VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

SCHOOL OF ADVANCED SCIENCES

DEPARTMENT OF MATHEMATICS

M.Sc. Data Science

(MDT)

Curriculum & Syllabi

(2024 -2025 Admitted Students)

VISION STATEMENT OF VELLORE INSTITUTE OF TECHNOLOGY

Transforming life through excellence in education and research.

MISSION STATEMENT OF VELLORE INSTITUTE OF TECHNOLOGY

- **World class Education**: Excellence in education, grounded in ethics and critical thinking, for improvement of life.
- **Cutting edge Research**: An innovation ecosystem to extend knowledge and solve critical problems.
- **Impactful People**: Happy, accountable, caring and effective workforce and students.
- **Rewarding Co-creations**: Active collaboration with national & international industries & universities for productivity and economic development.
- **Service to Society**: Service to the region and world through knowledge and compassion.

VISION STATEMENT OF SCHOOL OF ADVANCED SCIENCES

To be an internationally renowned science school in research and innovation by imparting futuristic education relevant to the society.

MISSION STATEMENT OF SCHOOL OF ADVANCED SCIENCES

- To nurture students from India and abroad by providing quality education and training to become scientists, technologists, entrepreneurs and global leaders with ethical values for a sustainable future.

- To enrich knowledge through innovative research in niche areas.

- To ignite passion for science and provide solutions for national and global challenges.

M.Sc. Data Science

PROGRAMME EDUCATIONAL OBJECTIVES (PEOs)

PEO_01: Graduates will demonstrate proficiency with statistical analysis of Data.

PEO_02: Graduates will execute statistical analyses with professional statistical software.

PEO_03: Graduates will demonstrate skill in Data management.

PEO_04: Graduates will develop the ability to build and assess Data based models.

PEO_05: Graduates will apply data science concepts and methods to solve problems in real-world contexts and will communicate these solutions effectively.

M.Sc. Data Science

PROGRAMME OUTCOMES (POs)

PO_01: Having a clear understanding of the subject related concepts and of contemporary issues.

PO_02: Having problem solving ability to address social issues.

PO_03: Having a clear understanding of professional and ethical responsibility.

PO_04: Having cross cultural competency exhibited by working in teams.

PO_05: Having a good working knowledge of communicating in English.

M.Sc. Data Science

PROGRAMME SPECIFIC OUTCOMES (PSOs)

On completion of M.Sc. Data Science programme, graduates will be able to

PSO1: Exhibit statistical competence in data analysis to develop and evaluate data-based models.

PSO2: Specialises in developing AI/ML and optimisation tools for decision-making using professional data science software.

PSO3: Undertake the research necessary to create data science solutions that delve into analytics and the data lifecycle.

M.Sc. Data Science

PROGRAMME CREDIT STRUCTURE

Category	Credits
Discipline Core Courses	28
Skill Enhancement Courses	05
Discipline Elective Courses	21
Open Elective Courses	06
Project/ Internship	20
Total Credits	**80**

M.Sc. Data Science

DETAILED CURRICULUM

Discipline Core Courses		28			
		L	**T**	**P**	**C**
PMDS501L	Applied Linear Algebra	3	0	0	3
PMDS502L	Probability and Distribution Models	3	0	0	3
PMDS503L	Statistical Inference	2	0	0	2
PMDS503P	Statistical Inference Lab	0	0	2	1
PMDS504L	Regression Analysis and Predictive Models	3	0	0	3
PMDS504P	Regression Analysis and Predictive Models Lab	0	0	2	1
PMDS505L	Data Mining and Machine Learning	3	0	0	3
PMDS505P	Data Mining and Machine Learning Lab	0	0	2	1
PMDS506L	Database Management Systems	3	0	0	3
PMDS506P	Database Management Systems Lab	0	0	2	1
PMDS507L	Big-Data Analytics	2	0	0	2
PMDS507P	Big-Data Analytics Lab	0	0	2	1
PMDS508L	Python Programming	2	0	0	2
PMDS508P	Python Programming – Lab	0	0	4	2

