time series analysis to real-world business and economic data. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Classify and analyze the components of time series data. | II/III | Understanding/ Applying |
| **CO2** | Examine and mitigate trends, seasonality, and irregularities in time series datasets. | III | Applying |
| **CO3** | Apply time series analysis techniques using various models to derive actionable insights from business and economic data. | III/V | Applying/ Evaluating |
| **CO4** | Develop the ability to build and evaluate time series models for forecasting. | V/VI | Evaluating/ Creating |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Basic Properties of time-series data:** Distribution and moments, Stationary, Autocorrelation, Heteroscedasticity, Normality Autoregressive models and forecasting: Auto-Regressive (AR) , Auto Regressive Moving Average (ARMA), Auto-Regressive Integrated Moving Average (ARIMA) models | 7 |
| II | **Models:** Random walk model, Non-stationary and unit-root process, Drift and Trend models. Regression analysis with time-series data using R programming | 6 |
| III | **Analysis:** Implementation of Regression analysis with time-series data using R programming. Basics of Principal Component Analysis (PCA) and Factor Analysis | 7 |
| IV | **Conditional Heteroscedastic Models:** Analyze volatility in high frequency data using various models. Autoregressive Conditional Heteroscedasticity (ARCH), Generalized autoregressive conditional Heteroscedasticity (GARCH). Threshold generalized autoregressive conditional | 6 |

| | | |
|---|---|---|
| | Heteroscedasticity (T-GARCH), Baba, Engle, Kraft and Kroner Generalized autoregressive conditional Heteroscedasticity (BEKK- GARCH) model. | |
| V | **Introduction to Non-linear and regime-switching models:** Nonlinear Models, Nonlinearity Tests, Modeling, Forecasting, Application, Introduction to Regime Shift Models in Time Series. Types of Regime Shift Models, Markov Switching Autoregressive Model, Quantile regression, Contagion models | 6 |
| VI | **Introduction to Vector Auto-regressive (VAR) models:** Vector Autoregressive Models (VAR) details, VARs with Exogenous Variables , Example of VAR model ,Impulse Response Function (IRF), Error Correction Models, Co-integration ,Introduction to Panel data models: Fixed-Effect and Random-Effect models | 7 |

| **Textbooks** | |
|---|---|
| 1 | Ruey S. Tsay "Analysis of Time-series data," Third Edition, Wiley,2014 |
| 2 | John Fox and Sanford Weisberg "An R Companion to Applied Regression," Third Edition, SAGE, 2018 |
| 3 | Chris Brooks "Introductory Econometrics for Finance," Fourth Edition, Cambridge University Press, 2019 |

| **References** | |
|---|---|
| 1 | Yves Croissant and Giovanni Millo "Panel Data Econometrics with R," First Edition, Wiley, 2018 |

| **Useful Links** | |
|---|---|
| 1 | https://archive.nptel.ac.in/courses/103/106/103106123/ |
| 2 | https://www.youtube.com/@statisticsfordatascience_i793 |

| **CO-PO Mapping** | | | | | | |
|---|---|---|---|---|---|---|
| | **Programme Outcomes (PO)** | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | 2 | | | | | |
| **CO2** | | 2 | | | | |
| **CO3** | | | 3 | | | |
| **CO4** | | | 2 | | | 2 |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.
For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

| **Course Information** | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem I/II |
| **Course Code** | 7DS513 |
| **Course Name** | Multi-Criteria Decision Making |
| **Desired Requisites:** | |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

| **Course Objectives** | |
|---|---|
| **1** | Understand the fundamental principles and frameworks of multi-criteria decision making (MCDM). |
| **2** | Learn to apply various MCDM methods to evaluate and prioritize alternatives in complex decision scenarios. |
| **3** | Develop skills to incorporate stakeholder preferences and conflicting criteria in the decision-making process |
| **4** | Acquire the ability to use MCDM software tools for solving real-world decision problems. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Demonstrate a comprehensive understanding of MOO and MCDM principles and frameworks. | II/III | Understanding/ Applying |
| **CO2** | Apply different MCDM methods to analyze, to optimize and rank alternatives effectively. | III | Applying |
| **CO3** | Examine stakeholder preferences and manage conflicting criteria in decision-making using data manipulation. | IV | Analyzing |
| **CO4** | Use MCDM software tools to solve complex decision problems and present their analysis. | II/IV | Applying/ Analyzing |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Introduction**:<br>Defining criteria and alternatives in the context of decision-making. Overview of the history and evolution of multi-criteria decision making (MCDM), Key milestones and influential figures in the development of MCDM, Review of decision making process in optimization and operations research models; overview of machine learning algorithms; ranking methods. | 6 |
| II | **Multi Objective Optimization (MOO):** Introduction to multi objective optimization, Linear and Nonlinear MOOP -Convex and No convex MOOP, Principles of Multi-Objective Optimization- Illustrating Pareto-Optimal Solutions Objectives in Multi-Objective Optimization, Non-Conflicting Objectives, Difference with Single-Objective Optimization - Two Goals, Two Search Spaces No Artificial | 6 |

| | | |
|---|---|---|
| | Fix-Ups | |
| III | **Classical and recent methods**: Multi objective optimization -Definition and significance, Differences between single-objective and multi-objective optimization, Commonly used terminologies (Pareto front, Pareto optimality, dominance, trade-offs), Genetic Algorithms (GA)-Overview of GA and Representation of solution. Particle Swarm Optimization (PSO)-Overview of PSO, Swarm intelligence and behavior of particles. | 7 |
| IV | **Multi Criteria Decision Making (MCDM):**<br>Introduction to MCDM methods; group decision making, Single Criterion Methods-Cost-benefit analysis, Utility theory. Multi-Criteria Methods:-Weighted Sum Model (WSM),Weighted Product Model (WPM), Analytic Hierarchy Process (AHP),Preference Ranking Organization Method for Enrichment Evaluations (PROMETHEE) | 7 |
| V | **Data Manipulation:**<br>Data wrangling and data management for large sized multi objective and multi criteria problems ,Structuring data for multi-objective optimization, Preparing data for multi-criteria decision analysis (MCDA), Criteria weighting and normalization, Creating decision matrices | 6 |
| VI | **Implementation**:<br>Python libraries commonly used in Multi Criteria Decision Making (MCDM)-NumPy, Pandas, Scipy, Scikit-learn, Matplotlib, Installing necessary libraries. Setting up Jupyter Notebook or any preferred Python IDE. Data Handling and Preparation-Data Collection, Data Cleaning, Data Transformation. Implementing Basic MCDM Methods- Python code to implement Weighted Sum Model (WSM) and Weighted Product Model (WPM). Detailed implementation of MCDM on real-world datasets | 7 |
| **Textbooks** | | |
| 1 | G.H. Tzeng, J.J. Huang, "Multiple Attribute Decision Making: Methods and Applications", CRC Press. 1 st Edition, 2011 | |
| 2 | A.A.Keller "Multi-Objective Optimization in Theory and Practice I: Classical Methods", Bentham Science Publishers, 1st edition 2017. | |
| 3 | M. Köksalan. J. Wallenius, S. Zionts, "Multiple Criteria Decision Making. From Early History to the 21st Century", World Scientific, 1st edition 2011. | |
| 4 | J. Branke, K. Deb, K. Miettinen, R. Slowinski (Eds.), "Multiobjective Optimization: Interactive and Evolutionary Approaches", Springer-Verlag, Berlin, Heidelberg, 2008 | |
| **References** | | |
| 1 | A. Ishizaka, P. Nemery, "Multicriteria Decision Aid: Methods and software", Wiley, Chichester, 2013 | |
| 2 | K. Deb, "Multi-Objective Optimization Using Evolutionary Algorithms", J.Wiley & Sons, 2001. | |
| 3 | Michael Carter, Camille C. Price and Ghaith Rabadi "Operations Research, A Practical Introduction", CRC Press,2nd edition 2023 | |
| **Useful Links** | | |
| 1 | https://onlinecourses.nptel.ac.in/noc24_ge01/preview | |

