Shirpur Education Society's

R. C. Patel Institute of Technology, Shirpur

(An Autonomous Institute)

Syllabus Booklet

B. Tech. Computer Science and Engineering
(Data Science)

With effect from Year 2021-22

Shahada Road, Near Nimzari Naka, Shirpur, Maharashtra 425405
Ph: 02563 259802, Website: www.rcpit.ac.in

Semester-III(w.e.f. 2021-22)													
Sr	Course Category	Course Code	Course Title	Teaching Scheme			Evaluation Scheme					Total	Credit
							Continuous Assessment (CA)				ESE		
							TA	Term Test 1 (TT1)	Term Test 2 (TT2)	Average (TT1 & TT2)			
				L	T	P	[A]			[B]	[C]	[A+B+C]	
1	BS	BSCS3010T	Mathematics for Intelligent Systems	3	1		20	15	15	15	65	100	4
2	PC1	PCCS3020T	Data Structures and Algorithms	3			20	15	15	15	65	100	3
3	PC1L	PCCS3020L	Data Structures and Algorithms Laboratory			2	25				25	50	1
4	PC2	PCCS3030T	Foundations of Data Analysis	3			20	15	15	15	65	100	3
5	PC2L	PCCS3030L	Foundations of Data Analysis Laboratory			2	25				25	50	1
6	PC3	PCCS3040T	Database Management Systems	3			20	15	15	15	65	100	3
7	PC3L	PCCS3040L	Database Management Systems Laboratory			2	25				25	50	1
8	PC4	PCCS3050T	Statistics for Data Science	3			20	15	15	15	65	100	3
9	PC4L	PCCS3050L	Statistics for Data Science Laboratory			2	25				25	50	1
10	PC5L	PCCS3060L	Programming with Python Laboratory			2	25				25	50	1
11	PJ	PJCS3070L	Semester Project-I			2	25				25	50	1
12	MC	MCCS3080T	Constitution of India	1									Audit Course
13			Field/Internship/Industry Training#										Audit Course
			Total	16	1	12	250			75	475	800	22
# Minimum 6 weeks internship should be done during winter/summer vacation of semester III to VI. Report to be submitted in Semester VII													

Mathematics for Intelligent Systems (BSCS3010T)

Teaching Scheme	**Examination Scheme**
Lectures : 03 Hrs./week	Term Test : 15 Marks
Tutorial : 01 Hr/week	Teacher Assessment : 20 Marks
Credits : 04	End Sem Exam : 65 Marks
	Total Marks : 100 Marks

Pre-requisites: Concepts of basic matrices, partial derivatives and basic probability.

Course Objectives:

To build the strong foundation in learners of mathematics needed for building concepts of machine learning.

CO	Course Outcomes	Blooms Level	Blooms Description
CO1	Analyze probability of random variables and probability distributions.	L4	Analyze
CO2	Demonstrate knowledge of linear algebra.	L3	Apply
CO3	Apply concepts of matrix theory.	L3	Apply
CO4	Demonstrate concepts of calculus.	L3	Apply
CO5	Analyze different optimization techniques.	L4	Analyze

Course Contents

Unit-I Probability, Random Variables and Probability Distributions 10 Hrs.

Probability: Conditional Probability, Mutually and Pair Wise Independent Events, Bayes' Theorem

Random Variables: Discrete Random Variable, Probability Mass Function, Discrete Distribution Function, Continuous Random Variable, Probability Density Function, Continuous Distribution Function, Mathematical Expectation, Moment Generating Function, Two-Dimensional Random Variable and its Joint Probability Mass and Density Function, Marginal Distribution Function, Conditional Distribution Functions, Covariance, Joint Moments.

Probability Distributions: Discrete Probability Distribution: Binomial Distribution, Poisson Distribution, Hypergeometric Distribution.

Continuous Probability Distribution: Uniform Distribution, Exponential Distribution, Normal Distribution, Beta Distribution, Gamma Distribution, Central Limit Theorem.

Unit-II Linear Algebra 08 Hrs.

Vectors in N-Dimensional Vector Space, Properties, Dot Product, Cross Product, Norm and Distance,Vector Spaces over Real Field, Properties of Vector Spaces over Real Field, Subspaces, Linear Independence and Dependence of Vectors, Span of Vectors, Basis of a Vector Space, Dimension of a Vector Space, Cauchy Schwarz Inequality, Linear Transformation, Norms and Spaces, Orthogonal Compliments and Projection Operator, Kernel Hilbert Spaces.

Unit-III Matrix Theory 08 Hrs.

Characteristic Equation, Eigen Values and Eigen Vectors, Properties of Eigen Values and Eigen Vectors, Cayley-Hamilton Theorem, Examples Based on Verification of Cayley Hamilton Theorem, Similarity of Matrices, Diagonalization of Matrices, Functions of Square Matrix, Derogatory and Non-derogatory Matrices, Least Squared and Minimum Normed Solutions.

Unit-IV Calculus 04 Hrs.

Gradient, Directional Derivatives, Jacobian, Hessian, Convex Sets, Convex Functions and its Properties.

Unit-V Optimization 12 Hrs.

Unconstrained and Constrained Optimization, Convergence.

Unconstrained Optimization Techniques: Newton's Method, Quasi Newton Method.

Constrained Optimization Techniques: Gradient Descent, Stochastic Gradient Descent, Penalty Function Method, Lagrange Multiplier Method, Karush–Kuhn–Tucker Method, Simplex Method,

Penalty and Duality, Dual Simplex Method, Downhill Simplex Method.

List of tutorials: (any 8)

1. To solve numerical on discrete probability distributions.
2. To solve numerical on continuous probability distributions.
3. To solve numerical on vector spaces (basis and dimension).
4. To solve numerical on cauchy-schwarz inequality and linear transformation.
5. To solve numerical on diagonalizability using eigenvalues and eigenvectors.
6. To solve numerical on minimal polynomial and functions of a matrix.
7. To solve numerical on calculus.
8. To solve numerical on Gradient descent and Lagrange's multiplier method.
9. To solve numerical on KKT method.
10. To solve numerical on all forms of simplex method.

Any other tutorial based on syllabus may be included which would help the learner to understand topic/concept.

