R23 Curriculum and Syllabus for Under Autonomy (NEP 2020 Implemented)

M.Sc. in Data Science (MDS)

(Effective from 2024-25 admission batch)

JIS University
Agarpara, Kolkata

SEMESTER-1							
Sl. No.	**Type**	**Course No.**	**Course Name**	**L**	**T**	**P**	**Credits**
THEORY							
1		PDT1001	Probability and Statistics	3	1	0	4
2		PDT1002	Computer Programming with Python	3	0	0	3
3		PDT1003	Introduction to Data Science	3	0	0	3
4		PDT1004	Data Structure and Algorithms	3	0	0	3
5		PDT1005	Data Base Management system	3	0	0	3
PRACTICAL							
6		PDT1104	Data Structure and Algorithms Laboratory	0	0	3	1.5
7		PDT1105	Data Base Management System Laboratory	0	0	3	1.5
8		PDT1102	Computer Programming with Python Laboratory	0	0	3	1.5
MANDATORYACTIVITIES/COURSES(Non-CGPA)							
9		JSC1501	Universal Human Values	0	0	0	1
10		JSC1502	Seminar / Group Discussion	1	0	0	1
11		JSC1503	Skill X	0	0	0	1
TOTAL				**15**	**1**	**8**	**20.5**

SEMESTER-2							
Sl. No.	**Type**	**Course No.**	**Course Name**	**L**	**T**	**P**	**Credits**
THEORY							
1		PDT2001	Mathematics for Data Science	3	1	0	4
2		PDT2002	Artificial Intelligence	3	0	0	3
3		PDT2003	R for Data Science	3	0	0	3
4		PDT2004	Multivariate Data Analysis	3	0	0	3
5		PDT2005	Data Warehousin6g and Data Mining	3	0	0	3
PRACTICAL							
6		PDT2102	Artificial Intelligence Laboratory using Python	0	0	3	1.5
7		PDT2103	R for Data Science Laboratory	0	0	3	1.5
9		PDT2105	Data Warehousing and Data Mining Lab	0	0	3	1.5
9		PDT2104	Multivariate Data Analysis Lab	0	0	3	1.5
MANDATORYACTIVITIES/COURSES(Non-CGPA)							
10		JSC2502	Seminar / Group Discussion	1	0	0	1
11		JSC2503	Skill X	0	0	0	1
12		JSC2504	NSS/2Physical Activities /Meditation Yoga/Photography	0	0	3	1
TOTAL				**15**	**1**	**9**	**22**

SEMESTER-3							
Sl. No.	**Type**	**Course No.**	**Course Name**	**L**	**T**	**P**	**Credits**
THEORY							
1		PDT3001	Machine Learning	3	0	0	3
2		PDT3002	Big Data Analytics	3	0	0	3
3		PDT3003	Data visualization	3	0	0	3
4		PDT3004	Operations Research and Optimization Techniques	3	0	0	3
5		**Elective I**		3	0	0	3
		PDT3005	Time Series Analysis				
		PDT3006	Security for Data Science				
PRACTICAL							
6		PDT3101	Machine Learning Laboratory	0	0	3	1.5
7		PDT3102	Big Data Analytics Laboratory	0	0	3	1.5
8		PDT3103	Data Visualization Laboratory	0	0	3	1.5
9		PDT3104	Seminar on Project I	0	0	0	2
MANDATORYACTIVITIES/COURSES(Non-CGPA)							
10		JSC3502	Group Discussion	1	0	0	1
11		JSC3503	Skill X	0	0	0	1
12		JSC3504	NSS/2Physical Activities /Meditation Yoga/ Photography	0	0	3	1
TOTAL				**15**	**0**	**9**	**21.5**

SEMESTER-4							
Sl. No.	**Type**	**Course No.**	**Course Name**	**L**	**T**	**P**	**Credits**
THEORY							
1		PDT4001	Deep learning	3	0	0	3
2		**Elective II**		3	0	0	3
		PDT4002	Predictive Analytics				
		PDT4003	Natural Language Processing				
PRACTICAL							
3		PDT4101	Deep Learning Lab	0	0	3	1.5
		PDT4002/4003	Predictive Analytics Lab/ Natural Language Processing Lab	0	0	3	1.5
4		PDT4105	Project / Dissertation Work	0	5	10	10
5		JSC4502	Seminar / Group Discussion	1	0	0	1
6		JSC4503	Skill X	0	0	0	1
7		JSC4504	NSS/2Physical Activities /Meditation Yoga/ Photography	0	0	3	1
TOTAL				**6**	**5**	**13**	**19**

SCHEME OF INSTRUCTION

Credits in Each Semester					
Cat. Code	Sem-I	Sem-II	Sem-III	Sem-IV	Total
Total	20.5	22	21.5	19	83

Semester 1
Curriculum and Detailed Syllabus

SEMESTER-1							
Sl. No.	**Type**	**Course No.**	**Course Name**	**L**	**T**	**P**	**Credits**
THEORY							
1		PDT1001	Probability and Statistics	3	1	0	4
2		PDT1002	Computer Programming with Python	3	0	0	3
3		PDT1003	Introduction to Data Science	3	0	0	3
4		PDT1004	Data Structure and Algorithms	3	0	0	3
5		PDT1005	Data Base Management system	3	0	0	3
PRACTICAL							
6		PDT1104	Data Structure and Algorithms Laboratory	0	0	3	1.5
7		PDT1105	Data Base Management System Laboratory	0	0	3	1.5
8		PDT1102	Computer Programming with Python Laboratory	0	0	3	1.5
MANDATORYACTIVITIES/COURSES(Non-CGPA)							
		JSC1501	Universal Human Values	0	0	0	1
		JSC1502	Seminar / Group Discussion	1	0	0	1
		JSC1503	Skill X	0	0	0	1
TOTAL				**15**	**1**	**8**	**20.5**

Course Code	PDT1001			
Course Title	Probability and Staistics			
Category	Basic Science			
LTP & Credits	L	T	P	Credits
	3	0	1	4
Total Contact Hours	48			
Pre-requisites	None			

Learning Objective:

The main objectives of this course are to:
1. Inculcate the knowledge on descriptive statistics
2. Impart the concept of probability and its applications
3. Know the distributions and its variants
4. Learn the various sampling techniques
5. Understand the concept of inferential statistics for decision making

Course Outcome:

CO1: Visualize and summarize the data
CO2: Know the usage of probability concept in a given situation
CO3: Select a suitable distribution and also to generate random sample
CO4: Draw sample by choosing suitable sampling techniques and estimate the parameters
CO5: Formulate hypothesis and perform suitable tests.

Course Content:

Module 1: Descriptive Statistics [8L]
Measures of Central Tendency: Mean, Median, Mode, Trimmed Mean – Measures of Dispersion: Range, Standard Deviation, Quartile Deviation, Mean and Median Absolute Deviation – Moments - Measures of Skewness and Kurtosis – Correlation and Linear Regression – Simple Problems.

Module 2: Basic Probability, Random Variables and Probability Distributions [10L]
Concept of Probability – Axioms of Probability - Conditional Probability – Simple Problems - Independent Events - Bayes' Rule (without proof) and Simple Applications. Discrete and Continuous Random Variables, Probability Distributions for Discrete and Continuous Random Variables – Distribution Functions for Discrete and Continuous Random Variables - Joint Distributions - Independent Random Variables - Probability Distributions of Functions of Random Variables – Marginal and Conditional Distributions – Mathematical Expectation.

Module 3: Special Probability Distributions[8L]
Notions of Binomial, Poisson Distribution and Normal Distributions – Properties – Relationship Between Binomial and Normal Distributions, Poisson and Normal Distributions – Uniform, Exponential, Gamma Distributions, t, Chi-square and F Distributions - Bivariate Normal Distribution – Simulation: Random Number Generation from Exponential, Gamma and Normal Distributions.

Module 4: Sampling Theory and Statistical Estimation Theory[10L]

Population and Sample - Random Samples – Sampling with and without Replacement, Sampling Distributions, Sampling distributions of Mean, Proportion and Difference of Means, Standard Error. Estimation of Parameters, Properties of Estimators: Unbiasedness, Consistency, Efficiency, Sufficiency. Point and Interval Estimates and Their Reliability, Confidence Interval Estimates of Population Parameters Based on Normal, t and Chi-square Distributions

Module 5: Statistical Decision Theory[12L]

Statistical Decisions, Statistical Hypothesis, Tests of Hypothesis and Significance, One-tail and Two-tail Tests. Parametric Tests: Tests Involving Normal, t, Chi-square and F Distributions - Test for Goodness of Fit, Contingency Tables, Tests for Independence of Attributes, One-way and Two-way Analysis of Variance. Non-parametric Tests: Sign Test, Run Test, Wilcoxon Signed Rank Test, Mann-Whitney U test, Kruskal-Wallis Test.

Reference Books

1. Montgomery, D. C., and Runger, G. C. (2018). Applied Statistics and Probability for Engineers, Seventh Edition, John Wiley & Sons, Inc
2. Bruce, P., Bruce, A., and Gedeck, P. (2020). Practical Statistics for Data Scientists, Second Edition, O'Reilly Media, Inc.
3. Spiegel, M. R., Schiller, J. J., and Alu Srinivasan, R. (2013). Probability and Statistics, Fourth Edition, Schaum's Outline Series, McGraw Hill Companies, Inc.

CO-PO Mapping:

	PO 1	**PO 2**	**PO 3**	**PO 4**	**PO 5**	**PO 6**	**PO 7**	**PO 8**	**PO 9**	**PO10**	**PO11**	**PO12**
CO1	3	2	-	-	-	-	-	-	-	-	-	1
CO2	3	2	3	-	-	-	-	-	-	-	-	1
CO3	3	2	2	-	-	-	-	-	-	-	-	1
CO4	3	3	1	-	-	-	-	-	-	-	-	1
CO5	3	3	2	-	-	-	-	-	-	-	-	1

Course Code	PDT1002			
Course Title	Computer Programming in Python			
Category	Basic Science			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites	None			

Learning Objective:

In this practical course, the students will be learning Python programming basics and paradigm. python looping, control statements and string manipulations. Students will be made familiar with the concepts of various modules, packages and python libraries used for various applications (Machine learning, Deep learning etc.).

Course Outcomes:

CO1: Understand and explain the basic principles of Python programming language and object-oriented concepts.
CO2: Define and demonstrate the use of built-in data structures along with the help of condition checking and looping structures.
CO3: Understand and apply various applications of different modules and packages in Python.
CO4: Learn to handle exceptions and files in Python.
CO5: Apply Python programming concepts to develop a computer game for teaching data structures like stacks and queues, enhancing visualization and understanding of abstract concepts through game-based learning.

Contents

Module 1:
Introduction to Python: Python variables, Python basic Operators, Python Data Types, variables, Declaring and using Numeric data types: int, float etc., Basic Input-Output Operations, Basic Operators

Module 2:
Conditionals and loops: Boolean Values, if, else and else if, Simple for loops in python, for loop using ranges, string, list and dictionaries. Use of while loops in python, Loop manipulation using pass, continue, break and else.

Module 3:
Strings: Assigning values in strings, String manipulations, String special operators, String formatting operators, Triple Quotes, Raw String, Unicode String, Build-in-String methods,

Module 4:
Lists: Lists Introduction, accessing values in list, List manipulations, List Operations, Indexing, slicing & matrices, use of tuple data type. string, list and Dictionary, string manipulation methods, programming using string, list and dictionary in-built functions.

Module 5:
Functions: Built –in Functions and methods, Functions, writing functions in Python, returning a result from a function, Pass by value & pass by reference, function arguments & its types, recursive functions.

Module 6:
Python packages: Simple programs using the built-in functions of packages matplotlib, numpy, pandas etc.,

Text Books:

1. Introduction to Python Programming, William Mitchell, Povel Solin, Martin Novak et al., NCLab Public Computing, 2012.
2. Introduction to Python Programming, ©Jacob Fredslund, 2007.

Reference Books:

1. An Introduction to Python, John C. Lusth, The University of Alabama, 2011.
2. Introduction to Python, ©Dave Kuhlman, 2008.