| CO-PO Mapping | | | | | | |
|---|---|---|---|---|---|---|
| | Programme Outcomes (PO) | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | 1 | 2 | | | | |
| **CO2** | 2 | 2 | | | | |
| **CO3** | | | 2 | | | |
| **CO4** | | 3 | | | | 1 |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.
ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.
For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem I/II |
| **Course Code** | 7DS514 |
| **Course Name** | Data Modeling and Simulation |
| **Desired Requisites:** | |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To understand the core concepts and methodologies of data modeling and simulation. |
| **2** | Learn to develop and validate various models to represent real-world systems and processes. |
| **3** | Gain proficiency in using simulation techniques to analyze the behavior and performance of complex systems. |
| **4** | Acquire the ability to apply data modeling and simulation tools to solve practical problems in business and engineering. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Demonstrate understanding of data modeling and simulation principles and methodologies. | II | Understanding |
| **CO2** | Build, Validate, and refine models that accurately represent real-world systems. | IV | Analyzing |
| **CO3** | Evaluate, Optimize and predict the performance of complex systems using various tools/techniques. | III/V | Applying/ Evaluating |
| **CO4** | Apply data modeling and simulation tools to address and solve real-world business and engineering challenges | III/V | Applying/ Evaluating |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Introduction**:<br>Introduction to spreadsheets; historical development; basic capabilities of spreadsheets and their usage for creating models; types of data used in spreadsheets; spreadsheet notations for mathematical operations; common built-in formulas and functions; conditional expressions; relative and absolute references. | 7 |
| II | **Model building**:<br>Designing spreadsheets reflecting assumptions; decision variables; and outcomes, creating basic cash-flow models; revaluating small business opportunities; incorporating what-if analysis; identifying key variables using sensitivity analysis; linear programming models and deterministic models. | 7 |
| III | **Optimization with Spreadsheets using Solver**:<br>Spreadsheet Solvers, Solving Linear Programming (LP) Problems in a Spreadsheet, Implementing an LP Model in a Spreadsheet, A Spreadsheet Model for the Blue Ridge Hot Tubs Problem, Express Optimization Problems | 7 |

| | | |
|---|---|---|
| | Mathematically-Decisions, Constraints. The Purpose of Sensitivity Analysis. | |
| IV | **Optimization with Spreadsheets using Solver**<br>Integer and nonlinear programming, multi-objective optimization, applications of optimization in different areas. Optimizing Existing Financial Spreadsheet Models, Implementing the Model, Optimizing the Spreadsheet Model Analyzing the Solution. | 6 |
| V | **Simulation and Optimization**:<br>Use of spreadsheets to implement Monte Carlo simulations and linear programs for optimization; model uncertainty and risk in spreadsheets; Spreadsheet Simulation Using Analytic Solver Platform, Preparing the Model for Simulation | 6 |
| VI | **Case Studies-**<br>Healthcare-Disease Spread Modeling, Finance-Risk Management, Environmental Science-Climate Change Modeling etc. | 6 |

| **Textbooks** | |
|---|---|
| 1 | Hillier and Hillier "Introduction to Management Science: A Modeling and Case Studies Approach with Spreadsheets", McGraw-Hill/Irwin, 6th edition 2019 |
| 2 | Cliff Ragsdale "Spreadsheet Modeling and Decision Analysis: A Practical Introduction to Business Analytics", Cengage India,8th edition 2017(module 3,4) |
| 3 | Barry Render, Nagraj Balakrishnan, and Ralph Stair, "Managerial Decision Modelling with Spreadsheets", Pearson, 1st edition 2003 |
| **References** | |
| 1 | S. Christian Albright and Wayne Winston "Spreadsheet Modeling and Applications: Essentials of Practical Management Science", Cengage. 1st edition 2004 |
| **Useful Links** | |
| 1 | https://archive.nptel.ac.in/courses/112/107/112107220/ |
| 2 | https://onlinecourses.nptel.ac.in/noc19_mg45/preview |

| **CO-PO Mapping** | | | | | | |
|---|---|---|---|---|---|---|
| | **Programme Outcomes (PO)** | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | 3 | | | | | 1 |
| **CO2** | 2 | 1 | 1 | | | |
| **CO3** | 1 | 1 | 2 | | | 2 |
| **CO4** | | 3 | 3 | | 2 | 2 |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.
ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6
For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem I/II |
| **Course Code** | 7DS515 |
| **Course Name** | Data-driven Analytics |
| **Desired Requisites:** | |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To familiarize with the applications of data science in traffic and transportation engineering |
| **2** | To Analyze large datasets to uncover meaningful patterns and trends. |
| **3** | To develop predictive models using statistical and machine learning techniques. |
| **4** | To optimize and Implement data-driven strategies to enhance decision-making. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Demonstrate skills in data collection, preparation, and visualization for transportation systems, | III | Applying |
| **CO2** | Experiment urban transportation planning using diverse data techniques and analytical tools | III | Applying |
| **CO3** | Apply IoT and machine learning to solve real-time problems. | III | Applying |
| **CO4** | Analyze the planning and analysis of urban mass transit systems using real-time data and etc. | IV | Analyzing |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Overview and Practical Applications:** Data Sources; Planning and Modelling., Characteristics of Problems ,Data Collection; Data Preparation and Visualization. | 6 |
| II | **Data Analytics and Planning:** Basics of Planning, Data Collection and Advanced Data Sources, Surveys, Demand Modeling using WiFi/ Bluetooth/ Call Data Record, Data Extraction and Analysis using APIs, Modeling Approaches, Genetic Algorithms, Case study for Planning -Example using Data-Driven Modeling and Simulation. | 7 |
| III | **Data-driven Analytics using Intelligent Systems:** Internet of Things (IOT), Machine Learning, Real-Time Monitoring and Data Analysis, Analysis of Key Parameters, and Development of Policy Framework. | 6 |
| IV | **Crash Data Analytics:** Crash Data, Data Preparation, Model Estimation, Real-Time Data-Driven Analysis; Emergency Data, Crash Prone Stretches, Conflict Data, Surrogate Approach, Proactive Assessment and Safety Interventions | 7 |
| V | **Urban Mass Transit System:** Basics of Urban Mass Transit System, Static and Dynamic | 6 |