M.Sc. Data Science

DETAILED CURRICULUM

Discipline Elective Courses		21			
		L	T	P	C
PMDS601L	Artificial Intelligence	3	0	0	3
PMDS601P	Artificial Intelligence Lab	0	0	2	1
PMDS602L	Advanced Machine Learning	3	0	0	3
PMDS602P	Advanced Machine Learning Lab	0	0	2	1
PMDS603L	Deep Learning	3	0	0	3
PMDS603P	Deep Learning Lab	0	0	2	1
PMDS604L	Exploratory Data Analysis	2	0	0	2
PMDS604P	Exploratory Data Analysis Lab	0	0	2	1
PMDS605L	Data Structures and Algorithms	3	0	0	3
PMDS605P	Data Structures and Algorithms Lab	0	0	2	1
PMDS606L	Natural Language Processing	3	0	0	3
PMDS607L	Optimization Techniques	3	0	0	3
PMDS608L	Healthcare Analytics	2	0	0	2
PMDS608P	Healthcare Analytics Lab	0	0	2	1
PMDS609L	Block chain Technology	2	0	0	2
PMDS610L	Financial Analytics	2	0	0	2
PMDS610P	Financial Analytics Lab	0	0	2	1
PMDS611L	Multivariate Data Analysis	3	0	0	3
PMDS611P	Multivariate Data Analysis Lab	0	0	2	1
PMDS612L	Statistics for Managers	3	0	0	3

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

M.Sc. Data Science

DETAILED CURRICULUM

Open Elective Courses	06
Engineering Disciplines | Social Sciences	

Skill Enhancement Courses		05			
		L	**T**	**P**	**C**
PENG501P	Technical Report Writing	0	0	4	2
PSTS501P	Qualitative Skills Practice	0	0	3	1.5
PSTS501P	Qualitative Skills Practice	0	0	3	1.5

Project and Internship		20			
		L	**T**	**P**	**C**
PMDS696J	Study Oriented Project	0	0	0	2
PMDS697J	Research Project	0	0	0	2
PMDS698J	Internship I / Dissertation I	0	0	0	4
PMDS699J	Internship II/ Dissertation II	0	0	0	12

Indicative Programme Structure of Master of Science in Data Science

First Year

SEMESTER 1				SEMESTER 2			
Course Code	**Course**	**L-T-P**	**C**	**Course Code**	**Course**	**L-T-P**	**C**
PMDS501L	Applied Linear Algebra	3-0-0	3	PMDS503L	Statistical Inference	2-0-0	2
PMDS502L	Probability and Distribution Models	3-0-0	3	PMDS503P	Statistical Inference Lab	0-0-2	1
PMDS508L	Python Programming	2-0-0	2	PMDS504L	Regression Analysis and Predictive Models	3-0-0	3
PMDS508P	Python Programming – Lab	0-0-4	2	PMDS504P	Regression Analysis and Predictive Models Lab	0-0-2	1
PMDS506L	Database Management Systems	3-0-0	3	PMDS505L	Data Mining and Machine Learning	3-0-0	3
PMDS506P	Database Management Systems Lab	0-0-2	1	PMDS505P	Data Mining and Machine Learning Lab	0-0-2	1
PMDS601L	Artificial Intelligence	3-0-0	3	PMDS604L	Exploratory Data Analysis	2-0-0	2
PMDS601P	Artificial Intelligence Lab	0-0-2	1	PMDS604P	Exploratory Data Analysis Lab	0-0-2	1
PMDS696J	Study Oriented Project	0-0-0	2	PMDS605L	Data Structures and Algorithms	3-0-0	3
PENG501P	Technical Report Writing	0-0-4	2	PMDS605P	Data Structures and Algorithms Lab	0-0-2	1
PSTS501P	Qualitative Skills Practice	0-0-3	1.5	PSTS502P	Qualitative Skills Practice	0-0-3	1.5
				PMDS697J	Research Project	0-0-0	2
					OE1/NPTEL		3
Total Credits			**23.5**	Total Credits			**24.5**

Second Year

SEMESTER 3				SEMESTER 4			
Course Code	**Course**	**L-T-P**	**C**	**Course Code**	**Course**	**L-T-P**	**C**
PMDS507L	Big-Data Analytics	2-0-0	2	PMDS699J	Internship II/ Dissertation II	0-0-0	12
PMDS507P	Big-Data Analytics Lab	0-0-2	1				
PMDS603L	Deep Learning	3-0-0	3				
PMDS603P	Deep Learning Lab	0-0-2	1				
PMDS606L	Natural Language Processing	3-0-0	3				
PMDS607L	Optimization Techniques	3-0-0	3				
PMDS698J	Internship I / Dissertation I	0-0-0	4				
	OE2/NPTEL		3				
Total Credits			**20**	Total Credits			**12**

Discipline Core

Course Code	Course Title	L	T	P	C
PMDS501L	APPLIED LINEAR ALGEBRA	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. Understanding the matrix algebra and its applications.
2. Solving computational problems.
3. Provide an introduction to vectors - matrices - and least square methods – all basic topics in linear algebra - in the context of data science.
4. Application to various real-life data-driven problems.

Course Outcomes

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

1. Understand basic matrix properties like rank - determinant - inverse and special type of matrices.
2. Use computational techniques for singular value decomposition and g-inverses.
3. Understand the concepts of vector space - subspaces and linear transformation.
4. Compute inner products on a real vector space and compute angle and orthogonality in inner product spaces.
5. Apply the methods of linear algebra and matrices in several important - modern applications of research and industrial problems involving data science.