Text Books:

1. Dr. B. S. Grewal, Higher Engineering Mathematics, 44^{th} Edition, Khanna Publication, 1965.
2. Kanti B. Datta, Mathematical Methods in Science and Engineering, 1^{st} Edition, Cengage Learning India, 2011.
3. Hamdy A. Taha, Operations Research - An Introduction, Pearson, 10^{th} Edition, 2010.
4. Kanti Swarup, P. K. Gupta, Mohan Man, Operations Research, 2020 Edition, S Chand Publication, 2005.

Reference Books:

1. W. Cheney, Analysis for Applied Mathematics, 1^{st} Edition, New York: Springer Science+Business Media, 2001.
2. S. Axler, Linear Algebra Done Right, 3^{rd} Edition, Springer International Publishing, 2015.
3. J. Nocedal and S. J. Wright, Numerical Optimization, 2^{nd} Edition, New York: Springer Science+Business Media, 2006.

4. J. S. Rosenthal, A First Look at Rigorous Probability Theory, 2^{nd} Edition, Singapore: World Scientific Publishing, 2006.

5. Seymour Lipschutz and Marc Lipson, Linear Algebra Schaum's outline series, 4^{th} Edition, Mc-Graw Hill Publication, 2009.

6. Erwin Kreysizg, John Wiley & Sons, Inc, Advanced Engineering Mathematics, 10^{th} Edition, 2000.

Evaluation Scheme:

Theory :

Continuous Assessment (A):

Subject teacher will declare Teacher Assessment criteria at the start of semester.

Continuous Assessment (B):

1. Two term tests of 15 marks each will be conducted during the semester.

2. Average of the marks scored in both the tests will be considered for final grading.

End Semester Examination (C):

1. Question paper based on the entire syllabus, summing up to 65 marks.

2. Total duration allotted for writing the paper is 3 hrs.

Data Structures and Algorithms (PCCS3020T)

Teaching Scheme
Lectures : 03 Hrs./week
Credits : 03

Examination Scheme
Term Test : 15 Marks
Teacher Assessment : 20 Marks
End Sem Exam : 65 Marks
Total Marks : 100 Marks

Prerequisite: Computer Programming (C Programming)

Course Objectives:

The course intends to introduce and familiarize students with data structures, their use in solving real time complex problems and implementation of these data structures. The course also aims to provide mathematical approach for analyzing algorithms using asymptotic notation and for measuring efficiency of algorithms. Finally, the course intends to make students learn various sorting and searching techniques and choose efficient one based on their efficiency.

CO	Course Outcomes	Blooms Level	Blooms Description
CO1	Make use of various Operations like Searching, Insertion, Deletion, Traversal etc. on various Data Structures.	L3	Apply
CO2	Choose Appropriate (Efficient) Sorting, Searching and Hashing Technique for given Problem and Implement it.	L1, L3	Remember, Apply
CO3	Choose Appropriate (Efficient) Data Structure and Algorithm and Apply them to Solve Specified Problems.	L1, L3	Remember, Apply
CO4	Evaluate and Analyze the Efficiency of Algorithms based on Time and Space Complexity.	L4, L5	Evaluate, Analyze
CO5	Formulate New Solutions for given Problems or Improve Existing one for Better Efficiency and Optimization.	L6	Create

Course Contents

Unit-I Review and Introduction 06 Hrs.

Review: Pointers, Structures, Function, Recursion. **Introduction to Data Structures**: Need of Data Structures, Types of Data Structures, Abstract Data Type (ADT). **Introduction to Algorithms and Analysis**: Need of Writing Algorithm, SDLC (System Development Life Cycle) and role of algorithms, Asymptotic Notation (Big-Oh, Big Omega, Theta Notations), Order of Growth Functions, Complexity Analysis Techniques, Few examples of analysis of algorithms (like Fibonacci, prefix average, etc.)

Unit-II Linked Lists 06 Hrs.

Basic Concept of Linked List, Comparison of Sequential (Array-based) and Linked Organizations, Dynamic Memory Management, ADT of Linked List, Singly Linked List, Doubly Linked List, Circular Linked List, various basic and Advanced Operations on Linked List (Insertion, Deletion, Merge, Traversal, Copy, Reverse etc.) and their Analysis, Applications of Linked Lists.

Unit-III Stack and Queue 08 Hrs.

Stacks: Introduction to Stack, Stack as an ADT, Stack ADT Implementation using Array and Linked List with respective Analysis and Comparison, Applications of Stacks: Expression Conversion (Infix to Prefix and Postfix) and Evaluation(Postfix Expression Evaluation), Parenthesis Correctness etc. **Queues**: Introduction to Queue, Queue as an ADT, Queue ADT Implementation using Array and Linked List with respective Analysis and Comparison, Linear Queue, Circular Queue, Priority Queue: Heap based Implementation, Deques, Applications of Queues.

Unit-IV Trees 08 Hrs.

Introduction to Trees, Basic Terminology, Types of Trees, Binary Tree Representation, Traversal of Binary Tree, Expression Tree, Binary Search Tree, Operations on Binary Search Tree and their Analysis, AVL Tree, Applications of Trees.

Unit-V Graphs 06 Hrs.

Representation of Graph, Types of Graph, Breadth-First Search (BFS), Depth–First Search (DFS), Minimum Spanning Tree: Prim's & Kruskal's Algorithm, Applications of Graphs.

Unit-VI Sorting and Searching Techniques 08 Hrs.

Sorting: Bubble Sort, Selection Sort, Insertion Sort, Quick Sort, Merge Sort, Heap Sort, Radix sort. Analysis of Sorting Techniques. **Searching**: Linear Search, Binary Search, Hashing Techniques and Collision Resolution Techniques, Linear Hashing, Hashing with Chaining, Separate Chaining, Open

Addressing, Rehashing, Analysis of Searching Techniques.

Text Books:

1. R. F. Gilberg and B. A. Forouzan, Data Structures – A Pseudocode Approach with C, 2^{nd} Edition,Cengage Learning, 2005.
2. Ellis Horowitz, Sartaj Sahni and Susan Anderson-Freed, Fundamentals of Data Structures in C, 2^{nd} Edition, W. H. Freeman and Company, 2008.

Reference Books:

1. Mark A. Weiss, Data Structures and Algorithm Analysis in C, 4^{th} Edition, Pearson, 2014.
2. M. T. Goodrich, R. Tamassia, D. Mount, Data Structures and Algorithms in C++, Wiley, 2004.
3. Tenenbaum, Langsam, Augenstein, Data Structures using C, Pearson, 2004.
4. Aho, Hopcroft, Ullman, Data Structures and Algorithms, Addison-Wesley, 2010.
5. Reema Thareja, Data Structures using C, Oxford, 2017.