| | | |
|---|---|---|
| | General Transit Feed Specification (GTFS), Real-Time Transit, Travel Time Variability, Transit Reliability, Transit Planning using Smart-Card Data, Real-Time Accessibility Analysis. | |
| VI | **Sensing and Data Mining for Smart Transportation Systems:** Intelligent Systems, Incident Management Program, Efficient Emergency Vehicle Movement (Pre-Emption), Crash Detection, Reporting, and Clearance; Traffic Surveillance, Identification of Hotspots, Violation Detection; Road Network Asset Management, Inventory of Potholes, other Deficiencies; Adaptive Traffic Signal. | 7 |

| **Textbooks** | |
|---|---|
| 1 | Fumitaka Kurauchi, Jan-Dirk Schmöcker "Public transport planning with smart card data" CRC Press, 1st edition 2021 |
| 2 | Juan de Dios Ortúzar, Luis G. Willumsen "Modelling Transport", Wiley,4th edition,20011 |
| 3 | Constantinos Antoniou, Loukas Dimitriou, Francisco Pereira "Mobility Patterns, Big Data and Transport Analytics" Elsevier, 1st edition 2019 |
| **References** | |
| 1 | Sara Moridpour, Alireza Toran Pour, Tayebeh Saghapour "Big Data Analytics in Traffic and Transportation Engineering: Emerging Research and Opportunities" IGI Global,2019 |
| 2 | Khaled R. Ahmed, Aboul-Ella Hassanien "Deep Learning and Big Data for Intelligent Transportation" Springer, 1st edition 2021 |
| 3 | Davy Janssens, Ansar-Ul-Haque Yasar and Luk Knapen "Data Science and Simulation in Transportation Research" IGI Global,2013 |
| **Useful Links** | |
| 1 | https://onlinecourses.nptel.ac.in/noc22_ce34/preview |
| 2 | https://onlinecourses.nptel.ac.in/noc19_mg45/preview |

**CO-PO Mapping**

| | **Programme Outcomes (PO)** | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | 2 | | | | | |
| **CO2** | | 2 | 1 | | | |
| **CO3** | | 1 | 2 | | | 2 |
| **CO4** | | | 2 | | | |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.
For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech, Sem I/II |
| **Course Code** | 7DS516 |
| **Course Name** | AI-ML in Data Science |
| **Desired Requisites:** | |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To Utilize AI and machine learning algorithms to extract actionable insights from complex datasets. |
| **2** | To Develop robust models for classification, regression, clustering, and anomaly detection in data science applications. |
| **3** | To Implement scalable solutions for data preprocessing, feature engineering, and model deployment in AI-driven projects |
| **4** | To Enhance data-driven decision-making processes through the integration of AI and ML techniques in data science workflows. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Get knowledge of Earth Science domains and data types, and apply to support ML and AI applications | II/III | Understanding /Applying |
| **CO2** | Analyze automation in Data Mining based on sophisticated AI-driven analytics of Earth Sciences data. | IV | Analyzing |
| **CO3** | Explore classification of Earthquakes Sources operations through automated data processing and real-time insights. | II | Applying |
| **CO4** | Apply AI and ML techniques to discover new patterns and opportunities within various data science applications studies. | II /IV | Applying/ Analyzing |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Introduction:**<br>Major Domains and Data Types in Earth Sciences- Earthquake Seismology, Engineering Geology and Rock Mechanics, Reservoir Characterization, **Machine learning (ML)** and statistical pattern recognition- Supervised learning (generative/ descriptive learning, parametric/ non-parametric learning, neural networks, Support vector machines) | 6 |
| II | **Introduction to Machine learning (ML) and Artificial Intelligence (AI) in Earth Sciences:**<br>Unsupervised learning (clustering, dimensionality reduction, kernel methods); time series modelling, linear regression, regularization, linear classifiers, ensemble methods, neural networks, model selection and evaluation, scalable algorithms for big data, and data ethics. Data science: Extreme value statistics, multi-vitiate analysis, factor analysis, and compositional data analysis. | 7 |

| | | |
|---|---|---|
| III | **Automating Data Mining and Analysis:**<br>Basics of earthquake detection and phase picking using short-term average (STA)/long-term average (LTA); Detection using waveform similarity- Network Matched Filtering/template matching, Fingerprint And Similarity Thresholding (FAST), Association of seismic phases across all stations using deep-learning techniques | 6 |
| IV | **Classification:**<br>Classification using supervised learning to classify earthquakes, finding occurrence mechanism. Training dataset (waveforms) on different kinds of sources: -earthquake, glacial, volcanic, landslide, explosion, etc. Seismic sources and radiation pattern of emerging waves. Deep learning (DL) based Seismic Inversion- Theory of Seismic Inversion, Convolutional neural network (CNN) and fully connected network (FCN) architectures, Performance evaluation. | 7 |
| V | **Automation in 3D Reservoir Property Prediction:**<br>Data Mining, Automated Petrophysics, Statistical and Regression Methods for Elastic Property Prediction, ML and AI application in Geostatistics, Convoluted Neural Networks for Seismic Interpretation, Deep Learning for Impedance Inversion and Porosity Prediction. Data-Driven Analytics in Shale Resources | 6 |
| VI | **Machine learning (ML) Applications in Engineering Geology and Rock Mechanics:**<br>ML in rock mass characterization, Rock Mass Rating, Slope Mass Rating, Engineering properties of rock and various rock engineering applications, Artificial Intelligence in Landslides study. Separation and Taxonomic Identification of Microfossil: 3D object recognition and segmentation applied to X-ray MicroCT images. | 7 |