Module:1	Introduction to Matrices	5 hours

Algebra of Matrices - Trace - Rank of a Matrix and Their Properties - Row Echelon Form - Determinants and Their Properties - Inverse - Symmetric - Orthogonal and Idempotent Matrices and Their Properties - Eigen Values and Eigen Vectors.

Module:2	Systems of Linear Equations	5 hours

Cramer's Rule - Pivoting - Row Canonical Form - Diagonal Form - Gauss and Gauss-Jordan Elimination - Elementary Matrices - Solution of Systems of Linear Equations.

Module:3	Matrix Factorization and Decomposition	8 hours

LU Decomposition - Matrix Decompositions for PCA And Least Squares – Eigen Decomposition - QR Decomposition - Singular Value Decomposition - Quadratic Forms and Related Results.

Module:4	Vector Spaces	6 hours

Vector Spaces - Subspaces - Linear Dependence and Linear Independence - Spanning Set - Basis and Dimension of a Vector Space.

Module:5	Linear transformation	6 hours

Linear Transformation - Kernel - Range - Matrix Representation of a Linear Transformation - Rank-Nullity Theorem - Change Of Basis And Similar Matrices.

Module:6	Inner product spaces	5 hours

Inner-Product Spaces - Orthogonal Sets and Bases - Orthogonal Projection - Gram-Schmidt Orthogonalization Process.

Module:7	Applications in Data Science	8 hours

Generalized Inverses (G-Inverses) - Methods of Constructing G-Inverses - General Solution to A System of Linear Equations. Sparse Matrices - Linear Discriminant Analysis and Canonical Correlation Analysis. Conceptual Discussion of Classical Cryptosystems –Plain Text - Cipher Text - Encryption – Decryption.

Module:8	Contemporary Issues	2 hours

	Total Lecture hours:	**45 hours**

Text Book(s)	
1	Gilbert Strang, Introduction to Linear Algebra, 2023, 6th Edition, Wellesley-Cambridge Press, U.S.
2	David C. Lay, Linear Algebra and Its Applications, 2019, 5th Edition, Pearson.
Reference Book(s)	
1	Friedberg, S., Insel, A. and Spence, L., Linear Algebra, 2019, 5th Edition, Pearson.
2	Nick Fieller, Basics of Matrix Algebra for Statistics with R, 2015, CRC Press.

Mode of Evaluation: CAT, Quiz, Assignment and FAT.			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS502L	PROBABILITY AND DISTRIBUTION MODELS	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To incorporate the concepts of probability theory and its applications as the core material in building theoretical ideas along with the practical notion.
2. To integrate the intrinsic ideas of preliminary and advanced distributions to correlate with the real-world scenarios.

Course Outcomes

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

1. Develop the problem-solving techniques needed to calculate probability and conditional probability.
2. Describe and construct the probability distribution functions and illustrate the mathematical expectation.
3. Demonstrate the various types of generating functions used in statistics.
4. Apply the commonly used univariate discrete and continuous probability distributions.
5. Illustrate the sampling distributions and its importance in Inferential statistics.

Module:1	Probability	4 hours

Introduction – Random Experiments, Empirical basis of probability, Algebra of events, laws of probability; Conditional Probability, Independence, Bayes' law; Application of probability to business and economics.

Module:2	Random Variables	7 hours

One-dimensional Random variable- Discrete and Continuous; Distribution functions and its properties; Bivariate Random Variables- Joint Probability functions, marginal distributions, conditional distribution functions; Notion of Independence of Random variables. Functions of random variables: introduction, distribution function technique, transformation technique: one variable, transformation technique: several variables, theory and applications.

Module:3	Mathematical Expectation	9 hours

Expectation, Variance, and Co-variance of random variables; Conditional expectation and conditional variance; Markov, Holder, Jensen and Chebyshev's Inequality; Weak Law of Large numbers, Strong law of large numbers and Kolmogorov theorem; Central Limit Theorem.

Module:4	Generating Functions	4 hours

Moment Generating Function, Characteristic Function and Probability Generating Function - Properties and Applications.

Module:5	Discrete Distributions	8 hours

Bernoulli, Binomial, Poisson, Geometric, Hyper-geometric, Negative Binomial, Multinomial, distributions and Discrete Uniform distribution - definition, properties and applications with numerical problems.