Evaluation Scheme:

Theory :

Continuous Assessment (A):

Subject teacher will declare Teacher Assessment criteria at the start of semester.

Continuous Assessment (B):

1. Two term tests of 15 marks each will be conducted during the semester.
2. Average of the marks scored in both the tests will be considered for final grading.

End Semester Examination (C):

1. Question paper based on the entire syllabus, summing up to 65 marks.
2. Total duration allotted for writing the paper is 3 hrs.

Data Structures and Algorithms Laboratory (PCCS3020L)

Practical Scheme
Practical : 02 Hrs./week
Credit : 01

Examination Scheme
Teacher Assessment : 25 Marks
End Sem Exam : 25 Marks
Total : 50 Marks

Course Objectives:

The course intends to introduce and familiarize students with data structures, their use in solving real time complex problems and implementation of these data structures. The course also aims to provide mathematical approach for analyzing algorithms using asymptotic notation and for measuring efficiency of algorithms. Finally, the course intends to make students learn various sorting and searching techniques and choose efficient one based on their efficiency.

CO	Course Outcomes	Blooms Level	Blooms Description
CO1	Make use of Different Searching and Sorting Operations.	L3	Apply
CO2	Examine Different Operations on Stack and Queue Data Structure.	L4	Analyze
CO3	Experiment with Single Linked List and Perform Various Operations.	L3	Apply
CO4	To Construct Various Hashing Techniques.	L3	Apply
CO5	To Construct Tree Data Structure.	L6	Create

List of Laboratory Experiments(At Least 12)

Suggested Experiments:

Note: Students are required to complete 12 experiments. At least one experiment is mandatory from each topic .

Any other experiment based on syllabus may be included, which would help the learner to understand topic/concept.

- Recursion
  - Implementation of Recursive Algorithms to Solve Various Fundamental Problems like: Addition of elements in an Array, Reversing an Array, Adding all digits of a given Numeral, Prefix Average, Factorial of a given number, Fibonacci Sequence etc.
- Sorting and Searching
  - Implementation of Insertion Sort, Selection Sort Menu Driven Program.
  - Implementation of Quick Sort.
  - Implementation of Merge Sort.
  - Implementation of Heap Sort.
  - Implementation of Binary Search.
  - Implementation of Hashing Functions with Different Collision Resolution Techniques.
- Linked List
  - Implementation of Linked Lists Menu Driven Program.
  - Implementation of different operations on Linked List: Copy, Concatenate, Split, Reverse, Count number of Nodes etc.
  - Implementation of Polynomial Operations (Addition, Subtraction) using Linked List.
- Stack and Queue
  - Implementation of Infix to Postfix Transformation and its Evaluation Program.
  - Implementation of Infix to Prefix Transformation and its Evaluation Program.
  - Implementation of Double Ended Queue Menu Driven Program.
  - Implementation of Queue Menu Driven Program.
  - Implementation of Circular Queue Menu Driven Program.
  - Implementation of Priority Queue Program using Array.
  - Implementations of Linked Lists Menu Driven Program (Stack and Queue).

- Implementations of Double Ended Queue using Linked Lists.
- Implementation of Priority Queue program using Heap.

- Trees
  - Implementation of BT (Binary Tree) Program.
  - Implementation of BST Program.
  - Implementation of Various Operations on Tree like: Copying Tree, Mirroring a Tree, Counting the Number of Nodes in the Tree, Counting only Leaf Nodes in the Tree.
  - Implementation of Construction of Expression Tree using Postfix Expression.

- Graphs
  - Implementation of Graph Menu Driven Program (DFS & BFS).

Evaluation Scheme:

Laboratory:

Continuous Assessment (TA):

Laboratory work will be based on PCCS3020T with minimum 12 experiments to be incorporated. The distribution of marks for term work shall be as follows:

1. Performance in Experiments: 05 Marks
2. Journal Submission: 05 Marks
3. Viva-voce: 05 Marks
4. Subject Specific Lab Assignment/Case Study: 10 Marks

The final certification and acceptance of term work will be subject to satisfactory performance of laboratory work and upon fulfilling minimum passing criteria in the term work.

End Semester Examination (ESE):

Oral/ Practical examination will be based on the entire syllabus including, the practicals performed during laboratory sessions.

Foundations of Data Analysis(PCCS3030T)

Teaching Scheme
Lectures : 03 Hrs./week
Credits : 03

Examination Scheme
Term Test : 15 Marks
Teacher Assessment : 20 Marks
End Sem Exam : 65 Marks
Total Marks : 100 Marks

Prerequisite: Basic Mathematics

Course Objectives:

To develop skills of data analysis techniques for data modelling.

CO	Course Outcomes	Blooms Level	Blooms Description
CO1	Identify visualization techniques to understand Data.	L3	Apply
CO2	Make use of ETL and perform OLAP operation.	L3	Apply
CO3	Perform various techniques to improve quality of data.	L6	Create
CO4	Choose appropriate feature engineering technique to prepare data for modelling.	L3	Apply
CO5	Make use of sampling techniques to sample data for modelling.	L3	Apply

Course Contents

Unit-I Data 06 Hrs.

Data Objects and Attributes: Nominal, Binary, Ordinal, Numeric, Discrete, Continuous. Characteristics of Data Sets: Dimensionality, Sparsity, Resolution. Types of Data Sets: Record Data, Data Matrix, Graph-based Data, Sequential Data, Sequence Data, Time Series Data, Spatial Data.

Data visualization: Temporal: Scatter Plots, Time Series Sequences, Line Graphs; Hierarchical: Tree Diagrams, Ring Charts; Network: Matrix Charts, Node-link Diagrams, Word Clouds, Alluvial Diagrams; Multidimensional: Pie Chart, Venn Diagrams, Stacked Bar Graph, Histograms; Geospatial: Flow Map, Density Map, Heat Maps.

Unit-II ETL Process and OLAP 08 Hrs.

Major steps in ETL Process, Data Extraction: Techniques, Data Transformation: Basic Tasks, Major transformation types, Data Loading: Applying Data, OLTP Vs OLAP, OLAP definition, Dimensional Analysis, Hypercubes.