| **Textbooks** | |
|---|---|
| 1 | Patrick Wong, Fred Aminzadeh, and Masoud Nikravesh, "Soft Computing for Reservoir Characterization and Modeling" , Springer-Verlag Berlin Heidelberg GmbH , 1st edition 2002. |
| 2 | William Sandham & Miles Leggett, "Geophysical Applications of Artificial Neural Network and Fuzzy Logic", Springer ,3rd Edition, 2003. |
| 3 | C. Cranganu, H. Luchian, M. E. Breaban, Artificial Intelligent Approached in Petroleum Geosciences, Springer 1st edition 2015. |
| **References** | |
| 1 | Shahab D. Mohaghegh , Data-Driven Analytics in Unconventional Resources, Springer 2017 |
| **Useful Links** | |
| 1 | https://nptel.ac.in/courses/106105238 |
| 2 | https://onlinecourses.nptel.ac.in/noc19_cs82/preview- ML for Engineering and Science Applications |

| CO-PO Mapping | | | | | | |
|---|---|---|---|---|---|---|
| | Programme Outcomes (PO) | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | 2 | 2 | | | | |
| **CO2** | | | 3 | | | |
| **CO3** | 2 | 2 | | | | 2 |
| **CO4** | | | 3 | | | |
| The strength of mapping is to be written as 1: Low, 2: Medium, 3: High<br>Each CO of the course must map to at least one PO. | | | | | | |

| Assessment |
|---|
| The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.<br>ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.<br>ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.<br>For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing) |

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem I/II |
| **Course Code** | 7DS517 |
| **Course Name** | Numerical Optimization in Data Science |
| **Desired Requisites:** | |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To provide knowledge about basic concepts of Numerical Optimization. |
| **2** | Apply numerical optimization techniques to minimize/maximize objective functions in data science problems. |
| **3** | Develop algorithms for efficiently solving constrained and unconstrained optimization problems in large-scale datasets. |
| **4** | Enhance computational efficiency and accuracy in data-driven decision-making through advanced optimization methodologies. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Identify proficiency in various programming methods and their applications. | I/II | Understanding/ Applying |
| **CO2** | Demonstrate and apply one or multi-dimensional unconstrained or constrained minimization methods. | II | Applying |
| **CO3** | Investigate the scalability of algorithms for faster processing and optimize processes within data-driven workflows. | IV | Analyzing |
| **CO4** | Explore/Evaluate optimal solutions in complex data science challenge using various software packages and case studies. | V | Evaluating |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Introduction:** Optimization, Types of Problems and Algorithms Linear Programming:-Review of various methods of linear programming. Basic properties of solutions and algorithms, Global convergence. Line Search Methods, Steepest Descent and Newton Methods | 6 |
| II | **Nonlinear Programming 1-D Unconstrained Minimization Methods:** Overview of Nonlinear Programming, 1-D Unconstrained Minimization, Applications and Examples, Objective Functions and Their Properties-Continuity, Differentiability. Iterative Search Methods-Golden Section Search, Fibonacci Search, Derivative-Based Methods-Newton's Method, Secant Method, Fibonacci Search, Bisection. | 6 |
| III | **Multi-dimensional Unconstrained Minimization Methods:** Cyclic Coordinate Method, Hookes & Jeeves continuous and discrete methods, Rosenbrock method, Nelder & Mead method, Box's Complex method, Powell method, Steepest descent method, Newton's method, conjugate gradient method. | 7 |

| | | |
|---|---|---|
| IV | **Constrained Minimization:**<br>Constrained Optimization- First Order Necessary Conditions, Second Order Necessary Conditions, Duality, Constraint Qualification, Rosen's gradient projection method for linear constraints, Zoutendijk method of feasible directions for nonlinear constraints, generalized reduced gradient method for nonlinear constraints. | 7 |
| V | **Penalty function methods:**<br>Barrier methods, properties of penalty and barrier functions , newton's method and penalty functions , conjugate gradients and penalty methods , normalization of penalty functions ,penalty functions and gradient projection ,exterior point penalty, interior point penalty. | 6 |
| VI | **Case studies :**<br>Case studies from Engineering and Industry, Use of software packages such as Linear, Interactive, and Discrete Optimizer (LINDO), Temporally Ordered Routing Algorithm (TORA), EXCEL, MATLAB etc | 7 |

| **Textbooks** | |
|---|---|
| 1 | David Luenberger and Yinyu Ye, "Linear and Nonlinear Programming", 3rd Edition, Springer, 2008. |
| 2 | Bazaraa, M. S., Sherali, H. D. and Shetty, C. M "Nonlinear Programming Theory and Algorithms", 2nd Edition, John Wiley and Sons,2006 |
| 3 | Fletcher R., "Practical Methods of Optimization", John Wiley,2nd edition, 2000 |
| 4 | Belegundu, A. D. and Chandrupatla, T. R. :"Optimization Concepts and Applications in Engineering", Cambridge university space., 2nd edition 2011(unit v) |
| | |

| **References** | |
|---|---|
| 1 | Mohan, C. and Deep, K.: "Optimization Techniques", New Age India Pvt. Ltd., 2nd edition ,2009. |
| 2 | Nocedal, J. and Wright, S. J.: "Numerical Optimization", Springer Series in Operations Research, Springer-Verlag, 1st edition 2006. |
| 3 | Deb, K.: "Optimization for Engineering Design Algorithms and Examples", Prentice Hall of India. 2001 |

| **Useful Links** | |
|---|---|
| 1 | https://nptel.ac.in/courses/106108056 |
| 2 | https://www.iitg.ac.in/rkbc/ce602-2012.htm |
| | |

| **CO-PO Mapping** | | | | | | |
|---|---|---|---|---|---|---|
| | **Programme Outcomes (PO)** | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | 1 | 2 | | | | |
| **CO2** | | 1 | 2 | | | |
| **CO3** | | 2 | 2 | | | |
| **CO4** | | 2 | 2 | | | 1 |
| The strength of mapping is to be written as 1: Low, 2: Medium, 3: High.<br>Each CO of the course must map to at least one PO. | | | | | | |

| Assessment |
| --- |
| The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.<br>ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.<br>ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.<br>For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing) |