CO-PO Mapping:

	PO 1	**PO 2**	**PO 3**	**PO 4**	**PO 5**	**PO 6**	**PO 7**	**PO 8**	**PO 9**	**PO1 0**	**PO1 1**	**PO1 2**
CO1	3	1	1	2	3	2	0	0	3	0	0	3
CO2	2	2	1	1	1	1	0	0	0	0	0	3
CO3	1	3	1	2	2	1	0	0	1	0	0	3
CO4	2	2	2	1	1	1	0	0	1	0	0	3
CO5	2	1	1	2	2	0	0	0	2	0	0	3

Course Code	PDT1003			
Course Title	Introduction to Data Science			
Category	Professional Core			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites	None			

Learning Objective:

In this course, the students will learn about the fundamentals of data science. The course will also impart design thinking capability to build big-data. Also, developing design skills of models for big data problems shall be covered.

After the completion of this course, the students will be in a better position to learn and understand the basic programming tools for data sciences.

Course Outcome:

CO1: To apply data visualization in big-data analytics

CO2: To explain and utilize Exploratory Data Analysis

CO3: To explain and utilize matrix decomposition techniques to perform data analysis

CO4: To explain and demonstrate data pre-processing techniques

CO5: To apply basic machine learning algorithms in various applications

Course Content:

Module 1: Introduction [4L]

Big Data and Data Science: Big Data Analytics, Business intelligence vs. Big data, big data frameworks, Current landscape of analytics, data visualization techniques, visualization software.

Module 2: Exploratory Data Analysis (EDA) [6L]

Exploratory Data Analysis (EDA), statistical measures, Basic tools (plots, graphs and summary statistics) of EDA, Data Analytics Lifecycle, Discovery.

Module 3: Basic Statistical Inference [5L]

Developing Initial Hypotheses, Identifying Potential Data Sources, EDA case study, testing hypotheses on means, proportions and variances.

Module 4: Regression models [5L]

Regression models: Simple linear regression, least-squares principle, MLR, logistic regression, Multiple correlation, Partial correlation.

Module 5: Linear Algebra Basics **[4L]**

Matrices to represent relations between data, Linear algebraic operations on matrices – Matrix decomposition: Singular Value Decomposition (SVD) and Principal Component Analysis (PCA).

Module 6: Data Pre-processing and Feature Selection **[6L]**

Data cleaning, Data integration, Data Reduction, Data Transformation and Data Discretization, Feature Generation and Feature Selection, Feature Selection algorithms: Filters, Wrappers, Decision Trees, Random Forests.

Module 7: Basic Machine Learning Algorithms **[6L]**

Classifiers: Decision tree, Naive Bayes classifier, k-Nearest Neighbors (k-NN), k-means, Support Vector Machine. Association Rule mining – Ensemble methods.

Text / Reference Books:

1. J. Leskovek, A. Rajaraman and J. Ullman, "Mining of Massive Datasets. v2.1", Cambridge University Press.
2. S. Acharya and S. Chellappan, "Big Data Analytics", Wiley.
3. J. Han, K. Kamber and J. Pei, "Data Mining: Concepts and Techniques", Morgan Kaufmann.
4. J. Liebowitz, "Big Data and Business Analytics", CRC Press.
5. C. Rajan, "Data mining methods, 2nd edition", Narosa.

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	1	1	1	2	2	-	-	-	-	-	-	3
CO2	1	2	2	2	2	-	-	-	-	-	-	3
CO3	1	2	2	2	1	-	-	-	-	-	-	3
CO4	2	1	1	1	1	-	-	-	-	-	-	1
CO5	2	1	1	1	1	-	-	-	-	-	-	3

Course Code	PDT1004			
Course Title	Data Structures and Algorithms			
Category	Professional Core			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites	a) Fundamentals of Programming			

Learning Objective:

In this course, the students will be taught about the significance of non-linear data structures with respect to the access and organization of data, various algorithmic approaches to write programs to solve problems in different engineering domains by using different data structures, merits and demerits of altered algorithms in terms of time-complexity.

Course Outcome:

CO1: To differentiate how the choices of data structure & algorithm methods impact the performance of program.

CO2: To solve problems based upon different data structure & also write programs.

CO3: To identify appropriate data structure & algorithmic methods in solving problem.

CO4: To discuss the computational efficiency of the principal algorithms for sorting, searching, and hashing

CO5: To compare and contrast the benefits of dynamic and static data structures implementations.

Course Content:

Module 1: Introduction of Data Structure [11L]

Concepts of data structures, Abstract Data Type.
Algorithms and programs, basic idea of pseudo-code, Properties of an algorithm.
Algorithm efficiency and analysis, time and space analysis of algorithms – order notations.
Array: Different representations – row major, column major.
Sparse matrix - its implementation and usage, Array representation of polynomials.
Linked List: Singly linked list – operations, Doubly linked list – operations.
Circular linked list – operations, Linked list representation of polynomial and applications.

Module 2: Linear Data Structure [10L]

Stack and its implementations (using array and linked list).
Applications (Infix, Prefix, and Postfix with their conversions, Postfix Evaluation).
Queue, circular queue, de-queue.
Implementation of queue- linear and circular (using array and linked list).

Recursion:Principles of recursion - use of stack, tail recursion.
Applications - The Tower of Hanoi, Eight-queen problem.

Module 3: Nonlinear Data structures [17L]

Trees: Basic terminologies, forest, tree representation (using array and linked list).

Binary trees - binary tree traversal (pre-, in-, post- order).

Threaded binary tree – operations.

Binary search tree- operations (creation, insertion, deletion, searching).

Concept of Max-Heap and Min-Heap (creation, deletion).

Height balanced binary tree – AVL tree (insertion, deletion with examples only).

Graph traversal and connectivity – Depth-first search (DFS), Breadth-first search (BFS) – concepts of edges used in DFS and BFS (tree-edge, back-edge, cross-edge, and forward-edge). Minimal spanning tree – Prim's algorithm, Kruskal's algorithm (basic idea of greedy methods).

Module 4: Searching and Sorting [9L]

Sorting Algorithms: Bubble sort, Insertion sort, Selection sort – with notion of complexity, Quick sort, Merge sort – with complexity, Radix sort – with complexity.

Searching: Sequential search, Binary search, Interpolation Search– with complexity.

Hashing: Hashing functions, Collision resolution techniques.

Text/Reference Books:

1. E. Horowitz, S. Sahni and S. Anderson-Freed, "Fundamentals of Data Structures of C", Universities Press.
2. S. Lipschutz, "Data Structures", Tata McGraw Hill Education (India) Private Limited.
3. A. M. Tanenbaum, "Data Structures in C", Pearson.
4. R. Thareja, "Data Structures Using C", Oxford.
5. A.K. Rath, A. K. Jagadev, "Data Structure Using C", Scitech Publications.
6. T. H. Coreman, "Introduction to Algorithms", MIT Press.

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	3	2	2	3	-	-	-	-	3	-	-
CO2	3	2	-	-	-	2	3	2	2	-	-	-
CO3	3	-	1	-	-	-	-	-	2	3	-	-
CO4	3	1	-	-	-	-	-	-	-	-	-	-
CO5	3	2	-	-	-	3	-	-	-	-	-	-

Course Code	PDT1005			
Course Title	Data Base Management System			
Category	Professional Core			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites	a) Data Structures and Algorithms			

Learning Objective:

In this course, the students will be able to learn the data models, conceptualize and depict a database system; design system using E-R diagram; learn SQL & relational database design; understand the internal storage structures using different file and indexing techniques; know the concepts of transaction processing, concurrency control techniques and recovery procedure.

Course Outcome:

CO1: To apply the knowledge of E-R diagram for an application

CO2: To explain the creation of the normalized relational database model

CO3: To analyze real world queries to generate reports from it

CO4: To determine whether the transaction satisfies the ACID properties

CO5: To create and maintain the database of an organization

Course Content:

Module 1: Introduction
[3L]

Concept and overview of DBMS, data models.
Database languages, database administrator, database users, three-schema architecture of DBMS.

Module 2: Entity-Relationship and Relational Database Model [9L]

Basic concepts, design issues, mapping constraints, keys, entity-relationship diagram, weak entity sets, extended E-R features, case study on E-R model.
Structure of relational databases, relational algebra, relational calculus, extended relational algebra operations, views, modifications of the database.

Module 3: SQL and Integrity Constraints [6L]

Concept of DDL, DML, DCL.
Basic structure, set operations, aggregate functions, null values, domain constraints, referential integrity constraints, assertions, views, nested sub-queries.
Database security application development using SQL, stored procedures and triggers.

Module 4: Relational Database Design [6L]

Functional dependency, Different anomalies in designing a Database. Normalization using functional dependencies, decomposition, Boyce-Codd Normal Form, 3NF.
normalization using multi-valued dependencies, 4NF, 5NF, Case Study.

Module 5: Internals of RDBMS [6L]

Physical data structures, query optimization: join algorithm, statistics and cost based optimization. Transaction processing, concurrency control and recovery management: transaction model properties, state serializability, lock base protocols; two phase locking, deadlock handling.

Module 6: File Organization & Index Structures [6L]

File and record Concept, placing file records on disk, fixed and variable sized records, Ttypes of single-level index (primary, secondary, clustering).
Multilevel indices, dynamic multilevel indices using B-tree and B+ tree.

Text / Reference Books:

1. R. Elmasri and S. B. Navathe, "Fundamentals of Database Systems", Addison Wesley Publishing.
2. C.J. Date, "Introduction to Database Management", Vol. I, II, III, Addison Wesley.
3. J.D. Ullman, "Principles of Database Systems", Galgottia Publication.
4. G. Jim and R. Address, "Transaction Processing : Concepts and Techniques", Morgan Kauffman.

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	2	3	2	-	-	-	-	-	-	-	-
CO2	3	3	3	1	-	-	-	-	-	-	-	-
CO3	3	3	3	1	2	-	-	-	-	-	-	-
CO4	3	3	3	1	-	-	-	-	-	-	-	-
CO5	3	2	2	2	-	-	-	-	-	-	-	-

Course Code	PDT1102			
Course Title	Computer Programming with Python Laboratory			
Category	Professional Core			
LTP & Credits	L	T	P	Credits
	0	0	3	2
Total Contact Hours	36			
Pre-requisites	a) Fundamentals of Programming b) Basic Problem Solving			

Learning Objective:

In this practical course, the students will be learning Python programming basics and paradigm. python looping, control statements and string manipulations. Students will be made familiar with the concepts of various modules, packages and python libraries used for various applications (Machine learning, Deep learning etc.).

Course Outcomes:

CO1: Understand and explain the basic principles of Python programming language and object oriented concept.

CO2: Define and demonstrate the use of built-in data structures along with the help of condition checking and looping structures.

CO3: Understand and apply various applications of different modules and packages in Python.

CO4: Learn to handle exceptions and files in Python.

CO5: Apply Python programming concepts to develop a computer game for teaching data structures like stacks and queues, enhancing visualization and understanding of abstract concepts through game-based learning.

Suggestive List of Experiments:

1. History, Features, Setting up path, working with Python, Basic Syntax, Variable and Data Types, Operator. **[1 day]**

2. Conditional Statements: If, If- else, Nested if-else, Looping, For, While, Nested loops , Control Statements : Break, Continue, Pass. **[1 day]**

3. String Manipulation: Accessing Strings, Basic Operations, String slices, Function and Methods, Lists: Introduction, Accessing list, Operations, Working with lists, Function and Methods. **[2 days]**

4. Tuple: Introduction, Accessing tuples, Operations, Working, Functions and Methods, Dictionaries: Introduction, Accessing values in dictionaries, Working with dictionaries, Properties. **[2 days]**

5. Functions: Defining a function, Calling a function, Types of functions, Function Arguments, Anonymous functions, Global and local variables. **[1 day]**

6. Modules: Importing module, Math module, Random module, Packages, Composition, Input-Output Printing on screen, Reading data from keyboard, Opening and closing file, Reading and writing files, Functions. **[2 days]**

7. Exception and File Handling: Exception, Exception Handling, Except clause, Try & finally clause, User Defined Exceptions. **[1 day]**

8. A case study on using a computer game for teaching data structures on stacks and queues. The computer game is developed to help students visualize the data structures and data access operations on stacks and queues. This game-based learning is engaging, fun and, more importantly, abstract concepts in data structures can be visualized and learnt through game playing. **[2 days]**

Text / Reference Books:

1. T. R. Padmanabhan, "Programming with Python (1st Ed.)", Springer.
2. R. Thareja, "Python Programming: using Problem Solving Approach (1st Ed.)", Oxford University Press.
3. W. McKinney, "Python Data Analysis (2nd Ed.)", O.Reilly.