OLAP Operations: Drill down, Roll up, Slice, Dice and Rotation, OLAP models: MOLAP, ROLAP.

Unit-III Data Preprocessing 10 Hrs.

Data Quality: Measurement Error, Data Collection Error, Noise, Artifacts, Precision, Bias, Accuracy, Outliers, Missing Values, Inconsistent Values, Duplicate Values.

Data Cleaning: Handling Missing Values and Noisy Data.

Data Transformation: Smoothing, Attribute Construction, Aggregation, Normalization.

Data Discretization: Binning, Histogram analysis, Clustering.

textbfOutlier Detection: Types of Outliers, Challenges, Statistical Method, Proximity-based Method, Clustering-based Method.

Unit-IV Feature Engineering 10 Hrs

Curse of Dimensionality, **Feature Selection**: Univariate methods (Pearson Correlation, F-Score, Chi-Square, Signal to Noise Ratio) and Multivariate methods (Forward Selection, Backward Selection and Stepwise Selection), **Feature Extraction**: Principal Component Analysis.

Unit-V Elementary Sampling Theory 08 Hrs.

Census and Sampling Survey, Steps in Sampling Design, Criteria of selecting a good sample procedure, Characteristics of a good Sample design, Types of sample design: Non Probability and Probability Sampling, Complex Random Sampling Design: Symmetric Sampling, Stratified Sampling, Cluster Sampling, Area Sampling, Sequential Sampling and Multi-stage Sampling.

Text Books:

1. Jason Browniee, Data Preparation for Machine Learning, Machine Learning Mastery.
2. Jason Osborne, Best Practices in Data Cleaning: A Complete Guide to Everything you Need to Do Before and After Collecting Your Data, Sage Publication, 2012.
3. Q. Ethan McCallum, Bad Data Handbook, O'Reilly, 2012.
4. Max Kuhn and Kjell Johnson, Feature Engineering and Selection: A Practical Approach for Predictive Models, CRC Press, 2020.

Reference Books:

1. Jeffrey Shaffer, Steve Wexler, Andy Cotgreave, The Big Book of Dashboards: Visualizing your Data using Real-World Business Scenarios, Wiley 2017.
2. C. R. Kothari, Research Methodology-Methods and Techniques, 2^{nd} Edition, New Age International.
3. S. C. Gupta and V. K. Kapoor, Fundamentals of Mathematical Statistics, 12^{th} Edition, Sultan Chand Publisher.
4. Paulraj Ponniah, Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals, 2^{nd} Edition, Wiley.
5. Rayan Sleeper, Practical Tableau, O'Reilly 2018.
6. Han, Kamber, Morgan Kaufmann, Data Mining Concepts and Techniques, 3^{rd} Edition.
7. Wes McKinney, Python for Data Analysis, 2^{nd} Edition, O'Reilly, 2018.

Evaluation Scheme:

Theory :

Continuous Assessment (A):

Subject teacher will declare Teacher Assessment criteria at the start of semester.

Continuous Assessment (B):

1. Two term tests of 15 marks each will be conducted during the semester.
2. Average of the marks scored in both the tests will be considered for final grading.

End Semester Examination (C):

1. Question paper based on the entire syllabus, summing up to 65 marks.
2. Total duration allotted for writing the paper is 3 hrs.

Foundations of Data Analysis Laboratory (PCCS3030L)

Practical Scheme
Practical : 02 Hrs./week
Credit : 01

Examination Scheme
Teacher Assessment : 25 Marks
End Sem Exam : 25 Marks
Total : 50 Marks

Course Objectives:

To analyze and visualize given data using various data analysis strategies.

CO	Course Outcomes	Blooms Level	Blooms Description
CO1	Utilize data preprocessing techniques on dataset.	L3	Apply
CO2	Perform basics data analysis strategies.	L6	Create
CO3	Perform basic mathematical tactics to analyze data.	L6	Create
CO4	Inspect large size datasets.	L4	Analyze
CO5	Predict suitable techniques to visualize the data.	L6	Create

List of Laboratory Experiments(At Least 8)

Suggested Experiments:(At least 8 experiments)

Visualization experiments can be performed using Tableau and Data Preprocessing experiments can be performed using Python/R.

- Create new measures on a given dataset and visualize them using a bar graph.
- Perform time series aggregation, apply filters on a given dataset, create line and area charts.
- Apply maps, scatter plots on a given dataset and create a dashboard.
- Perform joins, blends and create dual axis chart.
- Perform table calculations, bins, distributions and create Heat maps.
- Create an interactive data story.
- Perform Exploratory Data Analysis on a given dataset.
- Perform Data cleaning on a given dataset.
- Perform necessary Data Transformation on a given dataset.
- Perform correlation analysis on a given dataset.
- Perform dimensionality reduction using PCA.

Any other experiment based on syllabus may be included, which would help learners to understand the topic/concept.

Evaluation Scheme:

Laboratory:

Continuous Assessment (TA):

Laboratory work will be based on PCCS3030T with minimum 08 experiments to be incorporated. The distribution of marks for term work shall be as follows:

1. Performance in Experiments: 05 Marks
2. Journal Submission: 05 Marks
3. Viva-voce: 05 Marks
4. Subject Specific Lab Assignment/Case Study: 10 Marks

The final certification and acceptance of term work will be subject to satisfactory performance of laboratory work and upon fulfilling minimum passing criteria in the term work.

End Semester Examination (ESE):

Oral / Practical examination will be based on the entire syllabus including, the practicals performed during laboratory sessions.

Database Management Systems (PCCS3040T)

Teaching Scheme
Lectures : 03 Hrs./week
Credits : 03

Examination Scheme
Term Test : 15 Marks
Teacher Assessment : 20 Marks
End Sem Exam : 65 Marks
Total Marks : 100 Marks

Course Objectives:

The course intends to introduce the students to the management of database systems, with an emphasis on how to design, organize, maintain and retrieve information efficiently and effectively from a database.

CO	Course Outcomes	Blooms Level	Blooms Description
CO1	Design an optimized database.	L6	Create
CO2	Create and populate a relational database and retrieve information from the database by formulating SQL queries.	L5, L6	Evaluate, Create
CO3	Explain the concepts of transaction, concurrency and recovery.	L2	Understand
CO4	Apply indexing mechanisms for efficient retrieval of information from database.	L3	Apply

Course Contents

Unit-I Introduction to Database Concepts 03 Hrs.