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem I/II |
| **Course Code** | 7DS518 |
| **Course Name** | Graph Theory in Data Science |
| **Desired Requisites:** | |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To explain the requirements of graph based solutions and hence algorithms in data science and data analytics |
| **2** | To compare graph theoretic approaches modeling relationships and dependencies in complex datasets |
| **3** | To discuss programming aspects, tools and techniques for handling graphs |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Distinguish data relationships and patterns through graph-based representations | II | Understanding |
| **CO2** | Chose efficient path finding and graph traversal algorithms | III | Applying |
| **CO3** | Analyze various queries and retrieve data patterns | IV | Analyzing |
| **CO4** | Assess graph based models with real-life applications | V | Evaluating |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Fundamentals:** Introduction to various graph models, Basics of Paths, Cycles, and Trails Isomorphic graphs, Spanning tree, connectivity in graphs. | 6 |
| II | **Graph :** Eulerian and Hamiltonian Graphs, matching, vertex coloring and domination, random graphs, graph traversal mechanisms. | 6 |
| III | **Graph Modelling with Neo4j:** Graph Databases- directed vs undirected, weighted vs unweighted, cyclic vs acyclic, dense vs sparse, connected vs disconnected, Cypher Query Language, nodes and relationships, managing databases with Neo4j, creating, selecting a node, filtering, creating a relationship, selecting relationship, updating and deleting nodes and relationships, pattern matching and data retrieval, aggregation functions, importing data from CSV to JSON, Empowering business with pure Cypher, knowledge graphs, graph-based search, recommendation engines. | 7 |
| IV | **Graph Algorithms:** The Graph Data Science Library and Path finding, Dijkstra's shortest path algorithm, A-star algorithm, k-shortest path, optimizing processes using graphs, | 7 |

| | | |
|---|---|---|
| | travelling salesman problem, spanning tress, prims algorithm, minimum spanning tree in a Neo4j graph. | |
| V | **Spatial data:**<br>Node importance, representation spatial attributes, creating a geometry layer with Neo4j, spatial queries, visualization spatial data with Neo4j, Community detection and similarity measures. | 6 |
| VI | **Machine Learning on Graphs:**<br>Using graph-based features in machine Learning, predicting relationships, graph embedding from graphs to matrices, Applications of Neo4j in web applications. | 7 |

| **Textbooks** | |
|---|---|
| 1 | Jonathan Gross and Jay Yellen, "Graph Theory and its Applications", 2 nd Edition, CRC Press. 2018. |
| 2 | Estelle Scifo, Hands-On Graph Analytics with Neo4j, Kindle Edition, 2020. |
| 3 | Bondy J.A. and Murty U.S.R., Graph Theory Ⅱ, Springer. 1 st Edition 2013. |

| **References** | |
|---|---|
| 1 | Bela Bollobas, Random Graphs, Cambridge University Press. 2008 |
| 2 | Douglas B. West ―Graph Theory, Prentice Hall. 2014 |
| | |

| **Useful Links** | |
|---|---|
| 1 | https://onlinecourses.nptel.ac.in/noc21_cs48/preview |

**CO-PO Mapping**

| | **Programme Outcomes (PO)** | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | | 2 | | | | |
| **CO2** | 1 | 2 | | | | |
| **CO3** | | 1 | 2 | | | |
| **CO4** | | | 2 | | | 1 |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.
ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.
For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem I/II |
| **Course Code** | 7DS519 |
| **Course Name** | Pattern Recognition |
| **Desired Requisites:** | |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To discuss theoretical aspects of features and pattern recognition |
| **2** | To introduce classification models and evaluation metrics |
| **3** | To relate real-world complex problems for engineering solutions with domain expertise |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Describe features, patterns and classification | II | Understanding |
| **CO2** | Apply data pre-processing for handling data anomalies and outliers | III | Applying |
| **CO3** | Identify complex patterns and trends leading to actionable insights | IV | Analyzing |
| **CO4** | Select pattern recognition for real time problem-solving and optimization | V | Evaluating |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Introduction to Pattern Recognition and Bayesian Theory:** Pattern recognition systems, The design cycle, Modeling using continuous and discrete features, Discriminant functions, The Gaussian density, Error estimation, Some basic examples | 7 |
| II | **Parametric Models:** Maximum-likelihood estimation, Bayesian estimation, Expectation-Maximization and mixture density estimation, Hidden Markov Models, Bayesian Belief Networks | 6 |
| III | **Non-parametric Methods and Feature Reduction:** Density estimation, Parzen windows estimation, Nearest neighbor estimation, Curse of dimensionality, Principal Component Analysis, Linear Discriminant Analysis, Feature selection | 7 |
| IV | **Non-Bayesian Classifiers and Clustering:** K-nearest neighbor classifier, Linear discriminant functions, Support vector machines, Neural networks, Decision trees, Random Forests, Criterion functions for clustering, k-means clustering, Hierarchical clustering, Graph-theoretic clustering, Cluster validity | 7 |

| | | |
|---|---|---|
| V | **Algorithm-Independent Learning Issues:**<br>No Free Lunch Theorem, Resampling for classifier design, Comparing classifiers-metrics, test , Combining classifiers-Bagging, Boosting etc.. | 6 |
| VI | **Structural and Syntactic Pattern Recognition:**<br>Recognition with strings, Grammatical methods- Context-Free Grammars (CFG), Stochastic Grammars, Attributed Grammars, Graph Grammars, Graph-theoretic methods- Graph Matching, Graph Isomorphism etc.. | 6 |

| **Textbooks** | |
|---|---|
| 1 | R. O. Duda, P. E. Hart, D. G. Stork, Pattern Classification, 2nd edition, John Wiley & Sons, Inc 2000 |
| 2 | C. M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press 1st edition 1995 |
| 3 | K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press 2nd edition1990. |

| **References** | |
|---|---|
| 1 | R. Schalkoff, Pattern Recognition: Statistical, Structural and Neural Approaches, John Wiley & Sons, Inc. 1992 |
| 2 | A. K. Jain, R. C. Dubes, Algorithms for Clustering Data, Prentice Hall 1988 |
| 3 | |

| **Useful Links** | |
|---|---|
| **1** | https://archive.nptel.ac.in/courses/117/105/117105101/ |