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	1	1	2	3	2	0	0	3	0	0	3
CO2	2	2	1	1	1	1	0	0	0	0	0	3
CO3	1	3	1	2	2	1	0	0	1	0	0	3
CO4	2	2	2	1	1	1	0	0	1	0	0	3
CO5	2	1	1	2	2	0	0	0	2	0	0	3

Course Code	PDT1104			
Course Title	Data Structures and Algorithms Laboratory			
Category	Professional Core			
LTP & Credits	L	T	P	Credits
	0	0	3	2
Total Contact Hours	36			
Pre-requisites	a) Fundamentals of Programming			

Learning Objective:

In this course, the students will learn about C program based implementation of different algorithmic approaches by using non-linear and linear data structures to solve problems in different engineering domains.

Course Outcome:

CO1: To choose appropriate data structure as applied to specified problem definition

CO2: To compare operations like searching, insertion, deletion, traversing mechanism on various data structures

CO3: To explain various practical applications of data structures

CO4: To analyze how to store, manipulate and arrange data in an efficient manner

CO5: To demonstrate how to implement various data structures using arrays and linked list

Suggestive List of Experiments:

1. Experiments on arrays **[1 day]**
   Addition and Multiplication of Arrays
   Implementation of Sparse Matrices

2. Experiments on Abstract Data Types **[2 days]**
   Implementation of stack using Array
   Applications of stack – infix to postfix conversion, expression evaluation

3. Experiments on Linked List **[2 days]**
   Implementation of linked lists and its operations -- insertion, deletion and reverse
   Implementation of stacks and queues using linked list.
   Polynomial addition and polynomial multiplication.

4. Experiments on Searching and Sorting **[2 days]**
   Searching: Linear Search, Binary Search
   Sorting: Bubble Sort, Selection Sort, Insertion Sort, Quick Sort, Merge Sort and Heap Sort

5. Experiments on Non-linear Data Structures **[2 days]**
   Traversals of binary tree, Binary Search Tree (BST), Threaded binary tree

Height balanced binary tree – AVL tree (insertion, deletion)
B- Trees – insertion, deletion

6. Experiments on Hashing **[1 day]**
   Implementation of Hash tables and its operations -- searching, inserting, and deleting, handling collisions.

7. Innovative Experiments **[2 days]**
   Case study of solving complex problems from various engineering domains using suitable data structures (e.g., mesh analysis in electrical circuits, event-driven simulation, etc.).

Text/Reference Books:

1. C. E. Balagurusamy, "Data Structures using C", McGraw Hill.
2. E. Horowitz, S. Sahni and S. Anderson-freed, "Fundamentals of Data Structures of C", Universities Press.
3. A. K. Sharma, "Data Structures using C", Pearson.
4. R. Thareja, "Data Structures using C", Oxford University Press.

CO-PO mapping:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	3	2	2	2	-	-	-	-	1	-	-
CO2	-	2	2	-	2	-	-	-	-	1	-	2
CO3	2	1	1	-	-	-	-	-	-	-	-	-
CO4	3	2	-	2	-	-	-	-	-	-	1	-
CO5	-	-	2	1	2	-	-	-	-	-	1	2

Course Code	PDS1105			
Course Title	Data Base Management System Laboratory			
Category	Professional Core			
LTP & Credits	L	T	P	Credits
	0	0	3	2
Total Contact Hours	36			
Pre-requisites	a) Data Structures and Algorithms Laboratory			

Learning Objective:

In this course, the students will able to learn the data models, conceptualize and depict a database system; learn the fundamental concepts of SQL queries; understand the concept of designing a database with the necessary attributes; know the methodology of Accessing, Modifying and Updating data & information from the relational databases; learn database design as well as to design user interface and how to connect with database.

Course Outcome:

CO1: To understand the basic concepts regarding database, SQL queries

CO2: To explain the concepts of PL/SQL

CO3: To differentiate between DBMS and advanced DBMS.

CO4: To analyze database system concepts and apply normalization to the data base

CO5: To apply and create different transaction processing and concurrency control applications

Suggestive List of Experiments:

1. Experiments on fundamentals of data base systems **[2 days]**
   Creating a Database
   Creating a Table
   Specifying Relational Data Types
   Specifying Constraints
   Creating Indexes

2. Experiments on database Tables and Record handling **[2 days]**
   INSERT statement
   Use of SELECT and INSERT together
   DELETE, UPDATE, TRUNCATE statements
   DROP, ALTER statements

3. Experiments on retrieving data from database **[3 days]**
   The SELECT statement
   Use of the WHERE clause
   Use of the Logical Operators in the WHERE clause
   Use of IN, BETWEEN, LIKE , ORDER BY, GROUP BY and HAVING Clause
   Use of the Aggregate Functions

Combining tables using JOINS
Sub-queries

3. Experiments on Miscellaneous Database Management **[1 day]**
Creating Views
Creating Column Aliases
Creating Database Users
Use of GRANT and REVOKE

4. Experiments on PL/SQL **[1 day]**
Use of decision making statement, different loop structures to solve simple programs (e.g., sum of few numbers, pattern prints, etc.).
Inserting values into tables, reading data from a table. Basic working with CURSORS

5. Innovative Experiments **[3 days]**
Case study of handling complex databases (e.g., College Management System, Hospital management System, Library management System, Payroll management System, etc.)

Text / Reference Books:

1. H. F. Korth and A. Silberschatz, "Database System Concepts", McGraw Hill.
2. E. Ramez and S. Navathe, "Fundamentals of Database Systems", Benjamin Cummings Publishing Company.
3. C. J. Date, "Introduction to Database Management", Vol. I, II, III, Addison Wesley.
4. G. Jim and R. Address, "Transaction Processing : Concepts and Techniques", Moragan Kauffman.
5. J.D. Ullman, "Principles of Database Systems", Galgottia Publication.
6. I. Bayross , "SQL, PL/SQL the Programming Language of Oracle", BPB Publications.

CO-PO Mapping:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	3	2	1	-	-	-	-	-	-	-	-
CO2	3	2	2	1	2	-	-	-	-	-	-	-
CO3	1	2	3	-	-	-	-	-	-	-	-	-
CO4	3	1	2	2	1	-	-	-	-	-	-	-
CO5	2	2	3	1	-	-	-	-	-	-	-	-

Semester 2
Curriculum and Detailed Syllabus

SEMESTER-2							
Sl. No.	**Type**	**Course No.**	**Course Name**	**L**	**T**	**P**	**Credits**
THEORY							
1		PDT2001	Mathematics for Data Science	3	1	0	4
2		PDT2002	Artificial Intelligence	3	0	0	3
3		PDT2003	R for Data Science	3	0	0	3
4		PDT2004	Multivariate Data Analysis	3	0	0	3
5		PDT2005	Data Warehousin6g and Data Mining	3	0	0	3
PRACTICAL							
6		PDT2102	Artificial Intelligence Laboratory using Python	0	0	3	1.5
7		PDT2103	R for Data Science Laboratory	0	0	3	1.5
9		PDT2105	Data Warehousing and Data Mining Lab	0	0	3	1.5
9		PDT2104	Multivariate Data Analysis Lab	0	0	3	1.5
MANDATORYACTIVITIES/COURSES(Non-CGPA)							
10		JSC2502	Seminar / Group Discussion	1	0	0	1
11		JSC2503	Skill X	0	0	0	1
12		JSC2504	NSS/2Physical Activities /Meditation Yoga/ Photography	0	0	3	1
TOTAL				**15**	**1**	**9**	**22**

Course Code	PDT2001			
Course Title	Mathematics for Data Science			
Category	Basic Science			
LTP & Credits	L	T	P	Credits
	3	1	0	4
Total Contact Hours	48			
Pre-requisites	None			

Learning Objective:

In this course, the students will learn the basic concepts, theories and techniques of real analysis and also help students to learn the application of real analysis in different fields.

Course Outcome:

CO1: learn the basic matrix algebra and its uses for data science;

CO2: understand the tools that are used to diagnose the problems and solve them with new methods;

CO3: understand vector space and its application

CO4: learn the basics of linear transformation and its properties

CO5: learn the basic concept of inner product space and its application in real world applications

Course Content:

Module 1: (14L)
Algebra of Matrices, Trace and Rank of a Matrix and their properties, Determinants, Inverse, symmetric, orthogonal and idempotent matrices and their properties; Eigen values and Eigen vectors; Diagonalizing a matrix- Orthogonal diagonalization-, Applications to differential equations- Positive definite matrices- Similar matrices –Quadratic forms.

Module 2: (4L)
System of linear equations, Gauss elimination, Elementary matrices and a method for finding inverse of a matrix.

Module 3: (10L)
Vector spaces and subspaces – Linear combination, Span, Linear independence and dependence - Null space, Column space, and Row space – Basis and dimension of a vector space – Rank and nullity.

Module 4: (10L)
Introduction to linear transformations – General Linear Transformations – Kernel and range – Matrices of general linear transformation- Geometry linear operators-Change of basis

Module 5: (10L)
Inner product, Length, angle and orthogonality – Orthogonal sets – Orthogonal projections – Inner product spaces – Orthonormal basis: Gram-Schmidt process – QR Decomposition.

Text Books:

1. Nick Fieller (2015). Basics of Matrix Algebra for Statistics with R, Chapman and Hall/CRC.
2. Shayle R. Searle and Andre Khuri (2017). Matrix Algebra Useful for Statistics, John Wiley & Sons, Inc.:

Reference Books

1. Howard Anton and Chris Rorres, “Elementary Linear Algebra”, Wiley, 2011.
2. David C. Lay, “Linear Algebra and its Applications‟, Pearson Education, 2011.
3. Gilbert Strang, “Linear Algebra and its Applications”, Thomson Learning, 2009.
4. Steven J. Leon, “Linear Algebra with Applications”, Prentice Hall, 2006.

CO-PO Mapping:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	2	2	-	2	-	-	-	-	-	-	-	1
CO2	2	1	-	-	2	1	-	-	-	-	-	1
CO3	3	3	-	2	-	-	-	-	-	-	-	1
CO4	3	2	3	2	1	-	-	-	-	-	-	1
CO5	3	2	3	2	1	-	-	-	-	-	-	1

Course Code	PDT2002			
Course Title	Artificial Intelligence			
Category	Professional Elective			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites	a) Design and Analysis of Algorithms			

Learning Objective:

In this course, the students will learn the basic concepts, theories and techniques of artificial intelligence and also help students to learn the application of machine learning / AI algorithms in different fields of Computer Engineering.

Course Outcome:

CO1: Explain the basic concept of Artificial Intelligence and its applications.

CO2: Classify and analyze various AI tools and techniques.

CO3: Learn and evaluate various AI algorithms.

CO4: Apply the basic understanding of artificial intelligence in real-world applications.

CO5: Evaluate and assess the performance, effectiveness, and ethical implications of artificial intelligence systems in various real-world applications.

Course Content:

Module 1: Introduction to Artificial Intelligence (AI) [7L]

Overview: foundations, scope, problems, and approaches of AI. Intelligent agents: reactive, deliberative, goal-driven, utility-driven, and learning agents.

Module 2: AI Techniques [7L]

Artificial Intelligence programming techniques, Problem-solving through Search: forward and backward, state-space, blind, heuristic, problem-reduction, A, A*, AO*, minimax, constraint propagation, neural, stochastic, and evolutionary search algorithms, sample applications.

Module 3: Planning and Representation in AI [8L]

Planning: planning as search, partial order planning, construction and use of planning graphs, Representing and Reasoning with Uncertain Knowledge: probability, connection to logic, independence, Bayes rule, Bayesian networks, probabilistic inference, sample applications.

Module 4: Decision Making
[8L]

Decision-Making: basics of utility theory, decision theory, sequential decision problems, elementary game theory, sample applications.

Module 5: Knowledge Acquisition [6L]

Machine Learning and Knowledge Acquisition: learning from memorization, examples, explanation, and exploration. Learning nearest neighbor, naive Bayes, and decision tree classifiers, Q-learning for learning action policies, applications.

Text / Reference Books:

1. S. Russell, P. Norvig, "Artificial intelligence: A Modern Approach", Prentice Hall.
2. N. J. Nilsson, "Artificial Intelligence: A New Synthesis", Morgan-Kaufmann, 1998.
3. J. Pearl, "Heuristics: Intelligent Search Strategies for Computer Problem Solving", Addison-Wesley Publishing Company.
4. B, A. Heule, M.Van Maaren, H.Walsh, "T Handbook of Satisfiability", IOS Press.