Module:6	Continuous Distributions	8 hours

Uniform, Normal distribution function, Exponential, Gamma, Beta distributions (First and Second kind), Weibull, Cauchy and Laplace distribution functions - definition,

properties and applications, concept of truncated distributions.		
Module:7	**Sampling Distributions**	**3 hours**
Introduction, The sampling distribution of the Mean: Finite Populations, Sampling distribution of the proportion: Finite Populations, distribution of sample variance, Chi-square distribution, t- distribution, F distribution, order statistics: properties and its applications.		
Module:8	**Contemporary Issues**	**2 hours**
	Total Lecture hours	**45 hours**

Text Book(s)	
1	Sheldon M. Ross, A First course in Probability, 2020, 10th Edition, Pearson.
2	R.V. Hogg, J. W. McKean, and Allen T. Craig, An Introduction to Mathematical Statistics, 2019, 8th Edition, Pearson Education.
Reference Book(s)	
1	Rohatgi, V.K. and Ebsanes Saleh, A.K. Md., An Introduction to Probability and Statistics, 2002, 2nd Edition, John Wiley & Sons.
2	Krishnamoorthy, K., Handbook of Statistical Distributions with Applications, 2006, Chapman & Hall/CRC.
3	Gupta, S.C. and Kapoor V.K., Fundamentals of Mathematical Statistics, 2020, Sultan Chand & sons.
4	Maurits Kaptein, Edwin van den Heuvel, Statistics for Data Scientists: An Introduction to Probability and Statistics and Data Analysis, 2022, Springer.

Mode of Evaluation: CAT, Quiz, Assignment and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS503L	Statistical Inference	2	0	0	2
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

- Understand the types of questions that the statistical method addresses for decision making.
- Apply statistical methods to hypotheses testing and inference problems.
- Interpret the results in a way that addresses the question of interest.
- Communicate the purposes of the analyses, the findings from the analysis, and the implications of those findings.

Course Outcomes

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

1. Describe the criteria required of good point estimators, and check whether or not a proposed estimator within a stated statistical model satisfies these criteria.
2. Apply the principle of maximum likelihood, minimum variance and moment estimation methods to obtain point and interval estimates of parameters in one-parameter and multi-parameter statistical models.
3. Construct the hypothesis tests in some common models (including Normal models), correctly using the terms null hypothesis, alternative hypothesis, test statistic, rejection region, significance level, power and p-value.
4. Apply the parametric (Z, t, F, Chi-square) tests and interpret the results.
5. Develop the non-parametric tests, with due regard to the underlying assumptions.

Module:1	Introduction to Estimation	6 hours

Population, sample, parameter and statistic; characteristics of a good estimator; Consistency – Invariance property of Consistent estimator, Sufficient condition for consistency; Unbiasedness; Sufficiency – Factorization Theorem – Minimal sufficiency; Efficiency – Most efficient estimator, likelihood equivalence and uniformly minimum variance unbiased estimator.

Module:2	Methods of Estimation	6 hours

Methods of point estimation – Maximum likelihood estimation, method of minimum variance, method of moment estimator, concept of BLUE.

Module:3	Interval Estimation	2 hours

Methods of Interval estimation - Confidence limits and confidence coefficient, Construction of confidence intervals for population parameters.

Module:4	Testing of hypotheses	2 hours

Null Hypothesis, Alternative Hypothesis, Types of errors, power of a test, most powerful tests; Neyman-Pearson Fundamental Lemma and its applications; Uniformly most powerful tests; Likelihood Ratio tests.

Module:5	**Large sample tests**	**4 hours**
Large sample properties; Tests of significance (under normality assumption)- Test for a population mean, proportion; Test for equality of two means and proportions; Test for variance**.** Sequential Probability Ratio Test**.**		
Module:6	**Small sample tests**	**4 hours**
Student's t-test, test for a population mean, equality of two population means, paired t-test, F-test for equality of two population variances; Chi-square test for goodness of fit - test for independence of attributes.		
Module:7	**Non-parametric tests**	**4 hours**
Sign test, Signed rank test, Median test, Mann-Whitney -test, Run test, Kolmogorov –Smirnov test and Kruskal – Wallis-H-test**.**		
Module:8	**Contemporary issues**	**2 hours**
	Total Lecture hours	**30 hours**

Text Book(s)	
1	Robert V Hogg, Elliot A Tannis and Dale L. Zimmerman, Probability and Statistical Inference, 2020, 10th Edition, Pearson.
2	Manoj Kumar Srivastava and Namita Srivastava, Statistical Inference- Testing of Hypotheses, 2014, Prentice Hall of India.
Reference Book(s)	
1	Rajagopalan M and Dhanavanthan P, Statistical Inference, 2012, PHI Learning.
2	B. K. Kale and K. Muralidharan, Parametric Inference, 2016, Narosa Publishing House.
3	Marc S. Paolella, Fundamental statistical inference: A computational approach, 2018, Wiley.