**CO-PO Mapping**

| | **Programme Outcomes (PO)** | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | | 2 | | | | |
| **CO2** | 1 | 3 | | | | |
| **CO3** | | 2 | | | | |
| **CO4** | | | 3 | | | 1 |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.
ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.
For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

| **Course Information** | | | | | |
|---|---|---|---|---|---|
| **Programme** | M.Tech. (Data Science) | | | | |
| **Class, Semester** | First Year M. Tech., Sem I/II | | | | |
| **Course Code** | 7DS520 | | | | |
| **Course Name** | Financial Data Science | | | | |
| **Desired Requisites:** | | | | | |
| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

| **Course Objectives** | |
|---|---|
| **1** | To understand and apply the knowledge of data science related applications in the domain of finance |
| **2** | To utilize statistical modeling and machine learning to analyze financial data and predict market trends. |
| **3** | To develop algorithms for risk management, portfolio optimization, and trading strategies in financial markets. |
| **4** | To apply advanced data science techniques to improve financial decision-making and asset management. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Understand forecasting in market movements and investment returns using data-driven models. | II | Understanding |
| **CO2** | Analyze risk assessment and mitigation models/algorithms through advanced analytics in financial data. | IV | Analyzing |
| **CO3** | Identify fraudulent activities and examine regulatory adherence in financial operations using various modeling techniques. | II/III | Applying/ Analyzing |
| **CO4** | Justify better investment decisions and portfolio management by leveraging insights from financial data science. | V | Evaluating |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Data Science basics:** Preparation, organizing, and visualization of financial market data and examination of basic properties of security prices , Quantitative models of risk-return framework in financial market. | 6 |
| II | **Modeling:** Linear and non-linear price dynamics and modeling of security prices, Stock market prediction modeling, portfolio optimization, and wealth market maximization | 7 |
| III | **Role of different parameters:** Role of latent factor and commonality models in the data science. Application of latent factor and commonality models in financial markets | 6 |

| | | |
|---|---|---|
| IV | **Financial Modeling:**<br>Modeling of financial market volatility using Conditional Heteroscedastic Models, Introduction to Crisis/Non-crisis models, Non-linearity, extreme-value modeling. | 7 |
| V | **Financial Modeling:**<br>Markov regime-switching models, Quantile regression, Contagion models Introduction to data modeling for high-frequency algorithmic trading. | 7 |
| VI | **Use cases:**<br>Use cases for application of data science in Finance: Investment Management, Sharpe ratio analysis, Capital Asset Pricing Model, etc. (using R/ python programming) | 6 |

| **Textbooks** | |
|---|---|
| 1 | Chris Brooks "Introductory Econometrics," Fourth Edition, Cambridge University Press 4th edition,2019 |
| 2 | Ruey S. Tsay "Analysis of Time-series data," Third Edition, Wiley 2014 |
| 3 | John Fox and Sanford Weisberg "An R Companion to Applied Regression," Third Edition, SAGE 2018 |
| **References** | |
| 1 | Yves Croissant and Giovanni Millo "Panel Data Econometrics with R," 1st Edition, Wiley,2018 |
| **Useful Links** | |
| 1 | https://nptel.ac.in/courses/111103126 |
| 2 | https://onlinecourses.nptel.ac.in/noc21_mg93/preview |
| 3 | https://www.codecademy.com/learn/paths/finance-python |

| **CO-PO Mapping** | | | | | | |
|---|---|---|---|---|---|---|
| | **Programme Outcomes (PO)** | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | | 2 | | | | |
| **CO2** | 1 | 3 | | | | |
| **CO3** | | | 2 | | | |
| **CO4** | | | 3 | | | 1 |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.
ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.
For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem I/II |
| **Course Code** | 7DS531 |
| **Course Name** | Social Data Analysis |
| **Desired Requisites:** | |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To introduce the basic notions used for social network analysis and analyze social media data to understand trends, sentiment, and user behavior. |
| **2** | To develop algorithms for network analysis, community detection, and influence measurement in social networks. |
| **3** | To implement natural language processing techniques to extract insights from textual data in social media |
| **4** | To apply data-driven approaches to study social phenomena, cultural trends, and public opinion dynamics. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Understand consumer preferences and market sentiment through social data analysis. | II | Understanding |
| **CO2** | Compare and classify social issues and public sentiment through data-driven research. | II | Understanding |
| **CO3** | Analyze marketing strategies and customer engagement based on social media insights. | IV | Analyzing |
| **CO4** | Apply and analyze solutions to crisis management and response strategies by monitoring and analyzing social media trends. | II/IV | Applying/ Analyzing |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Social Network Analysis:** Preliminaries and definitions, Erdos Number Project, Centrality measures, Balance and Homophily. Random graph models: Random graphs and alternative models. | 6 |
| II | **Network details:** Models of network growth, Navigation in social Networks, Network topology and diffusion, Contagion in Networks, Complex contagion, Percolation and information, Epidemics and information cascades. | 6 |
| III | **Network Structure:** Graph theory, Centrality, Clustering, Node-Edge Diagrams, Matrix representation, Visualizing online social networks, Visualizing social networks with matrix-based representations, Matrix and Node-Link Diagrams, Cohesive subgroups, Multidimensional Scaling, Structural equivalence. | 7 |

| | | |
|---|---|---|
| IV | **Network Connectivity:**<br>Roles and positions, Ego networks, Weak ties, Structural holes, Heavy tails, Small Diameter, Clustering of connectivity, The Erdos Renyi Model, Clustering Models, Preferential Attachment. | 7 |
| V | **Navigation in Networks:**<br>Navigation in Networks Revisited, Important vertices and page rank algorithm, Towards rational dynamics in networks, Basics of game theory. | 6 |
| VI | **Behavior of Network:**<br>Coloring and consensus biased voting, network formation games, network structure and equilibrium, behavioral experiments, Spatial and agent-based models. | 7 |

| **Textbooks** | |
|---|---|
| 1 | Wasserman, Stanley, and Joseph Galaskiewicz. "Advances in social network analysis: Research in the social and behavioral sciences", Sage, SAGE Focus Editions, 1994. |
| 2 | Knoke, David, and Song Yang. "Social network analysis" Sage Publications 3rd edition, 2019. |
| 3 | Tanmoy Chakraborty, "Social Network Analysis," Wiley, 1st edition 2021. |

| **References** | |
|---|---|
| 1 | Carrington, Peter J., John Scott, and Stanley Wasserman, eds. Models and methods in social network analysis. Vol. 28.Cambridge university press.2005 |
| 2 | Liu, Bing. "Social network analysis." In Web data mining, pp. 269-309. Springer, Berlin, Heidelberg, 1st edition 2011 |

| **Useful Links** | |
|---|---|
| 1 | https://onlinecourses.nptel.ac.in/noc22_cs117/preview |
| | |

| **CO-PO Mapping** | | | | | | |
|---|---|---|---|---|---|---|
| | **Programme Outcomes (PO)** | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | | 2 | | | | |
| **CO2** | | | 3 | | | |
| **CO3** | | 2 | | | | 1 |
| **CO4** | | 2 | | | | |
| The strength of mapping is to be written as 1: Low, 2: Medium, 3: High<br>Each CO of the course must map to at least one PO. | | | | | | |

| **Assessment** |
|---|
| The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.<br>ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6. For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing) |