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	1	1	2	3	2	0	0	3	0	0	3
CO2	2	2	1	1	1	1	0	0	0	0	0	3
CO3	1	3	1	2	2	1	0	0	1	0	0	3
CO4	2	2	2	1	1	1	0	0	1	0	0	3
CO5	2	1	1	1	2	2	0	0	2	0	0	3

Course Code	PDT2003			
Course Title	R for Data Science			
Category	Professional Core			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites	a) Fundamentals of Programming b) Data Structures and Algorithms			

Course Outcome:

CO1: Understand R and R studio

CO2: Create reports using R markdown

CO3: Analyse data for a given problem

CO4: Apply probability and statistics in real life problems

CO5: Draw scientific inference from data using R

Course Content:

Module 1: R AND R STUDIO
[4L]

Getting started with R - installing R and R studio - getting help - installing and loading packages - simple arithmetic calculations - data structure – expressions - conditional statements – functions – loops - R–markdown - introduction to Statistics - probability and data with R.

Module 2: EXPLORATORY DATA ANALYSIS
[6L]

Visualizing numerical data - graphing systems available in R - descriptive Statistics - measures of central tendency and dispersion – correlation - transforming data - exploring categorical variables

Module 3: PROBABILITY AND PROBABILITY DISTRIBUTIONS
[12L]

Introduction - disjoint events - general addition rule – independence - probability examples - disjoint vs. Independent - conditional probability - probability trees - normal distribution - evaluating the normal distribution - working with the normal distribution - binomial distribution - normal approximation to binomial - working with the binomial distribution.

Module 4: ESTIMATION
[8L]

Introduction to Inference - sampling from population - maximum likelihood estimator - least square estimator - confidence interval (CI) (for a mean) - accuracy vs. Precision - required sample size for mean, CI (for the mean) examples.

Module 5: TESTING OF HYPOTHESIS **[6L]**

Introduction - hypothesis testing (HT) - decision errors - large sample and small sample tests - inference for other estimators - significance vs. confidence level - statistical vs. practical significance - inference for proportions

Text/Reference Books:

[1] Gupta S. C., & Kapoor V. K., Fundamental of Mathematical Statistics, Sultan Chand & Sons, 2018.

[2] Peng R. D, Exploratory data analysis with R, Lulu.Com, 2012.

[3] Peng R. D, R programming for data science, Leanpub, 2016.

[4] Teetor P, R cookbook: Proven recipes for data analysis, statistics, and graphics, O' Reilly Media Inc., 2011.

[5] Crawley M. J., The R book, John Wiley & Sons, 2012.

CO-PO Mapping:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	3	3	3	-	-	-	-	-	-	-	3
CO2	3	3	3	3	-	-	-	-	-	-	-	3
CO3	3	2	2	3	-	-	-	-	-	-	-	3
CO4	3	3	3	3	-	-	-	-	-	-	-	3
CO5	3	2	2	3	-	-	-	-	-	-	-	3

Course Code	PDT2004			
Course Title	Multivariate Data Analysis			
Category	Basic Science			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites	Basics of Mathematics and Statistics			

Learning Objective:

The main objectives of this course are to:

1. Distinguish between dependence and interdependence techniques
2. Fit the various regression models and predict the results
3. Perform the dimension reduction techniques and interpret the results
4. Discriminate and classify the given objects by using target variable
5. Form the groups by using suitable clustering techniques

Course Outcome:

CO1: Distinguish between dependence and interdependence techniques

CO2: Fit the various regression models and predict the results

CO3: Perform the dimension reduction techniques and interpret the results

CO4: Discriminate and classify the given objects by using target variable

CO5: Form the groups by using suitable clustering techniques

Course Content:

Module 1: Introduction to Multivariate Analysis **[8L]**

Meaning of Multivariate Analysis – Multivariate Analysis in Statistical Terms – Basic concepts: Variate, Measurment Scales, Measurement Error, Multivariate Measurement, Statistical Significance and Statistical Power. Classification of Multivariate Techniques: Dependence and Independence Techniques – Applications of Multivariate Techniques.

Module 2: Multiple Regression Analysis **[8L]**

Concept of Simple and Multiple Regressions – Illustrations. Prediction using Single and Several Independent Variables – Decision Process in Multiple Regression Analysis: Objectives, Research Design, Assumptions, Estimation of Regression Model – Assessing Model Fit – Interpretation of Regression Variate using Regression Coefficients and Assessing Multicollinearity.

Module 3: Factor Analysis **[8L]**

Notion of Principal Components and Factors – Concept of Data Summarization and Data Reduction - Introduction to Principal Component Analysis and Factor Analysis – Illustrations. Decision Process in Factor Analysis: Objectives, Design, Assumptions, Deriving Factors, Interpretation of Factors, Validation of Factors – Illustrations.

Module 4: Discriminant Analysis **[6L]**

Concept of Discriminant Function – Meaning of Discriminant Analysis – Decision Process in Discriminant Analysis: Objectives, Research Design, Assumptions, Estimation of Discriminant Model, Assessing Model Fit, Interpretation.

Module 5: Cluster Analysis **[6L]**

Meaning and Conceptual Development of Cluster Analysis – Decision Process in Cluster Analysis: Objectives, Research Design, Assumptions, Deriving Clusters, Interpretation of Clusters, Validation and Profiling of Clusters – Illustrations – Basic Notion of Hierarchical and Non-hierarchical Clusters

Reference Books:

1. Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2018). Multivariate Data Analysis, Eighth Edition, Pearson.
2. Johnson, R. A., and Wichern, D. W. (2015). Applied Multivariate Statistical Analysis, Sixth Edition, Pearson.
3. Johnson, D. E. (1998). Applied Multivariate Methods for Data Analysts, First Edition, Duxbury Press.

CO-PO Mapping:

	PO 1	**PO 2**	**PO 3**	**PO 4**	**PO 5**	**PO 6**	**PO 7**	**PO 8**	**PO 9**	**PO1 0**	**PO1 1**	**PO1 2**
CO1	3	2	-	-	-	-	-	-	-	-	-	1
CO2	3	2	-	-	-	-	-	-	-	-	-	1
CO3	3	2	2	-	-	-	-	-	-	-	-	1
CO4	3	3	2	-	-	-	-	-	-	-	-	1
CO5	3	3	3	-	-	-	-	-	-	-	-	1

Course Code	PDT2005			
Course Title	Data Warehousing and Data Mining			
Category	Professional Elective			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites	a) Data Structures and Algorithms b) Database Management System			

Learning Objective:

In this course, the students will understand classical models and algorithms in data warehousing and data mining. It enables students to analyze the data, identify the problems, and choose the relevant models and algorithms to apply. This course assesses the strengths and weaknesses of various methods and algorithms and analyze their behavior.

Course Outcomes:

CO1: Provide efficient distribution of information and easy access to data

CO2: Create user friendly reporting environment

CO3: Find the unseen pattern in large volume of historical data that helps to manage an organization efficiently

CO4: Understand the concepts of various data mining Techniques

CO5: Understand the concepts of Preprocessing.

Course Content:

Module 1: Introduction to Data Warehouse and Multi-dimensional Data [6L]

Introduction to Data Warehousing, Data warehouse Architecture and Infrastructure, Data cube and lattice structure. Star, Snowflakes and Fact Constellation models, Components. Building a Data warehouse – Mapping the Data Warehouse to a Multiprocessor Architecture, Reporting and Query tools and Applications, Tool Categories.

Module 2: Online Analytical Processing (OLAP) tools [6L]

Online Analytical Processing (OLAP) vs OLTP, Need –Multidimensional Data Model – OLAP Guidelines - ROLAP vs MOLAP vs HOLAP - Multidimensional versus Multirelational OLAP - Categories of Tools – OLAP Tools and the Internet.

Module 3: Data Mining and Knowledge Discovery Process [6L]

Introduction to Data Mining, Types of data, AI vs ML vs DL - Data Mining Functionalities, Data Mining Systems and Task Primitives - Integration of a Data Mining System with a Data Warehouse - Data Preprocessing, Data Mining vs. Machine learning, Prediction with Regression - Mining Frequent

Patterns, Associations and Correlations - Mining Methods (Apriori Algorithm) - Mining Methods-FP Growth Algorithm.

Module 4: Supervised and Unsupervised learning [13L]

Classification and Prediction - Basic Concepts - Decision Tree Induction - Bayesian Classification - Lazy Learners (KNN Classification) - Classification by Backpropagation - Support Vector Machines - Clustering and Applications and Trends in Data Mining - Categorization of Major Clustering Methods, Types of Data - Partitioning Methods - K-Means Clustering - K-Medoids Clustering - Density-Based Methods->DBSCAN - Hierarchical Methods (Agglomerative approach) - Hierarchical Methods (Divisive approach) - Grid Based Methods - Model-Based Clustering Methods.

Module 5: Data mining and Its Applications [5L]

Clustering High Dimensional Data - Outlier Analysis - Data Warehousing Applications - Data Mining Applications - Machine Learning Applications Towards Research.

Text / Reference Books:

1. J. Han and M. Kamber, “Data Mining Concepts and Techniques (2nd Ed.)”, Elsevier.
2. P. Tan, M. Steinbach and V. Kumar, “Introduction To Data Mining (3rd Ed.)”, PHI / Person Education.
3. D. T. Larose, “Data Mining Methods and Models (1st Ed.)”, Wiley.
4. M. H. Dunham, "Data Mining: Introductory and Advanced Topics (1st Ed.)", Prentice Hall.

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	2	3	2	-	-	-	-	-	-	-	-
CO2	3	3	3	1	-	-	-	-	-	-	-	-
CO3	3	3	3	1	2	-	-	-	-	-	-	-
CO4	3	3	3	1	-	-	-	-	-	-	-	-
CO5	3	2	2	2	-	-	-	-	-	-	-	-

Course Code	PDT2102			
Course Title	Artificial Intelligence Lab			
Category	Professional Elective			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites	b) Design and Analysis of Algorithms			

Learning Objective:
In this course, students will learn the basic principles, techniques, and applications of Artificial Intelligence and Machine Learning for problem-solving, inference, perception, knowledge representation, and puzzles design.

Course Outcomes:

CO1: Explain the working principles of PROLOG/LISP and apply LIST structure of PROLOG.

CO2: Apply reasoning and inference principles to real-world problems and design programs to solve various puzzles.

CO3: Design simple algorithms for data classification in Python/R and test them with benchmark datasets.

CO4: Design simple algorithms for data clustering in Python/R and test them with benchmark datasets.

CO5: Analyze and evaluate algorithms for estimation/prediction using regression.

Course Content:

Suggestive List of Experiments:

In this laboratory, students will be familiarized with PROLOG/LISP language. The experiments are structured into four modules.

1. Module 1 (4 days):

Introduction to PROLOG facts & rules using a simple family tree.
Explanation of how goals are given in PROLOG and simple queries on the family tree.
Formation of recursive definitions and how PROLOG executes goals.
Implementation of Graph Search algorithms like DFS, BFS.
Implementation of well-known puzzles like the 8-queens problem, Towers-of-Hanoi problem, etc.

2. Module 2 (4 days):

Implementation of Classifiers: KNN, Naive Bayes Classifier, Decision Tree, SVM, Perceptron, Multi-Layer Perceptron, Random Forest, etc., on Python/R platform and test them on benchmark datasets. Familiarization with ML Tools: Excel, WEKA, R, Python for classification.

3. **Module 3 (3 days):**

Implementation of data clustering algorithms: K-Means, DBSCAN, Hierarchical (AGNES/DIVISIVE), etc., on Python/R platform and test them on benchmark datasets. Familiarization with ML Tools: Excel, WEKA, R, Python for clustering.

4. Module 4 (1 day):

Implementation of Regression (single and Multiple Variables) linear and non-linear, Logistic regression for prediction tasks.

Text / Reference Books:

1. S. Russell, P. Norvig, "Artificial intelligence: A Modern Approach", Prentice Hall.
2. N. J. Nilsson, "Artificial Intelligence: A New Synthesis", Morgan-Kaufmann, 1998.
3. J. Pearl, "Heuristics: Intelligent Search Strategies for Computer Problem Solving", Addison-Wesley Publishing Company.
4. B, A. Heule, M.Van Maaren, H.Walsh, "T Handbook of Satisfiability", IOS Press.