Mode of evaluation: CAT, Assignment , Quiz and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS503P	Statistical Inference Lab	0	0	2	1
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives:

1. Understand the types of questions that the statistical method addresses for decision-making.
2. Apply statistical methods to hypotheses testing and inference problems.
3. Interpret the results in a way that addresses the question of interest.
4. Communicate the purposes of the analyses, the findings from the analysis, and the implications of those findings.

Course Outcomes:

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

1. Apply the concept hypothesis tests in some common models (including Normal models) and construct the null hypothesis, alternative hypothesis, test statistic, rejection region, significance level and develop the power curve and p-value.
2. Evaluate the parametric (Z, t, F, Chi-square) hypothesis testing and interpret the results.
3. Evaluate the non-parametric tests, with due regard to the underlying assumptions and interpret the results.

Indicative Experiments

1	Confidence intervals, *p*-value
2	Large Sample Tests- Test for Population mean & Population proportions
3	Small Sample Tests – t – test for population mean, Paired t test
4	F- test for population variances
5	Chi-square test for goodness of fit and Independence of Attributes
6	Sign Test
7	Mann Whitney Wilcoxon Test
8	Median Test and Run test
9	Kolmogorov – Smirnov test
10	Kruskal – Wallis-H-test
	Total Laboratory hours **30 hours**

Text Book(s)

1	Chester Ismay, Albert Y. Kim, Statistical Inference via Data Science: A Modern Dive into R, 2019, CRC Press.
2	Robert V Hogg, Elliot A Tannis and Dale L. Zimmerman, Probability and Statistical Inference, 2021, 9th Edition, Pearson publishers.

Reference Book(s)

1	Stefano Bonnini, Livio Corain, Marco Marozzi, Luigi Salmaso, Nonparametric Hypothesis Testing, 2014, John Wiley & Sons Ltd.
2	Rajagopalan M and Dhanavanthan P, Statistical Inference, 2012, PHI Learning.

Mode of evaluation: Assignment and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS504L	REGRESSION ANALYSIS AND PREDICTIVE MODELS	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. Understand the notions of regression and time series model building.
2. Impart application of regression and time series models in various domains.
3. Instruct the methodology to test assumptions and conditions involved in regression and time series models.

Course Outcomes

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

1. Understand the simple linear regression.
2. Apply the tests for assumption for checking normality and homoscedasticity.
3. Analyse the data using multiple linear and non-linear regression models
4. Apply an appropriate time series forecasting method in any given situation.
5. Analyse model validation of time series forecasting techniques.

Module:1	Simple Linear Regression	4 hours

Simple Regression Models with One Independent Variable - Assumptions, Estimation of Parameters - Standard Error of Estimator - Testing the Significance of Regression Coefficients - Standard Error of Prediction.

Module:2	Multiple Linear Regression	7 hours

Multiple Regression: Standard Gauss Markov Setup, Least Square (LS) Estimation, Error and Estimation Spaces - Variance - Covariance of LS Estimators - Estimation of Error Variance - Case with Correlated Observations - LS Estimation with Restriction on Parameters - Multicollinearity.

Module:3	Diagnostics	7 hours

Diagnostic Checks and Correction: Graphical Techniques, Tests for Normality, Uncorrelatedness, Homoscedasticity, Lack of Fit - Polynomial Regression - Transformations on Y or X - Inverse Regression.

Module:4	Nonlinear Regression	6 hours

Non-Linear Regression: Linearization Transforms, Advantages, Limitations, Non-Linear Least Squares, Parameter Estimation in a Non-Linear Systems - Generalized Linear Models: Logistic Regression, Poisson Regression.

Module:5	Introduction To Time Series Analysis	5 hours

Graphical Display - Classical Decomposition Model - Components and Various Decompositions of Time Series Models - Data Transformations - Methods of Estimation Trend, Seasonal and Exponential.

Module:6	Stationary Time Series Models	7 hours

Stationary and types of Stationary - White Noise Processes - Autocorrelation Function - Partial Autocorrelation Function and their Standard Errors - Autoregressive Model - Moving Average Model - Autoregressive Moving Average Model - Autoregressive Integrated Moving Average Model.