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem I/II |
| **Course Code** | 7DS532 |
| **Course Name** | Data science for Businesses |
| **Desired Requisites:** | |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To understand and apply the fundamental concepts of data science and business applications of data mining and machine learning. |
| **2** | To utilize data science to extract actionable insights and Develop predictive models to forecast market trends, customer behavior, and operational outcomes. |
| **3** | To implement data-driven strategies for optimizing marketing campaigns, customer acquisition, and retention. |
| **4** | To apply data analytics to enhance operational efficiency, resource allocation, and cost management in businesses |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Describe various data-driven business strategies. | II | Understanding |
| **CO2** | Estimate customer satisfaction and experiences based on data insights of Business Research (BR). | IV | Analyzing |
| **CO3** | Analyze operational efficiency and productivity through optimized predictive processes. | IV | Analyzing |
| **CO4** | Categorize data science techniques to drive business transformation and strategically assess model performance. | III/IV | Applying/ Analyzing |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Introduction:**<br>Introduction, Introduction to Data Analytics, Using Data Science Tools for Business Analytics, Concepts and Practices of business, Challenges- Technical Challenges-data quality, Quantity, integration etc.. Data Analytics. | 6 |
| II | **Nature and scope of Business Research (BR)**:<br>Types of businesses, Business Challenges: Internal ,external, Fulfilling customer needs, Recruitment, Retainment of Employees, Managing workflow, Achieving targets, Handling market Rivalry, Increasing sales etc Role of BR in decision making. Management decision problem Vs Business Research Objective | 7 |
| III | **Modeling:**<br>Intro to Predictive Modeling, Supervised Segmentation, Conceptual Predictive Analytics, From Correlation to Supervised Segmentation, Models, Induction, and | 6 |

| | | |
|---|---|---|
| | Prediction, Supervised Segmentation, Selecting Informative Attribute. Problem formulation, Fitting the data, Other Modeling methods. | |
| IV | **Model performance:**<br>Machine learning for financial stability, Model performance analytics & the science of predictive modeling Over fitting the data. Holdout testing, cross-validation and learning curves, domain knowledge validation | 7 |
| V | **Model performance analysis** :<br>Fundamental concepts, optimal model parameters based on data, Choosing the goal for data mining, Objective functions, Loss function, Exemplary techniques: Linear regression; Logistic regression; Support-vector machines etc. | 7 |
| VI | **Methods for data analysis:**<br>Methods for hierarchical data analysis , Integrating Multiple Evidence Sources for Accurate Predictions, reasoning, Joint Probability and independence Applying Bayes rules and other methods to data science<br>**Case Study**: Modeling consumer behavior for targeted marketing (banking and/or online advertising) | 6 |

| **Textbooks** | |
|---|---|
| 1 | Foster Provost, Tom Fawcett "Data Science for Business: What you need to know about data mining and data analytic thinking" O'Reilly Media, Inc. ISBN: 9781449361327, 1 st edition (O'Reilly, 2013) Foster's new update (as of 2020). |
| 2 | Probyto Data Science and Consulting Pvt. Ltd. "Data Science for Business Professionals: A Practical Guide for Beginners " (English Edition) 1st Edition,2020 |
| | |
| **References** | |
| 1 | Sergio Consoli ,Diego Reforgiato Recupero ,Michaela Saisana" Data Science for Economics and Finance-Methodologies and Applications" ISBN 978-3-030-66891-4 (eBook) https://doi.org/10.1007/978-3-030-66891-4, Kindle edition 2021 |
| | |
| **Useful Links** | |
| **1** | https://www.udacity.com/course/data-science-for-business-leaders--nd045 |

| **CO-PO Mapping** | | | | | | |
|---|---|---|---|---|---|---|
| | **Programme Outcomes (PO)** | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | | 2 | | | | |
| **CO2** | 1 | | 3 | | | |
| **CO3** | | | 3 | | | |
| **CO4** | | 2 | 1 | | | 1 |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.
ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.
For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

# Walchand College of Engineering, Sangli

*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem I/II |
| **Course Code** | 7DS533 |
| **Course Name** | Game theory |
| **Desired Requisites:** | |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Lecture** | 3 Hrs/week | **ISE** | **MSE** | **ESE** | **Total** |
| **Tutorial** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To understand algorithmic game theory and its applications using AI and machine learning techniques |
| **2** | To analyze strategic interactions and decision-making among rational agents in competitive scenarios. |
| **3** | To develop mathematical models to study equilibrium outcomes and optimal strategies in game theory. |
| **4** | To apply game-theoretic principles to understand behavior in economics, political science, and evolutionary biology. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Understand competitive dynamics and strategic including Nash equilibrium and dominant strategy equilibrium. | II | Understanding |
| **CO2** | Apply conflict resolution strategies and mixed strategy Nash equilibrium to analyze and predict outcomes in strategic interactions. | III | Applying |
| **CO3** | Analyze resource allocation and decision-making process through game-theoretic frameworks/theorems. | IV | Analyzing |
| **CO4** | Develop policies and mechanisms to promote cooperation and stability in competitive game environments. | VI | Creating |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Introduction:** Introduction to Game Theory, Introduction to Graph Strategy ,Dominant Strategy Equilibrium, Pure Strategy Nash Equilibrium, computing Nash equilibrium | 6 |
| II | **Graph Strategy:** Evidence on expected payoff functions, Strategic games, Mixed Strategy Nash Equilibrium, Max-min and Min-max Values, Dominated actions, Matrix Games. | 7 |
| III | **Correlated Strategy:** Correlated Strategies and Correlated Equilibrium, Nash Bargaining Problem, players division for a resource or surplus optimally. Nash bargaining solution, axiomatic approaches, Coalitional Games with Transferable Utility. | 7 |

| | | |
|---|---|---|
| IV | **Learning in game theory:**<br>The Core-conditions for non-emptiness and stability, Shapley Value-Formula and examples, Nucleolus- Integrative methods and examples. Sequential learning in games-Fictitious play, Bayesian learning, multi-agent learning using game theory | 6 |
| V | **Theorems:**<br>Introduction to Mechanism Design, Introduction to following theorems along with equation and examples-Arrows Impossibility theorem, Gibbard- Satterthwaite Theorem, Mechanisms with Money. | 7 |
| VI | **Other Mechanisms:**<br>Introduction to Vickrey-Clarke-Groves Mechanisms (VCG) Mechanism. Theory of Myerson's Lemma and VCG Mechanism details, The Groves' Theorem, Groves Mechanisms and Budget Balance, VCG examples. | 6 |