Introduction, Characteristics of Databases, File System v/s Database System, Users of Database System, Data Independence, DBMS System Architecture, Database Administrator.

Unit-II Entity–Relationship Data Model 08 Hrs.

The Entity-Relationship (ER) Model: Entity Types: Weak and Strong Entity Sets, Entity Sets, Types of Attributes, Keys, Relationship Constraints: Cardinality and Participation.

Extended Entity-Relationship (EER) Model: Generalization, Specialization and Aggregation.

Unit-III Relational Model and Relational Algebra 08 Hrs.

Introduction to the Relational Model, Relational Schema and Concept of Keys, Mapping the ER and EER Model to the Relational Model.

Relational Algebra: Unary and Set Operations, Relational Algebra Queries.

Unit-IV Structured Query Language (SQL) 09 Hrs.

Overview of SQL, Data Definition Commands, Data Manipulation Commands, Data Control Commands, Transaction Control Commands.

Integrity Constraints: Key Constraints, Domain Constraints, Referential Integrity, Check Constraints, Set and String Operations, Aggregate Function, Group By Clause, Having Clause.

Views in SQL, Joins, Nested and Complex Queries.

Introduction to PL/SQL

Unit-V Relational Database Design 10 Hrs.

Pitfalls in Relational-Database Designs, Concept of Normalization, Functional Dependencies, First Normal Form, 2NF, 3NF, BCNF.

Transactions Management and Concurrency:

Transaction Concept, Transaction States, ACID Properties, Concurrent Executions, Serializability – Conflict and View, Concurrency Control: Lock-Based, Timestamp-Based Protocols.

Recovery System: Introduction to Recovery System.

Unit-VI Indexing Mechanism 04 Hrs.

Hashing Techniques, Types of Indexes: Single Level Ordered Indexes, Multilevel Indexes, Overview of BTrees and B+ Trees.

Text Books:

1. Korth, Silberchatz, Sudarshan, Database System Concepts, 6^{th} Edition, McGraw – Hill.
2. Elmasri and Navathe, Fundamentals of Database Systems, 5^{th} Edition, Pearson Education.
3. Peter Rob and Carlos Coronel, Database Systems Design, Implementation and Management, 5^{th} Edition, Thomson Learning.
4. Chhanda Ray, Distributed Database System, Pearson Education India.
5. G. K. Gupta, Database Management Systems, McGraw – Hill.

Reference Books:

1. Dr. P.S. Deshpande, SQL and PL/SQL for Oracle 10g, Black Book, Dreamtech Press.
2. Gillenson, Paulraj Ponniah, Introduction to Database Management, Wiley Publication.
3. Raghu Ramkrishnan and Johannes Gehrke, Database Management Systems, 3^{rd} Edition, McGraw – Hill.
4. M. Tamer Ozsu, Patrick Valduriez, Principles of Distributed Database, 2^{nd} Edition, Pearson Education India.

Evaluation Scheme:

Theory :

Continuous Assessment (A):

Subject teacher will declare Teacher Assessment criteria at the start of semester.

Continuous Assessment (B):

1. Two term tests of 15 marks each will be conducted during the semester.
2. Average of the marks scored in both the tests will be considered for final grading.

End Semester Examination (C):

1. Question paper based on the entire syllabus, summing up to 65 marks.
2. Total duration allotted for writing the paper is 3 hrs.

Database Management Systems Laboratory (PCCS3040L)

Practical Scheme
Practical : 02 Hrs./week
Credit : 01

Examination Scheme
Teacher Assessment : 25 Marks
End Sem Exam : 25 Marks
Total : 50 Marks

Course Objectives:

1. To design an Entity-Relationship (ER) / Extended Entity-Relationship (EER) Model for a given application.
2. To define schema by converting conceptual model to relational model.
3. To understand the use of Structured Query Language (SQL) syntax for design of given application.
4. To retrieve information from database using different SQL operations.

CO	Course Outcomes	Blooms Level	Blooms Description
CO1	Build ER/EER diagram for the given application.	L3	Apply
CO2	Utilize ER/EER concepts to convert into relational schema with integrity constraints for given application.	L3	Apply
CO3	Design a database for given application using DDL and DML commands.	L6	Create
CO4	Apply string, SET and Join operations, Aggregate functions and nested queries on given application database.	L3	Apply
CO5	Identify, analyze and evaluate the project developed for an application.	L3, L4, L5	Apply, Analyze, Evaluate

List of Laboratory Experiments(At Least 10)

1. To draw an ER diagram for a problem statement.
2. To implement Basic SQL commands.
3. To access & modify data using SQL.
4. To implement Joins and Views.
5. To implement Subqueries.
6. To implement Integrity Constraints.
7. To implement triggers.
8. To implement procedures, functions and cursors.
9. To simulate ARIES recovery algorithm.
10. To demonstrate export-import commands.
11. To implement B-trees/B+ trees.

Any other experiment based on syllabus may be included which would help the learner to understand topic/concept.

Evaluation Scheme:

Laboratory:

Continuous Assessment (TA):

Laboratory work will be based on PCCS3040T with minimum 10 experiments to be incorporated. The distribution of marks for term work shall be as follows:

1. Performance in Experiments: 05 Marks
2. Journal Submission: 05 Marks
3. Viva-voce: 05 Marks
4. Subject Specific Lab Assignment/Case Study: 10 Marks

The final certification and acceptance of term work will be subject to satisfactory performance of laboratory work and upon fulfilling minimum passing criteria in the term work.

End Semester Examination (ESE):

Oral/ Practical examination will be based on the entire syllabus including, the practicals performed during laboratory sessions.

Statistics for Data Science (PCCS3050T)

Teaching Scheme
Lectures : 03 Hrs./week
Credits : 03

Examination Scheme
Term Test : 15 Marks
Teacher Assessment : 20 Marks
End Sem Exam : 65 Marks
Total Marks : 100 Marks

Prerequisite: Probability, Probability distribution

Course Objectives:

To build the strong foundation in statistics which can be applied to analyze data and make predictions.

CO	Course Outcomes	Blooms Level	Blooms Description
CO1	Interpret data using descriptive statistics.	L2	Understand
CO2	Demonstrate sampling distributions and estimate statistical parameters.	L2	Understand
CO3	Formulate hypothesis based on data and perform testing using various statistical techniques.	L6	Create
CO4	Develop analysis of variance on data.	L3	Apply
CO5	Examine relations between data.	L4	Analyze

Course Contents

Unit-I Introduction to Statistics 08 Hrs.