CO-PO Mapping:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	1	1	2	3	2	0	0	3	0	0	3
CO2	2	2	1	1	1	1	0	0	0	0	0	3
CO3	1	3	1	2	2	1	0	0	1	0	0	3
CO4	2	2	2	1	1	1	0	0	1	0	0	3
CO5	2	1	1	1	2	2	0	0	2	0	0	3

Course Code	PDT2103			
Course Title	R for Data Science Lab			
Category	Professional Core			
LTP & Credits	L	T	P	Credits
	0	0	3	1.5
Total Contact Hours	36			
Pre-requisites	c) Fundamentals of Programming d) Data Structures and Algorithms			

Course Outcome:

CO1: Understand R and R studio

CO2: Create reports using R markdown

CO3: Analyse data for a given problem

CO4: Apply probability and statistics in real life problems

CO5: Draw scientific inference from data using R

1. R Introduction: Reserved Words, Variables & Constants,
Operators, Operator Precedence.
2. Decision and Loop Structure: if...else, for loop, while Loop, break &
next, repeat Loop.
3. R Functions: Functions, Function Return Value, Environment & Scope,
Switch Function.
4. R Data Structures: Vectors, Matrix, List in R programming, Data Frame, Factor.
5. R Objects and Class: Object and Class, S3Class, S4 Class, R Reference
class, Inheritance.
6. R Graphs and Charts: Bar Plot, Histogram, Pie Chart, Box Plot, Strip Chart.
7. More on Plotting In R: Plot Function: Subplot, Saving Plot, 3DPlot.
8. R Program using a csv file to plot using the distributions like Normal
Distribution, Binomial Distribution, Poisson Distribution, Bernoulli Distribution
and inferencing the behaviour of data.
9. Program to normalize the data, replacing null columns either with mean or variance.
10. Program to import the data. And calculate covariance, correlation,
mean, standard deviation and generate plots.
11.Program to implement linear regression for predictive modeling check how well
the model fit the data.
12. Program to implement Multiple Regression for a given data.

13. Program to implement Cross Validation Methods

Validation set approach (or data split); Leave One Out Cross Validation; k-fold Cross Validation; Repeated k-fold Cross Validation

Text/Reference Books:

[1] Gupta S. C., & Kapoor V. K., Fundamental of Mathematical Statistics, Sultan Chand & Sons, 2018.
[2] Peng R. D, Exploratory data analysis with R, Lulu.Com, 2012.
[3] Peng R. D, R programming for data science, Leanpub, 2016.
[4] Teetor P, R cookbook: Proven recipes for data analysis, statistics, and graphics, O' Reilly Media Inc., 2011.
[5] Crawley M. J., The R book, John Wiley & Sons, 2012.

CO-PO Mapping:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	3	3	3	-	-	-	-	-	-	-	3
CO2	3	3	3	3	-	-	-	-	-	-	-	3
CO3	3	2	2	3	-	-	-	-	-	-	-	3
CO4	3	3	3	3	-	-	-	-	-	-	-	3
CO5	3	2	2	3	-	-	-	-	-	-	-	3

Course Code	PDT2104			
Course Title	Multivariate Data Analysis Lab			
Category	Basic Science			
LTP & Credits	L	T	P	Credits
	0	0	3	1.5
Total Contact Hours	36			
Pre-requisites	Basics of Mathematics and Statistics			

Learning Objective:

The main objectives of this course are to:
1. Distinguish between dependence and interdependence techniques
2. Fit the various regression models and predict the results
3. Perform the dimension reduction techniques and interpret the results
4. Discriminate and classify the given objects by using target variable
5. Form the groups by using suitable clustering techniques

Course Outcome:

CO1: Distinguish between dependence and interdependence techniques

CO2: Fit the various regression models and predict the results

CO3: Perform the dimension reduction techniques and interpret the results

CO4: Discriminate and classify the given objects by using target variable

CO5: Form the groups by using suitable clustering techniques

Problems relating to the following topics using R / Python programming shall form the basis for setting the question paper.
1. Computation of Mean vector and covariance matrix for multivariate data set
2. Generation of multivariate data using multivariate normal distribution
3. Fitting of linear, quadratic, exponential and logistic models
4. Principal Component analysis and factor analysis
5. Linear and quadratic discriminant analysis with classification of two and three groups.
6. Cluster analysis with hierarchical clustering (single linkage, average linkage, Wards method)
and non-hierarchical clustering (k-means)

Reference Books:

1. Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2018). Multivariate Data Analysis, Eighth Edition, Pearson.
2. Johnson, R. A., and Wichern, D. W. (2015). Applied Multivariate Statistical Analysis, Sixth Edition, Pearson.
3. Johnson, D. E. (1998). Applied Multivariate Methods for Data Analysts, First Edition, Duxbury Press.

CO-PO Mapping:

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO1 0	PO1 1	PO1 2
CO1	3	2	-	-	-	-	-	-	-	-	-	1
CO2	3	2	-	-	-	-	-	-	-	-	-	1
CO3	3	2	2	-	-	-	-	-	-	-	-	1
CO4	3	3	2	-	-	-	-	-	-	-	-	1
CO5	3	3	3	-	-	-	-	-	-	-	-	1

Course Code	PDT2105			
Course Title	Data Warehousing and Data Mining Lab			
Category	Professional Core			
LTP & Credits	L	T	P	Credits
	0	0	3	1.5
Total Contact Hours	36			
Pre-requisites				

Learning Objective:

Data mining is primarily used by the companies with a strong consumer focus. It enables these companies to determine the factors such as price, product positioning, or staff skills, and economic indicators, competition, and customer demographics.

Course Outcome:

CO1: Provide efficient distribution of information and easy access to data

CO2: Create user friendly reporting environment

CO3: Find the unseen pattern in large volume of historical data that helps to manage an organization efficiently

CO4: Understand the concepts of various data mining Techniques

CO5: Understand the concepts of Preprocessing.

Course Content:

List of Experiments12 Hours

1. Exploring Weka mining tool.

2.To study the file formats for the data mining.

3. Demonstration of preprocessing on dataset.

4. To convert ARFF (Attribute-Relation File Format) into text file and vice Versa.

5. To apply the concept of Linear Regression for training the given dataset.

6. Demonstration of Association rule process on dataset using apriori algorithm.

7. Demonstration of classification rule process on dataset using j48 algorithm.

8. Demonstration of classification rule process on dataset using id3 algorithm.

9. Demonstration of classification rule process on dataset using naïve bayes algorithm.

10.Demonstration of clustering rule process on dataset using simple k-means algorithm.

Text / Reference Books:

5. R. Elmasri and S. B. Navathe, “Fundamentals of Database Systems”, Addison Wesley Publishing.
6. C.J. Date, “Introduction to Database Management”, Vol. I, II, III, Addison Wesley.
7. J.D. Ullman, “Principles of Database Systems”, Galgottia Publication.
8. G. Jim and R. Address, “Transaction Processing : Concepts and Techniques”, Morgan Kauffman.

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	2	3	2	-	-	-	-	-	-	-	-
CO2	3	3	3	1	-	-	-	-	-	-	-	-
CO3	3	3	3	1	2	-	-	-	-	-	-	-
CO4	3	3	3	1	-	-	-	-	-	-	-	-
CO5	3	2	2	2	-	-	-	-	-	-	-	-

Semester 3
Curriculum and Detailed Syllabus

SEMESTER-3							
Sl. No.	**Type**	**Course No.**	**Course Name**	**L**	**T**	**P**	**Credits**
THEORY							
1		PDT3001	Machine Learning	3	0	0	3
2		PDT3002	Big Data Algorithms & Analytics	3	0	0	3
3		PDT3003	Data visualization	3	0	0	3
4		PDT3004	Operations Research and Optimization Techniques	3	0	0	3
5	PEC	**Elective I**		3	0	0	3
		PDT3005	Time Series Analysis				
		PDT3006	Security for Data Science				
PRACTICAL							
6		PDT3101	Machine Learning Laboratory	0	0	3	1.5
7		PDT3102	Big Data Algorithms & Analytics Laboratory	0	0	3	1.5
8		PDT3103	Data Visualization Laboratory	0	0	3	1.5
9		PDT3104	Seminar on Project	0	0	0	2
MANDATORYACTIVITIES/COURSES(Non-CGPA)							
10		JSC3502	Seminar / Group Discussion	1	0	0	1
11		JSC3503	Skill X	0	0	0	1
12		JSC3504	NSS/2Physical Activities /Meditation Yoga/ Photography	0	0	3	1
TOTAL				**15**	**0**	**9**	**23**

Course Code	PDT3001			
Course Title	Machine Learning			
Category	Professional			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites	a) Probability and Statistics b) Design and Analysis of Algorithms			

Learning Objective:

It covers some of the important regression, classification, clustering, rule-based and probabilistic models and algorithms. The Themes included linear and logistic regression, regularization, decision trees, probabilistic, SVMs and neural networks, clustering and reduction in feature dimensionality.

Course Outcome:

CO1: Understand the fundamental principles and concepts of machine learning.

CO2: Implement and evaluate supervised and unsupervised learning algorithms.

CO3: Apply appropriate machine learning techniques to solve classification, regression, clustering, and reinforcement learning problems.

CO4: Utilize popular machine learning libraries and frameworks such as scikit-learn, TensorFlow, and PyTorch

CO5: Design and execute experiments to analyze the performance of machine learning models.

Course Content:

Module 1: Introduction to Machine-based Learning [6L]

Overview of machine learning: definitions, types, and applications. Supervised vs. unsupervised learning. Machine learning pipeline: data preprocessing, feature engineering, model training, and evaluation. Evaluation metrics for classification and regression tasks., Bias-variance trade off and overfitting.

Module 2: Supervised Learning [8L]

Linear regression, polynomial regression, logistic regression, multi-class classification problem, Support Vector Machine for classification and regression, Decision trees and ensemble-based learning: bagging and boosting, k-nearest neighbors, Neural networks: theory, architectures, and activation functions. Deep learning fundamentals: feedforward neural networks.

Module 3: Un-Supervised Learning [8L]

Clustering algorithms: K-means, hierarchical clustering, DBSCAN. Dimensionality reduction techniques: Principal Component Analysis (PCA), t-distributed Stochastic Neighbor Embedding

(t-SNE). Association rule learning and Apriori algorithm. Gaussian Mixture Models (GMM) and Expectation-Maximization (EM) algorithm. Self-organizing maps (SOM) for dimensionality reduction. Introduction to autoencoders for unsupervised representation learning.

Module 4: Reinforcement Learning [8L]

Introduction to reinforcement learning: agent, environment, and reward. Markov Decision Processes (MDPs) and dynamic programming. Monte Carlo methods for policy evaluation and control. Temporal Difference (TD) learning: Q-learning and SARSA. Deep Q-Networks (DQN) for deep reinforcement learning. Policy gradient methods and actor-critic algorithms. Multi-armed bandits and exploration exploitation trade off

Module 5: Advanced Topics in Machine Learning [6L]

Transfer learning and domain adaptation techniques. Model interpretability and explainable AI. Hyperparameter tuning and model selection strategies. Time series analysis and forecasting. Anomaly detection and outlier detection methods. Recommendation systems: collaborative filtering and content based filtering. Adversarial machine learning and security implications.

Text/Reference Books:

1.P. Flach, "Machine Learning: The Art and Science of Algorithms that Make Sense of Data", Cambridge University Press.

2.M. Mohri, A. Rostamizadeh and A. Talwalkar, "Foundations of Machine Learning", MIT Press.

3.K. P. Murphy, "Machine Learning: A Probabilistic Perspective", MIT Press.

CO-PO Mapping:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	**3**	1	1	2	3	2	0	0	3	-	-	3
CO2	**2**	2	1	1	1	1	-	-	-	-	-	3
CO3	**1**	3	1	2	2	1	-	-	1	-	**-**	3
CO4	2	2	2	1	1	1	-	-	1	-	-	3
CO5	**2**	1	1	1	**2**	2	-	-	2	-	**-**	3

Course Code	PDT3002			
Course Title	Big Data Analytics			
Category	Professional			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites	a)Database Management Systems b) Data Science			

Learning Objective:

This course gives an overview of Big Data, i.e. storage, retrieval and processing of big data. In addition, it also focuses on the "technologies", i.e., the tools/algorithms that are available for storage, processing of Big Data. It also helps a student to perform a variety of "analytics" on different data sets and to arrive at positive conclusions.

Course Outcome:

CO1: Outline the importance of Big Data Analytics

CO2: Apply statistical techniques for Big data Analytics

CO3: Analyze problems appropriate to mining data streams.

CO4: Apply the knowledge of clustering techniques in data mining

CO5: Apply Hadoop map Reduce programming for handing Big Data.