| **Textbooks** | |
|---|---|
| 1 | Martin J. Osborne "An Introduction to Game Theory," First Edition, Oxford University Press.2003 |
| 2 | Y. Narahari "Game theory and mechanism design," First Edition, World Scientific. 2014 |
| 3 | Noam Nisan, Tim Roughgarden, Éva Tardos, Vijay V. Vazirani. "Algorithmic Game Theory," First Edition, Cambridge University Press, 2007 |

| **References** | |
|---|---|
| 1 | Ivan Pastine, Tuvana Pastine, and Tom Humberstone "Introducing Game Theory: A Graphic Guide," First Edition, Icon Books Ltd, 2017 |
| 2 | Michael Maschler, Eilon Solan, Shmuel Zamir "Game Theory," Second Edition, Cambridge University Press, 2020 |

| **Useful Links** | |
|---|---|
| **1** | https://onlinecourses.nptel.ac.in/noc19_ge32/preview |
| **2** | https://onlinecourses.nptel.ac.in/noc22_cs77/preview |
| **3** | https://www.cse.iitb.ac.in/~swaprava/courses/cs711/lecnotes.pdf |

**CO-PO Mapping**

| | **Programme Outcomes (PO)** | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | | 2 | | | | |
| **CO2** | | | 3 | | | |
| **CO3** | | 1 | 2 | | | 1 |
| **CO4** | | | 3 | | | |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3. ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.
ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.
For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)

# Open Elective

**Walchand College of Engineering, Sangli**
*(Government Aided Autonomous Institute)*

**AY 2024-25**

**Course Information**

| | |
|---|---|
| **Programme** | M.Tech. (Data Science) |
| **Class, Semester** | First Year M. Tech., Sem II |
| **Course Code** | |
| **Course Name** | Data Science for Engineers |
| **Desired Requisites:** | |

| **Teaching Scheme** | | **Examination Scheme (Marks)** | | | |
|---|---|---|---|---|---|
| **Practical** | 3 Hrs/Week | **ISE** | **MSE** | **ESE** | **Total** |
| **Interaction** | - | 20 | 30 | 50 | 100 |
| | | **Credits: 3** | | | |

**Course Objectives**

| | |
|---|---|
| **1** | To get acquaint with concepts in Machine Learning (ML). |
| **2** | To apprehend the recent trends in Data Science |
| **3** | To make able to understand the applications in Data Science |
| **4** | To implement python code and add visualization using various libraries. |

**Course Outcomes (CO) with Bloom's Taxonomy Level**

At the end of the course, the students will be able to,

| **CO** | **Course Outcome Statement/s** | **Bloom's Taxonomy Level** | **Bloom's Taxonomy Description** |
|---|---|---|---|
| **CO1** | Understand the mathematical foundation required for data science. | II | Understanding |
| **CO2** | Apply the first level data science algorithms to solve problems. | II | Applying |
| **CO3** | Evaluate data science problem-solving algorithms and frameworks through a practical case study. | III/V | Analyzing /Evaluating |
| **CO4** | Construct several types of plots using various libraries of python. | VI | Creating |

| **Module** | **Module Contents** | **Hours** |
|---|---|---|
| I | **Basics of Python:**<br>Tools required for Data Science, Introduction to Spyder, setting working directory, creating and saving a script file, file execution, clearing console, removing variables from environment, clearing environment, commenting script files, variable creation, arithmetic and logical operations, data types. | 6 |
| II | **Data types, Control structures and Libraries**:<br>Strings, lists, arrays, tuples, dictionary sets, range, Reading files, exploratory data analysis, data preparation and preprocessing, If-else family, for loop, for loop with if break, while loop and functions, Numpy, Pandas | 7 |
| III | **Data Visualization** :<br>Data Visualization using Matplotlib and Seaborn libraries.Scatter plot, line plot, bar plot, histogram, box plot, pie chart, pair plot | 6 |

| | | |
|---|---|---|
| IV | **Unsupervised Learning:**<br>Why data reduction?, key idea behind PCA, linear algebra behind PCA, PCA in practice, clustering algorithm in practice, case study of k-means algorithm | 6 |
| V | **Interactive Python dashboards with Plotly :**<br>Ploty Basic - scatter plot, bar plot, bubble plot, box plot, histograms, heat maps, dashboard components, interactive components in dashboard | 7 |
| VI | **Case Studies:**<br>Regression and Classification (Use of any case study using a dataset),Regression Datasets : Crime_in_India, Salary_Classification, Income_Data, Classification Datasets - Shopping_Mall, Social_Network_Ads | 7 |

| **Textbooks** | |
|---|---|
| 1 | R. Nageswara Rao, ―"Core Python Programming‖," Dreamtech Press, 2nd Edition, 2017 |
| 2 | Chun, J Wesley, ―"Core Python Programming"‖, Pearson, 2nd Edition, 2007 Reprint 2010 |
| 3 | Douglas Montgomery- "Applied statistics and probability for engineers", Wily, Pearson, 6th Edition, 2016 |
| 5 | Samir Madhavan -Mastering Python for data science- PACKT,1st edition 2015 |

| **References** | |
|---|---|
| 1 | Scikit-Learn User Guide, Release 0.23.1, scikitlearn developers, May 19,2020 |
| 2 | Python 3.x Documentation |
| 3 | Gilbert Strang- Introduction to linear algebra ,Pearson, 6th Edition, 2017 |

| **Useful Links** | |
|---|---|
| 1 | https://onlinecourses.nptel.ac.in/noc19_mg47/preview |
| 2 | https://docs.python.org/3/tutorial/ |
| 3 | https://www.learnpython.org/ |
| 4 | https://www.hackerrank.com/ |

**CO-PO Mapping**

| | **Programme Outcomes (PO)** | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| **CO1** | | 2 | | | | |
| **CO2** | 2 | | 1 | | | |
| **CO3** | | 1 | 3 | | | 1 |
| **CO4** | | 2 | 1 | | | |

The strength of mapping is to be written as 1: Low, 2: Medium, 3: High
Each CO of the course must map to at least one PO.

**Assessment**

The assessment is based on MSE, ISE and ESE. MSE shall be typically on modules 1 to 3.
ISE shall be taken throughout the semester in the form of teacher's assessment. Mode of assessment can be field visit, assignments etc. and is expected to map at least one higher order PO.
ESE shall be on all modules with around 40% weightage on modules 1 to 3 and 60% weightage on modules 4 to 6.
For passing a theory course, Min. 40% marks in (MSE+ISE+ESE) are needed and Min. 40% marks in ESE are needed. (ESE shall be a separate head of passing)