Module:7	**Non-Stationary time series models**	**7 hours**
Tests For Non-Stationarity: Random Walk, Unit Root Tests: Dickey Fuller Test, Augmented Dickey Fuller Test - ARIMA Models: Basic Formulation of The ARIMA Model and their Statistical Properties - Forecasting Model Selection Techniques: AIC, BIC And AICC Forecasting Model Monitoring.		
Module:8	**Contemporary Issues**	**2 hours**
	Total Lecture hours	**45 hours**

Text Book(s)	
1	Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, Introduction to Linear Regression Analysis, 2016, 3rd Edition, Wiley India Pvt. Ltd., New Delhi.
2	Douglas C. Montgomery, Cheryl L. Jennings, Murat Kulahci, Introduction to TimeSeries Analysis and Forecasting, 2016, 2nd Edition, Wiley India Pvt. Ltd., New Delhi.
Reference Book(s)	
1	George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung, Time Series Analysis: Forecasting and Control, 2016, 5th Edition, Wiley India Pvt. Ltd., New Delhi.
2	Norman R. Draper, Harry Smith, Applied Regression Analysis, 2015, 3rd Edition, Wiley India Pvt. Ltd., New Delhi.

Mode of Evaluation: CAT, Assignment, Quiz and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course code	Course Title	L	T	P	C
PMDS504P	REGRESSION ANALYSIS AND PREDICTIVE MODELS LAB	0	0	2	1
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. Understand the notions of regression and time series model building.
2. Impart application of regression and time series models in various domains.
3. Instruct the methodology to test assumptions and conditions involved in regression and time series models

Course Outcomes

On the successful completion of the course, student will be able to:

1. Understand the Simple Linear Regression
2. Apply the tests for assumption for checking normality and homoscedasticity.
3. Analyse the data using Multiple linear and non-linear regression models
4. Apply an appropriate time series forecasting method in any given situation.
5. Analyse model validation of time series forecasting techniques.

Indicative Experiments

1	Simple linear regression: model fitting, estimation of parameters, computing R^2 and adjusted R^2 and model interpretation.
2	Residual analysis and forecast accuracy for a given data set.
3	Validating Simple linear regression using t, F and p- test.
4	Developing confidence interval and testing the model simple and multiple regression.
5	Multiple regression: estimation of parameters, fitting of the model, error analysis, model validation, variable selection and testing.
6	Problem of multicollinearity and, determination of VIF. Diagnostic measures and outlier detection, Durbin Watson test, variable selection and model building
7	Visualization of Stationary and Non-stationary time series
8	Moving Average Time Series Model and Differencing
9	Exponential smoothing technique (Single, double and triple)
10	Auto-Regressive Model for Stationary Time Series
11	Autoregressive Integrated Moving Average for Non-Stationary Time Series
12	Forecasting model validations
	Total Laboratory hours **30 hours**

Text Book(s)

1	Ali S. Hadi, Samprit Chatterjee, Regression Analysis by Example Using R, 2023, 6th Edition, John Wiley & Sons, New York.
2	Chris Chatfield, Haipeng Xing, The Analysis of Time Series: An Introduction with R, 2019, 7th Edition, CRC Press.

Reference Book(s)

1	Manu Joseph, Modern Time Series Forecasting with Python, 2022, 1st Edition, Packt Publishing Ltd, United Kingdom.

Mode of assessment: Assignment and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS505L	DATA MINING AND MACHINE LEARNING	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. Understand the role of separate database for decision making.
2. Learn the core ideas of data mining techniques in different case studies.
3. Inculcate the concept learning and Machine learning theory.

Course Outcomes

At the end of the course -students will be able to:

1. Gain knowledge over the importance of KDD and Data Mining
2. Recognize the key areas and issues in data mining.
3. Identify data discrepancies and eliminate anomalies and comprehend different types of learning.
4. Predict the outcome based on regression and compute optimal hyperplane and support vectors for data classification.
5. Analyse the data using the machine learning methods to address social, engineering and business problems.

Module:1	Fundamentals of Data Mining	4 hours

Introduction to data mining - Data types -Measures of similarity and dissimilarity - Data mining tools supervised and unsupervised learning - Classification of Data Ming Systems - Data Mining Task Primitives - Major issues in Data mining.

Module:2	Data Warehousing	4 hours

Data Warehousing Components - Multi-Dimensional Data Model - Data Warehouse Architecture - Data Warehouse Implementation - Mapping the Data Warehouse to Multiprocessor Architecture - OLAP - Need - Categorization of OLAP Tools -Uses of data warehouse.

Module:3	The Ingredients of Machine Learning	8 hours

Machine Learning – Types; Data – Getting the data - visualizing the data - preparing the data; Selecting and Training a Model – Fine tuning a Model: Grid Search – Randomized Search - Main Challenges: Data Inadequacy – Non-representativeness – Irrelevant features – -Bias-Variance- Overfitting the Model – Underfitting the Model.

Module:4	Supervised Learning Techniques	8 hours

Binary Classifier – Performance Measures: Cross –Validation – Confusion Matrix – Precision and Recall – Multiclass classification – Mutli-label classification; Linear Regression – Gradient Descent: Batch Gradient – Stochastic Gradient Descent – Mini-batch Gradient Descent; Polynomial Regression –Logistic Regression – Bayes Classification - Estimating Probabilities -Decision Boundaries -Softmax Regression.