Course Content:

Module 1: Introduction to Big Data [5L]

Evolution of Big data - Best Practices for Big data Analytics - Big data characteristics - Validating - The Promotion of the Value of Big Data - Big Data Use Cases- Characteristics of Big Data Applications - Perception and Quantification of Value -Understanding Big Data Storage - Evolution Of Analytic Scalability - Analytic Processes and Tools - Analysis vs Reporting - Modern Data Analytic Tools - Statistical Concepts: Sampling Distributions - Re-Sampling - Statistical Inference - Prediction Error.

Module 2: Data Analysis, Clustering And Classification [9L]

Regression Modeling - Multivariate Analysis - Bayesian Modeling - Support Vector and Kernel Methods - Analysis of Time Series: Linear Systems Analysis - Nonlinear Dynamics - Rule Induction. Overview of Clustering - K-means - Use Cases - Overview of the Method-Determining the Number of Clusters - Diagnostics - Reasons to Choose and Cautions .- Classification: Decision Trees - Overview of a Decision Tree - The General Algorithm - Decision Tree Algorithms - Evaluating a Decision Tree - Decision Trees in R - Naïve Bayes - Bayes' Theorem - Naïve Bayes Classifier.

Module 3: Stream Memory [6L]

Introduction To Streams Concepts – Stream Data Model and Architecture - Stream Computing - Sampling Data in a Stream – Filtering Streams – Counting Distinct Elements in a Stream – Estimating Moments – Counting Oneness in a Window – Decaying Window - Real time

Analytics Platform(RTAP) Applications - Case Studies - Real Time Sentiment Analysis, Stock Market Predictions.

Module 4: Association And Graph Memory [8L]

Advanced Analytical Theory and Methods: Association Rules - Overview - Apriori Algorithm - Evaluation of Candidate Rules - Applications of Association Rules - Finding Association& finding similarity - Graph Analytics for Big Data: Graph Analytics - The Graph Model - Representation as Triples - Graphs and Network Organization - Choosing Graph Analytics - Graph Analytics Use Cases - Graph Analytics Algorithms and Solution Approaches - Technical Complexity of Analyzing Graphs- Features of a Graph Analytics Platform..

Module 5: Frameworks And Visualization [8L]

MapReduce – Hadoop, Hive, MapR – Sharding – NoSQL Databases - S3 - Hadoop Distributed File Systems – Visualizations - Visual Data Analysis Techniques - Interaction Techniques; Systems and Analytics Applications - Analytics using Statistical packages-Approaches to modeling in Analytics – correlation, regression, decision trees, classification, association Intelligence from unstructured information-Text analytics-Understanding of emerging trends and Technologies-Industry challenges and application of Analytics- Analyzing big data with twitter - Big data for E-Commerce Big data for blogs - Review of Basic Data Analytic Methods using R.

Text/Reference Books:

1. D. Loshin, "Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph", 2013.
2. A. Rajaraman and J. D. Ullman, “Mining of Massive Datasets”, Cambridge University Press, 2012.
3. M. Berthold, D. J. Hand, “Intelligent Data Analysis”, Springer, 2007.
4. B. Baesens, "Analytics in a Big Data World: The Essential Guide to Data Science and its Applications", Wiley Publishers, 2015.
5. K. H. Pries and R. Dunnigan, "Big Data Analytics: A Practical Guide for Managers " CRC Press, 2015

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	3	-	-	-	-	-	-	-	-	-	-
CO2	3	2	-	**1**	2	-	-	-	-	-	-	1
CO3	3	2	2	3	2	-	-	-	-	-	-	-
CO4	3	3	2	1	2	-	-	-	-	-	-	-
CO5	3	2	3	3	3	-	-	-	-	-	-	1

Course Code	PDT3003			
Course Title	Data visualization			
Category	Professional			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites	a) Data Structures and Algorithms b) Operating Systems c) Discrete Structures			

Learning Objective:

To understand the basic concepts of Data science, Data Visualization; ability to apply AI and Data Science in different domain; and do exploratory analysis on a given data.

Course Outcome:

CO1: Demonstrate understanding of Data Visualization and key Terms

CO2: Design Effective Data Visualization for viual Mapping and Design

CO3: Will demonstrate skills on creating visual representation of Data

CO4: Will demonstrate understanding of Visualization classification and its techniques

CO5: Will demonstrate skills in creating different types of representation

Course Content:

Module 1: Introduction to Data Visualization and Tool Overview [6L]

Introduction to Data Visualization: Importance, principles, and types of visualizations, Overview of Tableau: Features, interface, and capabilities for data exploration and visualization, Overview of Power BI: Features, components, and advantages for creating interactive dashboards and reports.

Module 2: Data Connection and Preparation [8L]

Connecting to Data Sources: Importing data from Excel, CSV, databases, and cloud services, Data Preparation: Cleaning, transforming, and shaping data for analysis and visualization, Data Joins and Blending: Combining data from multiple sources using joins and blending techniques.

Module 3: Basic Visualization Techniques [8L]

Chart Types: Bar charts, line charts, scatter plots, pie charts, and heatmaps, Customizing Visualizations: Formatting axes, colors, labels, and annotations, Adding Interactivity: Filters, parameters, and actions to enhance user interaction.

Module 4: Advanced Visualization Techniques [8L]

Advanced Charts and Graphs: Treemaps, box plots, histograms, and dual-axis charts, Dashboard Design rinciples: Layout, composition, and storytelling techniques, Dashboard Interactivity: Drill-downs, tooltips, and parameter actions for interactive dashboards.

Module 5: Data Analysis and Insights [10L]

Exploratory Data Analysis (EDA): Identifying patterns, trends, and outliers in data, Statistical

Analysis: Calculating summary statistics, correlations, and distributions, Insights and Actionable Findings: Communicating insights and making data-driven decisions.

Module 6: Advanced Topics and Integration [10L]

Advanced Visualization Techniques: Animations, forecasting, and custom scripting, Integration with Other Tools: Embedding visualizations in web applications, integrating with R and Python for advanced analytics, Best Practices and Future Trends: Optimization techniques, accessibility, and emerging trends in data visualization.

Text/Reference Books:

1. C. O'Neil and R. Schutt, "Doing Data Science, Straight Talk from The Frontline", O'Reilly. 2014.
2. P. Bruce and A. Bruce, "Practical Statistics for Data Scientists", O'Reilly Media, 2017.
3. G. James, D. Witten, T. Hastie, and R. Tibshirani, "An Introduction to Statistical Learning: with Applications in R", Springer,2013.
4. J. M. Chambers, "Software for Data Analysis: Programming with R (Statistics and Computing)", Springer.

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	**3**	1	2	2	**-**	-	-	-	-	-	-	2
CO2	1	1	2	-	2	-	-	-	-	-	-	3
CO3	**1**	1	2	-	2	-	-	-	-	-	-	3
CO4	**1**	1	3	2	1	-	-	-	-	-	-	3
CO5	**1**	2	2	-	3	2	-	-	-	-	-	3

Course Code	PDT3004			
Course Title	Operations Research and Optimization Techniques			
Category	Professional Core			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites	a) Mathematics b) Fundamentals of Programming			

Learning Objective:

In this course the students will learn about the basic knowledge of LPP, duality, transportation problem, assignment problem, game theory, queueing and inventory models. At the end of the course, the students will get knowledge about various decision making through operations research models.

Course Outcome:

CO1: To explain linear programming problems and appreciate their limitations

CO2: To analyze and solve linear programming problems using appropriate techniques and optimization solvers

CO3: To conduct and interpret post-optimal and sensitivity analysis and explain the primal-dual relationship

CO4: To develop mathematical skills to analyze and solve transportation, assignment problem and network models arising from a wide range of applications

CO5: To share and communicate ideas, explain procedures and interpret results and solutions in written and electronic forms to different audiences

Course Content:

Module 1: Linear Programming Problem [10L]

Linear Programming Problem (LPP): Basics of LPP and its applications. General mathematical formulation of LPP. Solution of LPP by Graphical Method, Simplex Method, Charnes' Big M-Method; Duality Theory.

Module 2: Transportation Problem and Assignment Problem [5L]

Transportation Problem, Assignment Problem – problem solving.

Module 3: Game Theory and Networking model [9L]

Game Theory: Introduction; Two-person Zero Sum game, Saddle Point; Mini-Max and Maxi-Min Theorems (statement only) and problems; Games without Saddle Point; Graphical Method; Principle of Dominance.

Network Optimization Models: CPM, PERT, Time estimates, earliest expected time, latest allowable occurrence time, latest allowable occurrence time and stack. Critical path, Probability of meeting scheduled date of completion of project. Calculation of CPM network.

Module 4: Sequencing and Queuing Theory [8L]

Sequencing: Johnson's Algorithm (1957) For n Jobs and two machines, n Jobs and three machines.

Queuing Theory: introduction and basic structure; Birth-and-Death Model (Poisson / Exponential distribution); Poisson Queue Models: (M/M/1) :(∞/FIFO) and (M/M/1) :(N/FIFO) and Problems.

Module 5: Decision Making [4L]

Introduction, Individual decision making, Multi person decision making, Multi criteria decision making

Text/Reference Books:

1. K. Swaroop and P. K. Manmohan, "Operations Research", Sultan Chand and Sons.
2. J. G. Chakraborty and P. R. Ghosh, "Linear Programming and Game Theory", Central Book Agency.
3. P. M. Karak, "Linear Programming and Theory of Games", ABS Publishing House
4. D. K. Jana and T. K. Roy, "Operations Research", Chhaya Prakashani Pvt. Ltd.
5. H. A. Taha, "Operations Research", Pearson.
6. J. K. Sharma, "Operations Research Theory and Applications", Macmillan India.

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	2	-	-	-	-	-	-	-	-	-	1
CO2	3	2	2	-	-	-	-	-	-	-	-	1
CO3	3	3	2	-	-	-	-	-	-	-	-	1
CO4	3	3	2	-	-	-	-	-	-	-	-	1
CO5	3	3	2	-	-	-	-	-	-	-	-	1

Course Code	PDT3005			
Course Title	Time Series Analysis			
Category	Professional			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites	a) Probability and Statistics b) Design and Analysis of Algorithms			

Learning Objective:

Present time series in an informative way, both graphically and with summary statistics, Model time series to analyses the underlying structure(s) in both the time and frequency domai

Course Outcome:

CO1: Forecast the trend pattern exhibited by the given data by using various methods

CO2: Run and interpret time series models and regression models for time series

CO3: Use the Box-Jenkins approach to model and forecast time series data empirically.

CO4: Analyze and estimate the cyclic components using special processes.

CO5: To implement the concept in real life problem

Course Content:

Module 1: Introduction to Trend [8L]

Introduction to times series data, application of time series from various fields, Components of a time series, Decomposition of time series. Trend: Estimation of trend by free hand curve method, method of semi averages, fitting a various mathematical curve, and growth curves.

Module 2: Trend and Seasonal Component [10L]

Method of moving averages. Detrending. Effect of elimination of trend on other components of the time series. Seasonal Component: Estimation of seasonal component by Method of simple averages, Ratio to Trend, Ratio to moving average and Link relatives.

Module 3: Forecasting [8L]

Variate component method: Stationary Time series: Weak stationary, auto correlation function and correlogram of moving average .Forecasting: Exponential smoothing methods, Short term forecasting methods: Brown's discounted regression, Box-Jenkins Method.

Module 4: Cyclic Component [10L]

Deseasonalization. Cyclic Component: Harmonic Analysis. Some Special Processes: Moving-average (MA) process and Autoregressive (AR) process of orders one and two, Estimation of the parameters of AR (1) and AR (2) – Yule-Walker equations.

Text/Reference Books:

1. Kendall M.G. (1976): Time Series, Charles Griffin.

2. Chatfield C. (1980): The Analysis of Time Series –An Introduction, Chapman & Hall.
3. Mukhopadhyay P. (2011): Applied Statistics, 2nd ed. Revised reprint, Books and Allied

CO-PO Mapping:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	**3**	1	1	2	3	2	0	0	3	-	-	3
CO2	**2**	2	1	1	1	1	-	-	-	-	-	3
CO3	**1**	3	1	2	2	1	-	-	1	-	**-**	3
CO4	2	2	2	1	1	1	-	-	1	-	-	3
CO5	**2**	1	1	1	**2**	2	-	-	2	-	**-**	3

Course Code	PDT3101			
Course Title	Machine Learning Lab			
Category	Professional Core			
LTP & Credits	L	T	P	Credits
	0	0	3	1.5
Total Contact Hours	36			
Pre-requisites				

Learning Objective:

It covers some of the important regression, classification, clustering, rule-based and probabilistic models and algorithms. The Themes included linear and logistic regression, regularization, decision trees, probabilistic, SVMs and neural networks, clustering and reduction in feature dimensionality.