Types of Statistics, Population vs Sample

Measures of Central Tendency: Arithmetic Mean, Properties, Weighted Mean, Properties, Median, Mode, Grouped and Ungrouped Data, Empirical Relation between the Mean, Median and Mode, Geometric Mean, Harmonic Mean, Relation between Arithmetic, Geometric and Harmonic Mean, Outlier.

Measures of Dispersion: Range, Quartile Deviation, Mean Deviation, Standard Deviation, Properties, Variance, Root Mean Square Deviation, Empirical Relations between Measures of Dispersion, Absolute and Relative Dispersion, Coefficient of Variation, Moments, Pearson's β and γ Coefficients, Skewness, Kurtosis, Population Parameters and Sample Statistics, Histogram, Frequency Polygon.

Measures of Position: Quartiles, Interquartile Range, Semi Interquartile Range, Percentiles, Percentile Rank, 10–90 Percentile Range, Box and Whisker Plot.

Unit-II Sampling Distribution and Estimation 07 Hrs.

Sampling Distribution: Central Limit Theorem, Population Distribution, Chi-Square Distribution, z-Distribution, Student's t-Distribution, f-Distribution.

Statistical Estimation: Characteristics of Estimators, Consistency, Unbiasedness, Unbiased Estimates, Efficient Estimates, Sufficient Estimators, Point Estimates, Interval Estimates, Determination of Sample Size for Estimating Mean and Proportions, Estimates of Population Parameters, Probable Error.

Unit-III Hypothesis Testing for Data Driven Decision Making 12 Hrs.

Hypothesis testing: Test of Significance, Null and Alternative Hypothesis, Type I and Type II Error, Factors Affecting Type II Error, Probability of Type II Error, Power of Test, p Value, Critical Region, Level of Significance.

Confidence Interval: Population Mean, Difference between Two Population Means, Population Proportion, Difference between Two Population Proportions, Variance, Ratio of Variances of Two Populations. Goodness of Fit Test using Kolmogorov-Smirnov Test and Anderson Darling Test.

Tests using z-Statistics: Difference between Sample Proportion and Population Proportion, Difference between Two Sample Proportion, Difference between Sample Mean and Population Mean with Known σ and Unknown σ, Difference between Two Sample Means, One Tailed and Two Tailed Tests.

Test using t-Statistics: Difference between Sample Mean and Population Mean, Difference between Two Independent Sample Means, Difference between Means from the Same Group.

Test using f-Statistics: Equality of Population Variance.

Test using Chi-Square Statistics: Test of Independence, Goodness of Fit.

Unit-IV Analysis of Variance (ANOVA) for data analysis 07 Hrs.

Sample Size Calculation, One Way ANOVA, POST-HOC Analysis (Tukey's Test), Randomized Block Design, Two Way ANOVA.

Unit-V Examining Relationship 08 Hrs.

Correlation: Scatter Plot, Covariance, Karl Pearson's Coefficient of Correlation, Hypothesis Test for Correlation, Correlation vs Causation, Extreme Data Values, Limits of Correlation Coefficient, Rank Correlation, Spearman's Rank Correlation Coefficient, Repeated Ranks, Partial and Multi Correlation.

Regression: Linear Regression Analysis, Lines of Regression, Regression Coefficients, Scatter Plot with Regression Lines, Hypothesis Test for Regression, Multiple Regression, Coefficient of Determination, Residuals, Collinearity, Influential Observations.

Text Books:

1. Thomas Haslwanter, "An Introduction to Statistics with Python", 3^{rd} Edition, Springer, 2016.
2. Allen B. Downey, "Think Stats: Probability and Statistics for Programmers", 1^{st} Edition, Green Tea Press, 2011.
3. Enrich L. Lehmann, Joseph P. Romano, "Testing Statistical Hypotheses", 3^{rd} Edition, Springer, 2008.
4. S. P. Gupta, "Statistical Methods", 43^{rd} Edition, Sultan Chand, 2014.

Reference Books:

1. Peter Bruce, Andrew Bruce, Peter Gedeck, "Practical Statistics for data scientists 50+ Essential Concepts Using R and Python", 2^{nd} Edition, O'Reilly Media, Inc, 2020.
2. David Freedman, Robert Pisani, Roger Purves, W. W. Norton, "Statistics", 4^{th} Edition, 2007.
3. S. C. Gupta, V. K. Kapoor, "Fundamentals of mathematical statistics", 10^{th} Edition, Sultan Chand, 2002.

Evaluation Scheme:

Theory :

Continuous Assessment (A):

Subject teacher will declare Teacher Assessment criteria at the start of semester.

Continuous Assessment (B):

1. Two term tests of 15 marks each will be conducted during the semester.
2. Average of the marks scored in both the tests will be considered for final grading.

End Semester Examination (C):

1. Question paper based on the entire syllabus, summing up to 65 marks.

2. Total duration allotted for writing the paper is 3 hrs.

Statistics for Data Science Laboratory (PCCS3050L)

Practical Scheme
Practical : 02 Hrs./week
Credit : 01

Examination Scheme
Teacher Assessment : 25 Marks
End Sem Exam : 25 Marks
Total : 50 Marks

Course Objectives:

To build the strong foundation in statistics which can be applied to analyze data and make predictions.

CO	Course Outcomes	Blooms Level	Blooms Description
CO1	Outline different types of data and its visualization.	L2	Understand
CO2	Choose appropriate descriptive statistics measures for statistical analysis.	L3	Apply
CO3	Solve Confidence Interval for different parameters	L3	Apply
CO4	Examine hypothesis test using various statistics.	L5	Analyze
CO5	Discuss nonparametric tests of hypotheses.	L6	Create
CO6	Solve Correlation and Regression Data Analytical Methods.	L3	Apply

List of Laboratory Experiments: (any 8 using Python)

1. To perform descriptive statistics on data.
2. To visualize descriptive statistics on data.
3. To prove central limit theorem.
4. To study sampling distributions and their parameters.
5. To perform statistical estimation tests on data.
6. To calculate confidence interval for different parameters.
7. To perform goodness of fit using Kolmogorov-Smirnov test and Anderson Darling test.
8. To perform hypothesis test using z statistics.
9. To perform hypothesis test using t statistics.
10. To perform hypothesis test using f statistics.
11. To perform hypothesis test using Chi Square.
12. To perform ANOVA on given data.
13. To perform Correlation on given data.
14. To perform Regression on given data Regression and evaluate the model.