Module:5	Ensemble Machine Learning	6 hours

Linear SVM with Soft Margin Classification – Non-linear SVM Classification: Polynomial features –Similarity features – Gaussian Kernel; SVM Regression. Decision Trees and Random Forests: Training and Visualizing a Decision Tree – CART Algorithm – Gini Impurity; Bagging – Pasting – Random Forests – Boosting: Adaboost and Gradient Boosting – Stacking - Explainability.

Module:6	Dimensionality Reduction	6 hours

Main approaches – Projection and Manifold Learning – PCA (Principal Component Analysis): Preserving the Variance – Principal Components – Projecting down to

Dimensions – Randomized PCA – Kernel PCA.		
Module:7	**Unsupervised Learning Techniques**	**7 hours**
Clustering: K-means Clustering – Limitations – Clustering for Image Segmentation - Preprocessing - Semi supervised learning – DBSCAN – Hierarchical – Partitional - Gaussian Mixtures.		
Module:8	**Contemporary Issues**	**2 hours**
	Total Lecture hours	**45 hours**

Text Book(s)	
1	Alpaydin Ethem, Introduction to Machine Learning, 2019, 3rd Edition, PHI Learning Private Limited.
2	Mohammed J. Zaki and Wagner Meira, Jr., Data Mining and Machine Learning: Fundamental Concepts and Algorithms, 2020, 2nd Edition, Cambridge University Press.
Reference Book(s)	
1	Balas K Natarajan, Machine Learning, 2014, Elsevier Science.
2	Deisenroth, Marc Peter, A., Aldo Faisal and Cheng Soon Ong., Mathematics for machine learning, 2019, Cambridge University Press.

Mode of Evaluation: CAT, Assignment, Quiz and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course code	Course Title	L	T	P	C
PMDS506P	DATA MINING AND MACHINE LEARNING LAB	0	0	2	1
Pre-requisite	NIL	Syllabus Version			
		1.0			

Course Objectives

1. Understand the implementation procedures for the machine learning algorithms using Matlab /R/Python, Weka (Machine Learning software in JAVA).
2. Understand modern notions in data analysis-oriented computing and conduct experiments to design a component or a product applying all the relevant standards with realistic constraints.

Course Outcomes

At the end of the course the student should be able to:

1. Understand the most popular machine learning algorithms.
2. Analyse and perform an evaluation of learning algorithms and model selection.
3. Compare the strengths and weaknesses of many popular machine learning approaches.
4. Appreciate the underlying mathematical relationships within and across machine learning algorithms and the paradigms of supervised and unsupervised learning.
5. Analyze and implement various machine learning algorithms in a range of real-world applications.

List of Challenging Experiments

1.	Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.
2.	Implement linear regression using python. Select appropriate data set for your experiment and plot the graphs
3.	Write a program to construct a Logistic Regression considering classification data.
4.	Implement the SVM using classification data.
5.	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
6.	Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions.
7.	Implement k-means clustering for classification.
8.	Implementation of Time Series Clustering and alignment algorithms
9.	Implement Principal Component Analysis (PCA) for dimensionality reduction.
10	Implement Linear Discriminant Analysis (LDA) for dimensionality reduction.
	Total Laboratory hours **30 hours**

Text Book(s)			
1	Mehryar Mohri, Afshin Rostamizadeh and Ameet Talwalkar, Foundations of Machine Learning, 2018, 2/Ed., MIT Press.		
Reference Book(s)			
1	Christopher Bishop, Pattern Recognition and Machine Learning, 2013, Springer.		
2	Balas K Natarajan, Machine Learning, 2014, Elsevier Science.		
Mode of assessment	Digital Assignment and FAT		
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS506L	Database Management Systems	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To understand the basic concepts of database, ER Modelling, normalization and query optimization.
2. To comprehend the concepts concurrency control, recovery and indexing.
3. To explore the concepts of NoSQL and main types of NoSQL databases.

Course Outcomes

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

1. Describe the concepts of database, construct entity-relationship (ER) model for the real world problems and transfer data model into database designs.
2. Analyze the fundamental concepts of normalization, transaction, concurrency control and recovery mechanisms.
3. Demonstrate the basic database storage structure and indexing techniques.
4. Organize the detailed architecture and primary benefits using NoSQL Databases.
5. Analyze the major types of NoSQL databases.

Module:1	DATABASE SYSTEMS CONCEPTS AND DATA MODELING	7 hours

Basic concepts of database systems- Entity Relationship Model - Structural Constraints- Relational Model- Relational Model Constraints- Mapping ER model to a Relational Schema and database integrity.