Course Outcome:

CO1: Understand the Basic operations of Linear Algebra in Machine Learning

CO2: Use various Supervised Learning techniques like Linear Regression and Nonlinear Regression

CO3: Apply Statistical approaches for multiple Learning techniques

CO4: Construct models for Classification

CO5: Build neural network models

Suggestive List of Experiments:

1. Write a Program to perform the following operations on matrices
a) Matrix addition
b) Matrix Subtraction
c) Matrix Multiplication
d) Matrix Inversion
e) Transpose of a Matrix
2. Write a Program to perform the following operations
a) Find the minimum and maximum element of the matrix
b) Find the minimum and maximum element of each row in the matrix
c) Find the minimum and maximum element of each column in the matrix
d) Find trace of the given matrix
e) Find rank of the given matrix
f) Find eigenvalues and eigenvectors of the given matrix.

3. Write a Program to find the mean, median, standard deviation and mode using user defined functions
4. Create a data frame with columns at least 5 observations
a) Retrieve a particular column from the Data Frame
b) Summarize the data frame and observe the statistics of the Data Frame created
c) Observe the mean and standard deviation of the data frame and print the values.
5. Write a program to implement the Linear Regression for a sample training data set stored as

a .CSV file. Compute Mean Square Error by considering few test data sets..

6. Write a program to implement the Non-linear Regression for a sample training data set stored as a .CSV file. Compute Mean Square Error by considering few test data sets.

7. Write a program to implement the Logistic Regression for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier.

8. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.

9. Write a program to implement k-Nearest Neighbor algorithm to classify the iris data set. Print both correct and wrong predictions.

10. Write a program to implement Support Vector Machine algorithm to classify the iris data set. Print both correct and wrong predictions.

11. Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.

12. Write a program to demonstrate the working of the decision tree based CART algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.

13. Write a program to construct a Regression tree for cost estimation by assuming any numerical dataset.

14. Write a program to calculate the accuracy, precision, and recall for your data set. Assume a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task.

15. Implement a single neural network and test for different logic gates.

16. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.

Text / Reference Books:

1. Vijayvargia, Abhishek, Machine Learning with Python: An Approach to Applied Machine Learning, BPB Publications, 1st edition,2018.

2. Aurelien Geron, Hands-On Machine Learning with Scikit-Learn and TensorFlow, Oreilly, March 2017.

3. Dr. M Gopal, Applied Machine Learning, 1st Edition, McGraw-Hill,2018

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	1	2	1	1	2	-	-	-	-	-	-	-
CO2	2	1	3	2	3	1	-	-	-	-	-	1
CO3	1	1	1	1	1	-	-	-	-	-	-	-
CO4	2	1	2	2	3	-	-	-	-	-	-	-
CO5	1	2	1	1	1	-	-	-	-	-	-	-

Course Code	PDT3102			
Course Title	Big Data Algorithms & Analytics Lab			
Category	Professional Core			
LTP & Credits	L	T	P	Credits
	0	0	3	1.5
Total Contact Hours	36			
Pre-requisites	None			

Learning Objective:

This course gives an overview of Big Data, i.e. storage, retrieval and processing of big data. In addition, it also focuses on the "technologies", i.e., the tools/algorithms that are available for storage, processing of Big Data. It also helps a student to perform a variety of "analytics" on different data sets and to arrive at positive conclusions.

Course Outcome:

CO1: Outline the importance of Big Data Analytics

CO2: Apply statistical techniques for Big data Analytics

CO3: Analyze problems appropriate to mining data streams.

CO4: Apply the knowledge of clustering techniques in data mining

CO5: Apply Hadoop map Reduce programming for handing Big Data.

Suggestive List of Experiments:

1. Install, configure and run python, numPy and Pandas.
2. Install, configure and run Hadoop and HDFS.
3. Visualize data using basic plotting techniques in Python.
4. Implement NoSQL Database Operations: CRUD operations, Arrays using MongoDB.
5. Implement Functions: Count – Sort – Limit – Skip – Aggregate using MongoDB.
6. Implement word count / frequency programs using MapReduce.
7. Implement a MapReduce program that processes a dataset.
8. Implement clustering techniques using SPARK.
9. Implement an application that stores big data in MongoDB / Pig using Hadoop / R.

Text / Reference Books:

1. Mark Dexter, Louis Landry, "Joomla Programming", 2012 Pearson Education.
2. Seema Acharya and Subhashini C, "Big Data and Analytics", Wiley Publication, 2015.
3. Judith Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman, "Big data for dummies", Wiley Publication, 2013.
4. Tom White, "Hadoop: The Definitive Guide", O"Rilly Publication, 2015.

5. Chuck Lam, “Hadoop in action”, Dreamtech Press, 2011.
6. Dirk Deroos, Paul C. Zikopoulos, Roman B. Melnyk, Bruce Brown, “Hadoop for dummies”, Wileypublication, 2015.

CO-PO Mapping:

PO-CO Mapping	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO10	PO11	PO12
CO1	2	2	2	2	3	1	-	-	1	1	1	1
CO2	2	2	2	2	3	1	-	-	1	1	1	1
CO3	2	2	2	2	2	1	-	-	1	1	1	1
CO4	2	2	2	2	3	1	-	-	1	1	1	1
CO5	2	2	2	2	3	1	-	-	1	1	1	1

Course Code	PDT3103			
Course Title	Data Visualization Lab			
Category	Professional Core			
LTP & Credits	L	T	P	Credits
	0	0	3	1.5
Total Contact Hours	36			
Pre-requisites	None			

Learning Objective:

To understand the need of different datasets visualization
Can perform descriptive and inferential analysis on data sets
Can create visualization using modern tool
Can develop analytical products
Identify opportunities for application of data visualization in various domains

Course Outcome:

CO1: Design and create data visualizations.
CO2: Conduct exploratory data analysis using visualization.
CO3: Craft visual presentations of data for effective communication.
CO4: Use knowledge of perception and cognition to evaluate visualization design alternatives.
CO5: Use JavaScript with D3.js to develop interactive visualizations for the Web.

Suggestive List of Experiments:

1. Defining data visualization

2. Visualization workflow: describing data visualization workflow, process in practice.

3. Data Representation: chart types: categorical, hierarchical, relational, temporal & spatial.

4. 2-D: bar charts, Clustered bar charts, dot plots, connected dot plots, pictograms, proportional shape charts, bubble charts, radar charts, polar charts, Range chart, Box-and-whisker plots, univariate scatter plots, histograms word cloud, pie chart, waffle chart, stacked bar chart, back-to-back bar chart, treemap and all relevant 2-D charts.

5. 3-D: surfaces, contours, hidden surfaces, pm3d coloring, 3D mapping

6. Multi-dimensional data visualization

7. Manifold visualization

8. Graph data visualization

Text / Reference Books:

1. A. Kirk, “Data Visualization A Handbook for Data Driven Design, Sage Publications, 2016.
2. P. K. Janert, “Gnuplot in Action, Understanding Data with Graphs”, Manning Publications, 2010.
3. E. Siegel, T. H. Davenport, “Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die”, Wiley, 2013

4. A.Cordoba, “Understanding the Predictive Analytics Lifecycle”, Wiley, 2014.
5. J. R. Evans, “Business Analytics – Methods, Models and Decisions”, Pearson 2013.
6. R. N. Prasad, S. Acharya, “Fundamentals of Business Analytics”, Wiley, 2015

CO-PO Mapping:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	**3**	1	2	2	**-**	-	-	-	-	-	-	2
CO2	1	1	2	-	2	-	-	-	-	-	-	3
CO3	**1**	1	2	-	2	-	-	-	-	-	-	3
CO4	**1**	1	3	2	1	-	-	-	-	-	-	3
CO5	**1**	2	2	-	3	2	-	-	-	-	-	3

Semester 4
Curriculum and Detailed Syllabus

SEMESTER-4							
Sl. No.	**Type**	**Course No.**	**Course Name**	**L**	**T**	**P**	**Credits**
THEORY							
1		PDT4001	Deep learning	3	0	0	3
2		**Elective II**		3	0	0	3
		PDT4002	Predictive Analytics				
		PDT4003	Natural Language Processing				
PRACTICAL							
3		PDT4101	Deep Learning Lab	0	0	3	1.5
		PDT4002/4003		0	0	3	1.5
4		PDT4105	Project / Dissertation Work	0	5	10	10
5		JSC4502	Seminar / Group Discussion	1	0	0	1
6		JSC4503	Skill X	0	0	0	1
7		JSC4504	NSS/2Physical Activities /Meditation Yoga/ Photography	0	0	3	1
TOTAL				**6**	**5**	**13**	**19**

Course Code	PDT4001			
Course Title	Deep Learning			
Category	Professional Elective			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites	a) Machine Learning			

Course Outcome:

CO1: Make use of deep learning APIs like Keras

CO2: Implement multiple conversions for Analysis

CO3: Apply deep learning techniques for object identification and segmentation

CO4: Implement RNN and CNN for multiple problems

CO5: Implement Autoencoders and GAN.

Course Content:

Module 1: Introduction **[8L]**

Introduction to deep learning, Neural Network Basics, Backpropagation, Feed forward Neural Network, Logistic Regression.

Module 2: Key Concepts **[6L]**

Key concepts on Deep Neural Networks, Shallow Neural Network, Planar data classification with a hidden layer, Building your Deep Neural Network: step by step

Module 3: Optimization **[8L]**

Hyperparameter Tuning, Batch Normalization, Regularization, Gradient Checking.

Generative Adversarial Networks: Practical aspects of deep learning, Generative Adversarial Networks (GAN), Conditional GAN, Super Resolution GAN, Cycle GAN.

Module 4: Deep Reinforcement Learning **[8L]**

Deep Reinforcement Learning, Hyperparameter Tuning, Batch Normalization.

Module 5: Convolutional Neural Network **[6L]**

Foundations of Convolutional Neural Network, Deep Convolutional Models.

Text/Reference Books:

1. I. Goodfellow, Y. Bengio and A. Courville, “Deep Learning”, MIT Press.
2. C. M. Bishop, “Pattern Recognition and Machine Learning”, Springer.

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	**3**	1	2	2	**-**	-	-	-	-	-	-	2
CO2	1	1	2	-	2	-	-	-	-	-	-	3
CO3	**1**	1	2	-	2	-	-	-	-	-	-	3
CO4	**1**	1	3	2	1	-	-	-	-	-	-	3
CO5	**1**	2	2	-	3	2	-	-	-	-	-	3

Course Code	PDT4002			
Course Title	Predictive Analytics			
Category	Professional Elective			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites				

The course enables students to: To learn, how to develop models to predict categorical and continuous outcomes, using such techniques as neural networks, decision trees, logistic regression, support vector machines and Bayesian network models. To know the use of the binary classifier and numeric predictor nodes to automate model selection. To advice on when and how to use each model. Also learn how to combine two or more models to improve prediction

Course Outcome:

CO1: Understand the process of formulating business objectives, data selection/collection, preparation

CO2: Understand the process to successfully design, build, evaluate and implement predictive models for a various business application.

CO3: Compare the underlying predictive modeling techniques.

CO4: Select appropriate predictive modeling approaches to identify cases to progress with.

CO5: Apply predictive modeling approaches using a suitable package such as SPSS Modeler

Course Content:

Module 1: [8L]

Introduction to Data Mining Introduction, what is Data Mining? Concepts of Datamining, Technologies Used, Data Mining Process, KDD Process Model, CRISP – DM, Mining on various kinds of data, Applications of Data Mining, Challenges of Data Mining.

Module 2: [10L]

Data Understanding and Preparation Introduction, Reading data from various sources, Data visualization, Distributions and summary statistics, Relationships among variables, Extent of Missing Data. Segmentation, Outlier detection, Automated Data Preparation, Combining data files, Aggregate Data, Duplicate Removal, Sampling DATA, Data Caching, Partitioning data, Missing Values.

Module 3: [10L]

Model development & techniques Data Partitioning, Model selection, Model Development Techniques, Neural networks, Decision trees, Logistic regression, Discriminant analysis, Support vector machine, Bayesian Networks, Linear Regression, Cox Regression, Association rules.