Any other experiment based on syllabus may be included which would help the learner to understand topic/concept.

Evaluation Scheme:

Laboratory:

Continuous Assessment (TA):

Laboratory work will be based on PCCS3050T with minimum 08 experiments to be incorporated. The distribution of marks for term work shall be as follows:

1. Performance in Experiments: 05 Marks
2. Journal Submission: 05 Marks
3. Viva-voce: 05 Marks
4. Subject Specific Lab Assignment/Case Study: 10 Marks

The final certification and acceptance of term work will be subject to satisfactory performance of laboratory work and upon fulfilling minimum passing criteria in the term work.

End Semester Examination (ESE):

Oral / Practical examination will be based on the entire syllabus including the practicals performed during laboratory sessions.

Programming with Python Laboratory (PCCS3060L)

Practical Scheme
Practical : 02 Hrs./week
Credit : 01

Examination Scheme
Teacher Assessment : 25 Marks
End Sem Exam : 25 Marks
Total : 50 Marks

Course Objectives:

1. To learn the basic and OOP concepts of Python.
2. To study various advance python concept like inheritance, exception handling, modules etc.
3. To learn to develop GUI based standalone and web application.

CO	Course Outcomes	Blooms Level	Blooms Description
CO1	Demonstrate basic data types and data structures in python.	L2	Understand
CO2	Demonstrate the concepts of Object-Oriented Programming.	L2	Understand
CO3	Experiment with file, directory handling and text processing concepts in python.	L3	Apply
CO4	Make use of database connectivity, client-server communication using python.	L3	Apply
CO5	Utilize various advance modules of Python for data analysis.	L3	Apply

Course Contents

Unit-I Python Basics 08 Hrs.

Data Types in Python, Operators in Python, Input and Output, Control Statement, Arrays in Python, String and Character in Python, Functions, List and Tuples, Dictionaries.

Unit-II Introduction to OOP 08 Hrs.

Classes, Objects, Constructor, Methods, Abstraction, Inheritance, Magic Methods, Exception Handling

Unit-III Advanced Python 09 Hrs.

Building Modules, Packages: Python Collections Module, Opening and Reading Files and Folders (Python OS Module, Python Datetime Module, Python Math and Random Modules, Text Processing and Regular expression in Python)

Unit-IV Python Integration Primer 08 Hrs.

Graphical User Interface using Tkinter: Form Designing Networking in Python: Client Server Socket Programming, Python Database Connectivity using SQL lite.

Unit-V Python Advance Modules 09 Hrs.

Numpy: Working with Numpy, Constructing Numpy Arrays, Printing Arrays, Arithmetic Operations on Matrix's, Numpy zeros(), Matplotlib: Matplotlib-Installation and Sample Code, Bar Chart Pandas: Data Processing, Pandas-Data structure, Pandas-Series Data, Data Frames

Suggested List of Laboratory Experiments:

1. Exploring basics of python like data types (strings, list, array, dictionaries, set, tuples) and control statements.
2. Demonstrate the concept of Abstraction in Python.
3. Demonstrate the concept of Inheritance.
4. Demonstrate exception handling.
5. Python program to explore different types of Modules
6. Exploring Files and directories -
   (a) Python program to append data to existing file and then display the entire file.
   (b) Python program to count number of lines, words and characters in a file.
   (c) Python program to display file available in current directory
7. Make use of RE module to do text processing.
8. Creating GUI with python containing widgets such as labels, textbox, radio, checkboxes and custom dialog boxes.
9. Program to demonstrate CRUD (create, read, update and delete) operations on database (SQLite/ MySQL)using python.
10. Creation of simple socket for basic information exchange between server and client.
11. Make use of advance modules of Python like Matplotlib, Numpy, Pandas.

Any other experiment based on syllabus may be included, which would help the learner to understand topic/concept.

Text Books:

1. Learn Python the Hard Way, Zed Shaw's Hard Way Series, 3rd Edition,2013.
2. Python Projects, Laura Cassell, Alan Gauld, wrox publication,2015.

Digital Resources:

1. The Python Tutorial, http://docs.python.org/release/3.0.1/tutorial/
2. http://spoken-tutorial.org
3. www.staredusolutions.org

Evaluation Scheme:

Laboratory:

Continuous Assessment (TA) 25 Marks:

Laboratory work will be based on PCCS3060L with minimum 10 experiments to be incorporated. The distribution of marks for term work shall be as follows:

1. Performance in Experiments: 05 Marks
2. Journal Submission: 05 Marks
3. Viva-voce: 05 Marks
4. Subject Specific Lab Assignment/Case Study: 10 Marks

The final certification and acceptance of term work will be subject to satisfactory performance of laboratory work and upon fulfilling minimum passing criteria in the term work.

End Semester Examination (ESE) 25 Marks:

Oral / Practical examination will be based on the entire syllabus including, the practicals performed during laboratory sessions.

Semester Project-I (PJCS3070L)

Practical Scheme
Practical : 02 Hrs./week
Credit : 01

Examination Scheme
Teacher Assessment : 25 Marks
End Sem Exam : 25 Marks
Total : 50 Marks

Course Objectives:

Students are expected to design, simulate/implement a project based on the knowledge acquired from current semester subjects.

CO	Course Outcomes	Blooms Level	Blooms Description
CO1	Conduct a survey of several available literatures in the preferred field of study.	L4	Analyze
CO2	Demonstrate various/alternate approaches to complete a project.	L2	Understand
CO3	Ensure a collaborative project environment by interacting and dividing project work among team members.	L3	Apply
CO4	Present their project work in the form of a technical report / paper and thereby improve the technical communication skill.	L3	Apply
CO5	Demonstrate the ability to work in teams and manage the conduct of the research study.	L2	Understand

Semester Project:

The purpose of introducing semester project at second year level is to provide exposure to students with a variety of projects based on the knowledge acquired from the semester subjects. This activity is supposed to enrich their academic experience and bring enough maturity in student while selecting the project. Students should take this as an opportunity to develop skills in implementation, presentation and discussion of technical ideas/topics. Therefore, proper attention shall be paid to the content of semester project report which is being submitted in partial fulfillment of the requirements of the Second Year and it is imperative that a standard format be prescribed for the report.

Each student shall work on project approved by departmental committee approved by the Head of Department, a group of 03 to 05 students (max allowed: 5 students in extraordinary cases, subject to the approval of the department committee and the Head of the department) shall be allotted for each Semester Project. Each group shall submit at least 3 topics for the Semester Project. The departmental committee shall finalize one topic for every group. Semester Project Title or Theme should be based on knowledge acquired during semester. The project work shall involve sufficient work so that students get acquainted with different aspects of knowledge acquired from semester subjects.

Student is expected to:

- Select appropriate project title based on acquired knowledge from current semester subjects.
- Maintain Log Book of weekly work done (Log Book Format will be as per Table 1).
- Report weekly to the project guide along with log book.

Assessment Criteria:

- At the end of the semester, after confirmation by the project guide, each project group will submit project completion report in prescribed format for assessment to the departmental committee (including project guide).
- Assessment of the project (at the end of the semester) will be done by the departmental committee (including project guide).

Prescribed project report guidelines:

Size of report shall be of minimum 25 pages. Project Report should include appropriate content for:

- Introduction
- Literature Survey
- Related Theory
- Implementation details

- Project Outcomes
- Conclusion
- References

Assessment criteria for the departmental committee (including project guide) for Continuous Assessment:

Guide will monitor weekly progress and marks allocation will be as per Table 2.

Assessment criteria for the departmental committee (including project guide) for End Semester Exam:

Departmental committee (including project guide) will evaluate project as per Table 3.

Each group shall present/publish a paper based on the semester project in reputed/peer reviewed Conference/Journal/TechFest/Magazine before end of the semester.

Table 1: Log Book Format

Sr	Week (Start Date:End Date)	Work Done	Sign of Guide	Sign of Coordinator
1				
2				

Table 2: Continuous Assessment Table

Sr	Exam Seat No	Name of Student	Student Attendance	Log Book Maintain	Literature Review	Depth of Understanding	Report	Total
			5	5	5	5	5	25

Table 3: Evaluation Table

Sr	Exam Seat No	Name of Student	Project Selection	Design/ Simulation/ Logic	PCB/ hardware/ programming	Result Verification	Presentation	Total
			5	5	5	5	5	25

Constitution of India (MCCS3080T)

Teaching Scheme **Audit Course**

Lecture : 01 Hr./week

Course Objectives:

1. To provide basic information about Indian Constitution.
2. To identify individual role and ethical responsibility towards society.
3. To understand human rights and its implications.

CO	Course Outcomes	Blooms Level	Blooms Description
CO1	Why general knowledge and legal literacy thereby to take up competitive examinations.	L1	Remember
CO2	Explain state and central policies, fundamental duties.	L2	Understand
CO3	Identify Electoral Process, special provisions.	L3	Apply
CO4	Relate powers and functions of Municipalities, Panchayat's and Co- operative Societies.	L1	Remember
CO5	Develop Engineering ethics and responsibilities of Engineers.	L3	Apply
CO6	Classify Engineering Integrity & Reliability.	L4	Analyze

Course Contents

Unit-I Introduction to the Constitution of India 2 Hrs.

The Making of the Constitution and Salient features of the Constitution. Preamble to the Indian Constitution Fundamental Rights & its limitations.

Unit-II Directive Principles of State Policy 3 Hrs.

Relevance of Directive Principles State Policy Fundamental Duties.

Union Executives – President, Prime Minister Parliament Supreme Court of India.

Unit-III State Executives 3 Hrs.

Governor, Chief Minister, State Legislature High Court of State.

Electoral Process in India, Amendment Procedures, 42^{nd}, 44^{th}, 74^{th}, 76^{th}, 86^{th} & 91^{st} Amendments.

Unit-IV Special Provisions 3 Hrs.

For SC & ST Special Provision for Women, Children & Backward Classes, Emergency Provisions.

Human Rights:

Meaning and Definitions, Legislation Specific Themes in Human Rights- Working of National Human Rights Commission in India Powers and functions of Municipalities, Panchyats and Cooperative Societies.

Unit-V Scope & Aims of Engineering Ethics 3 Hrs.

Responsibility of Engineers, Impediments to Responsibility.

Risks, Safety and liability of Engineers, Honesty, Integrity & Reliability in Engineering.

Text Books:

1. Durga Das Basu, "Introduction to the Constitution on India", Student Edition, Prentice –Hall EEE, 19^{th}/ 20^{th} Edition, 2001.
2. Charles E. Haries, Michael S Pritchard and Michael J. Robins, "Engineering Ethics", Thompson Asia, 2003.

Reference Books:

1. M.V.Pylee, "An Introduction to Constitution of India", Vikas Publishing, 2002.
2. M.Govindarajan, S.Natarajan, V.S.Senthilkumar, "Engineering Ethics", Prentice – Hall of India Pvt. Ltd. New Delhi, 2004.
3. Brij Kishore Sharma, " Introduction to the Constitution of India", PHI Learning Pvt. Ltd., New Delhi, 2011.

4. Latest Publications of Indian Institute of Human Rights, New Delhi.

Web Resources

1. www.nptel.ac.in
2. www.hnlu.ac.in
3. www.nspe.org
4. www.preservearticles.com

Evaluation Scheme:

1. Student should submit a report on the case study declared by teacher.
2. Audit point shall be awarded subject to submission of report of the case study declared by teacher.

Field/Internship/Industry Training

Guidelines

Minimum of six weeks in an Industry in the area of Computer Science and Engineering. The summer internship should give exposure to the practical aspects of the discipline. In addition, the student may also work on a specified task or project which may be assigned to him/her. The outcome of the internship should be presented in the form of a report.

1. Student shall undergo industrial training /internship for a minimum period of SIX weeks during summer vacations of third to sixth semester.
2. The industry in which industrial training/internship is taken should be a medium or large scale industry.
3. The paper bound report on training must be submitted by the student in the beginning of Seventh semester along with a certificate from the company where the student took training.
4. Every student should write the report separately.
5. Institute/Department/T&P Cell have to assist the students for finding Industries for the training/internship.
6. Students must take prior permission from department before joining for industrial training/internship.
7. Note that, the degree certificate will not be awarded if the certificate of field/industry/internship is not submitted to the department.
8. The field/industry/internship training will be reflected on the final marksheet/degree certificate in the section of audit points completed.