Module 4: [8L]

Model Evaluation and Deployment Introduction, Model Validation, Rule Induction Using CHAID, Automating Models for Categorical and Continuous targets, Comparing and Combining Models, Evaluation Charts for Model Comparison, MetaLevel Modeling, Deploying Model, Assessing Model Performance, Updating a Model.

Text/Reference Books:

1.Eric Siegel, Predictive analytics- the power to predict who will Click, buy, lie, or die, John Wiley & Sons, 2013.

2. Dean Abbott, Applied Predictive Analytics - Principles and Techniques for the Professional Data Analyst, 2014.

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	**3**	1	2	2	**-**	-	-	-	-	-	-	2
CO2	1	1	2	-	2	-	-	-	-	-	-	3
CO3	**1**	1	2	-	2	-	-	-	-	-	-	3
CO4	**1**	1	3	2	1	-	-	-	-	-	-	3
CO5	**1**	2	2	-	3	2	-	-	-	-	-	3

Course Code	PDT4003			
Course Title	Natural Language Processing			
Category	Professional Elective			
LTP & Credits	L	T	P	Credits
	3	0	0	3
Total Contact Hours	36			
Pre-requisites	a) Design and Analysis of Algorithms b) Compiler Design			

Learning Objective:

In this course, the students will learn about the various Natural Language techniques that are essential to understand how to build Language Processing systems. In particular, various security applications shall be discussed as case studies.

The course will be very helpful for the students in strengthening their basic knowledge in Language Processing.

Course Outcome:

CO 1: Use the NLTK and spaCy toolkit for NLP Programming.
CO 2: Analyze various corpora for developing programs.
CO 3: Develop various pre-processing techniques for a given corpus.
CO 4: Develop programming logic using NLTK functions.
CO 5: Build applications using various NLP techniques for a given corpus.

Course Content:

Module 1: Introduction to NLP [7L]

Introduction: Human languages, models, ambiguity, processing paradigms; Phases in natural language processing, applications. Text representation in computers, encoding schemes. Linguistics resources- Introduction to corpus, elements in balanced corpus, TreeBank, PropBank, WordNet, VerbNet etc. Resource management with XML.

Module 2: Management of Linguistic Data [7L]

Management of linguistic data with the help of GATE, NLTK. Regular expressions, Finite State Automata, word recognition, lexicon. Morphology, acquisition models, Finite State Transducer. N-grams, smoothing, entropy, HMM, ME, SVM, CRF.

Module 3: Speech Tagging and Applications [8L]

Part of Speech tagging: Stochastic POS tagging, HMM, Transformation based tagging (TBL), Handling of unknown words, named entities, multi word expressions. A survey on natural language grammars, lexeme, phonemes, phrases and idioms, word order, agreement, tense, aspect and mood and agreement, Context Free Grammar, spoken language syntax.

Module 4: Parsing [8L]

Parsing: Unification, probabilistic parsing, TreeBank. Semantics: Meaning representation, semantic analysis, lexical semantics, WordNet Word Sense Disambiguation- Selection restriction, machine learning approaches, dictionary based approaches.

Module 5: Discourse and applications of NLP [6L]

Discourse: Reference resolution, constraints on co-reference, algorithm for pronoun resolution, text coherence, discourse structure. Applications of NLP: Spell-checking, Summarization Information Retrieval- Vector space model, term weighting, homonymy, polysemy, synonymy, improving user queries. Machine Translation– Overview.

Text / Reference Books:

1. D. Jurafsky and J. H. Martin. "Speech and Language Processing", Pearson Education.
2. A. James, "Natural Language Understanding", Pearson Education.
3. A. Bharati, R. Sangal and V. Chaitanya, "Natural Language Processing: a Paninian Perspective", Prentice-Hall of India.
4. T. Siddiqui and U. S. Tiwary, "Natural language processing and Information Retrieval", OUP.

CO-PO Mapping:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	**3**	1	2	2	**-**	-	-	-	-	-	-	2
CO2	1	1	2	-	2	-	-	-	-	-	-	3
CO3	**1**	1	2	-	2	-	-	-	-	-	-	3
CO4	**1**	1	3	2	1	-	-	-	-	-	-	3
CO5	**1**	2	2	-	3	2	-	-	-	-	-	3

Course Code	PDT4101			
Course Title	Deep Learning Lab			
Category	Professional Core			
LTP & Credits	L	T	P	Credits
	0	0	3	1.5
Total Contact Hours	36			
Pre-requisites	None			

Learning Objective:

To understand the need of different datasets visualization
Can perform descriptive and inferential analysis on data sets
Can create visualization using modern tool
Can develop analytical products
Identify opportunities for application of data visualization in various domains

Course Outcome:

At the end of the Course the student shall be able to

CO1: Make use of deep learning APIs like Keras

CO2: Implement multiple conversions for Analysis

CO3: Apply deep learning techniques for object identification and segmentation

CO4: Implement RNN and CNN for multiple problems

CO5: Implement Autoencoders and GAN.

Suggestive List of Experiments:

1. Build a deep neural network model start with linear regression using a single variable.
2. Build a deep neural network model start with linear regression using multiple variables.
3. Write a program to convert speech into text.
4. Write a program to convert text into speech.
5. Write a program to convert video into frames.
6. Write a program for Time-Series Forecasting with the LSTM Model.
7. Build a feed forward neural network for prediction of logic gates.
8. Write a program to implement deep learning Techniques for image segmentation.
9. Write a program for object detection using image labeling tools.
10. Write a program to predict a caption for a sample image using LSTM.
11. Write a program for character recognition using CNN.
12. Write a program to predict a caption for a sample image using CNN.
13. Write a program for character recognition using RNN and compare it with CNN.
14. Write a program to detect Dog image using YOLO Algorithm.
15. Write a program to develop Autoencoders using MNIST Handwritten Digits.
16. Write a program to develop a GAN for Generating MNIST Handwritten Digits.

Text / Reference Books:

1. Navin Kumar Manaswi ,Deep Learning with Applications Using Python Chatbots

and Face, Object, and Speech Recognition With TensorFlow and Keras , Apress,2018.
2. Ian Goodfellow, Yoshua Bengio, Aaron Courville, “Deep Learning”, MIT Press,2016.
3. Josh Patterson and Adam Gibson, “Deep learning: A practitioner's approach”, O'Reilly Media, First Edition, 2017.

CO-PO Mapping:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	**3**	1	2	2	**-**	-	-	-	-	-	-	2
CO2	1	1	2	-	2	-	-	-	-	-	-	3
CO3	**1**	1	2	-	2	-	-	-	-	-	-	3
CO4	**1**	1	3	2	1	-	-	-	-	-	-	3
CO5	**1**	2	2	-	3	2	-	-	-	-	-	3

Course Code	PDT4102			
Course Title	Predictive Analytics Lab			
Category	Professional Core			
LTP & Credits	L	T	P	Credits
	0	0	3	1.5
Total Contact Hours	36			
Pre-requisites	None			

The course enables students to: To learn, how to develop models to predict categorical and continuous outcomes, using such techniques as neural networks, decision trees, logistic regression, support vector machines and Bayesian network models. To know the use of the binary classifier and numeric predictor nodes to automate model selection. To advice on when and how to use each model. Also learn how to combine two or more models to improve prediction

Course Outcome:

CO1: Understand the process of formulating business objectives, data selection/collection, preparation

CO2: Understand the process to successfully design, build, evaluate and implement predictive models for a various business application.

CO3: Compare the underlying predictive modeling techniques.

CO4: Select appropriate predictive modeling approaches to identify cases to progress with.

CO5: Apply predictive modeling approaches using a suitable package such as SPSS Modeler

Suggestive List of Experiments:

Following experiments to be carried out using Python/SPSS/SAS/R/Power BI

1. Simple Linear regression
2. Multiple Linear regression
3. Logistic Regression
4. CHAID
5. CART
6. ARIMA – stock market data
7. Exponential Smoothing
8. Hierarchical clustering
9. Ward's method of clustering
10. Crowdsource predictive analytics- Netflix data

Text / Reference Books:

1. Eric Siegel, Predictive analytics- the power to predict who will Click, buy, lie, or die, John Wiley & Sons, 2013.

2. Dean Abbott, Applied Predictive Analytics - Principles and Techniques for the Professional Data Analyst, 2014.

CO-PO Mapping:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	**3**	1	2	2	**-**	-	-	-	-	-	-	2
CO2	1	1	2	-	2	-	-	-	-	-	-	3
CO3	**1**	1	2	-	2	-	-	-	-	-	-	3
CO4	**1**	1	3	2	1	-	-	-	-	-	-	3
CO5	**1**	2	2	-	3	2	-	-	-	-	-	3

Course Code	PDT4103			
Course Title	Natural Language Processing Lab			
Category	Professional Core			
LTP & Credits	L	T	P	Credits
	0	0	3	1.5
Total Contact Hours	36			
Pre-requisites	None			

Learning Objective:

In this course, the students will learn about the various Natural Language techniques that are essential to understand how to build Language Processing systems. In particular, various security applications shall be discussed as case studies.

The course will be very helpful for the students in strengthening their basic knowledge in Language Processing.

Course Outcome:

CO 1: Use the NLTK and spaCy toolkit for NLP Programming.

CO 2: Analyze various corpora for developing programs.

CO 3: Develop various pre-processing techniques for a given corpus.

CO 4: Develop programming logic using NLTK functions.

CO 5: Build applications using various NLP techniques for a given corpus.

Suggestive List of Experiments:

1. Installation and exploring features of NLTK and spaCy tools. Download Word Cloud and few corpora.
2. (i)Write a program to implement word Tokenizer, Sentence and Paragraph Tokenizers.
(ii) Check how many words are there in any corpus. Also check how many distinct words are there?
3. (i) Write a program to implement both user-defined and pre-defined functions to generate
    (a) Uni-grams
    (b) Bi-grams
    (c) Tri-grams
    (d) N-grams

(ii) Write a program to calculate the highest probability of a word (w2) occurring after another word(w1).
4. (i)Write a program to identify the word collocations.
(ii) Write a program to print all words beginning with a given sequence of letters.
(iii) Write a program to print all words longer than four characters.
5. (i) Write a program to identify the mathematical expression in a given sentence.
(ii) Write a program to identify different components of an email address.
6. (i) Write a program to identify all antonyms and synonyms of a word.
(ii) Write a program to find hyponymy, homonymy, polysemy for a given word.

7. (i) Write a program to find all the stop words in any given text.
(ii) Write a function that finds the 50 most frequently occurring words of a text that are not stop words.
8. Write a program to implement various stemming techniques and prepare a chart with the performance of each method.
9. Write a program to implement various lemmatization techniques and prepare a chart with the performance of each method.
10. (i) Write a program to implement Conditional Frequency Distributions(CFD) for any corpus. (ii) Find all the four-letter words in any corpus. With the help of a frequency distribution (FreqDist), show these words in decreasing order of frequency. (iii) Define a conditional frequency distribution over the names corpus that allows you to see which initial letters are more frequent for males versus females.
11. (i) Write a program to implement Part-of-Speech (PoS) tagging for any corpus. (ii) Write a program to identify which word has the greatest number of distinct tags? What are they, and what do they represent? (iii) Write a program to list tags in order of decreasing frequency and what do the 20 most frequent tags represent? (iv) Write a program to identify which tags are nouns most commonly found after? What do these tags represent?
12. Write a program to implement TF-IDF for any corpus.
13. Write a program to implement chunking and chinking for any corpus.
14. (i) Write a program to find all the mis-spelled words in a paragraph. (ii) Write a program to prepare a table with frequency of mis-spelled tags for any given text.
15. Write a program to implement all the NLP Pre-Processing Techniques required to perform further NLP tasks.

Text / Reference Books:

1.D. Jurafsky and J. H. Martin. “Speech and Language Processing”, Pearson Education.
2.A. James, “Natural Language Understanding”, Pearson Education.
3.A. Bharati, R. Sangal and V. Chaitanya, “Natural Language Processing: a Paninian Perspective”, Prentice-Hall of India.
4.T. Siddiqui and U. S. Tiwary, “Natural language processing and Information Retrieval”, OUP.

CO-PO Mapping:

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	**3**	1	2	2	**-**	-	-	-	-	-	-	2
CO2	1	1	2	-	2	-	-	-	-	-	-	3
CO3	**1**	1	2	-	2	-	-	-	-	-	-	3
CO4	**1**	1	3	2	1	-	-	-	-	-	-	3
CO5	**1**	2	2	-	3	2	-	-	-	-	-	3