Module:2	DATABASE DESIGN AND QUERY PROCESSING	7 hours

Guidelines for Relational Schema- Functional Dependency- Normalization- Translating SQL Queries into Relational Algebra- Heuristic Query Optimization.

Module:3	TRANSACTION PROCESSING CONCEPTS	6 hours

Introduction to transaction processing- Transaction and system concepts - Desirable properties of transactions- Characterizing schedules based on recoverability- Characterizing schedules based on serializability- Test for serializability.

Module:4	CONCURRENCY CONTROL AND PHYSICAL DATABASE DESIGN	6 hours

Lock-based protocols- Techniques for concurrency control- Recovery concepts- File organization- and Indexing.

Module:5	NOSQL	6 hours

Database revolutions: First generation, second generation, third generation- Managing transactions and data integrity- ACID and BASE for reliable database transactions- Speeding performance by strategic use of RAM, SSD, and disk, Brewer's CAP theorem.

Module:6	KEY VALUE DATA STORES	6 hours

Essential features of key value databases- Key-Value architecture- Designing structured values- Limitations of key-value databases - Design patterns for key-value databases and Case study for Key-Value databases.

Module:7	**NOSQL DATA MODEL**	**5 hours**
Aggregate models- Document data model- Key- value data model- Columnar data model and Graph based data model.		
Module:8	**Contemporary Issues**	**2 hours**
	Total Lecture hours	**45 hours**

Text Book(s)	
1	Raghu Ramakrishnan, Johannes Gehrke, Database Management Systems, 2015, 4th Edition, Tata McGraw Hill.
Reference Book(s)	
1	Henry F Korth, Abraham Silberschatz, S. Sudharshan, Database System Concepts, 2006, 5th Edition, McGraw Hill.
2	R. Elmasri and S. B. Navathe, Fundamentals of Database Systems, 2016, 7th Edition, Addison Wesley.
3	Guy Harrison, Next Generation database: NoSQL New SQL and Big Data, 2015, 1st Edition, Apress.
4	Daniel G. McCreary and Ann M. Kelly, Making Sense of NoSQL, 2013, Manning publisher.

Mode of Evaluation: CAT, Assignment, Quiz and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course code	Course Title	L	T	P	C
PMDS506P	DATABASE MANAGEMENT SYSTEMS LAB	0	0	2	1
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To understand the concepts of SQL commands.
2. To comprehend the concepts of PL/SQL.
3. To explore the concepts of NoSQL using MongoDB.

Course Outcomes

At the end of the course the student should be able to:

1 Perform SQL and PL/ SQL queries.
2 Implement the MongoDB methods, insert, update, delete operations in a NoSQL database

Indicative Experiments

1	Basic SQL Commands (DDL, DML, DCL ,TCL and Constraints)
2	Operators, Views and Functions
3	Joins and Subqueries
4	PL/SQL Introduction and Control Structures
5	Exception handling
6	Functions and Procedures
7	Cursors and Triggers
8	Basics of MongoDB, Methods and operators
9	Working with documents and collections
10	Indexing in MongoDB
	Total Laboratory hours — **30 hours**

Text Book(s)

1	Manu sharma, MongoDB Complete Guide, 2021, BPB Publications.
2	David Hows, Peter Membrey , Eelco Plugge, DUPTim Hawkins, The Definitive Guide to MongoDB: A complete guide to dealing with Big Data using MongoDB, 2015, Aress.

Reference Book(s)

1	Rick Copeland, MongoDB Applied Design Patterns: Practical Use Cases with the Leading NoSQL Database, 2013, O'Reilly.
2	Amit Phaltankar, Juned Ahsan, Michael Harrison and Liviu Nedov, MongoDB Fundamentals, A hands-on guide to using MongoDB and Atlas in the real world, 2020, Packt Publishing.

Mode of evaluation	Assignment and FAT		
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course code	Course Title	L	T	P	C
PMDS507L	BIG DATA ANALYTICS	2	0	0	2
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To have knowledge on accessing, storing and manipulating the huge data from different resources.
2. To learn map reduce analytics using Hadoop and related tools.
3. To work with map reduce applications
4. To understand the working environment of Pig and Hive for processing the structured and unstructured data.
5. To understand the usage of Hadoop related tools for Big Data Analytics

Course Outcomes

At the end of the course the student should be able to:

1. Describe big data and use cases from selected business domains.
2. Install, configure, and run Hadoop and HDFS.
3. Perform map-reduce analytics using Hadoop.
4. Demonstrate the Pig architecture and evaluation of pig scripts
5. Use Hadoop-related tools such as HBase, Cassandra, Pig, and Hive for big data analytics

Module:1	Understanding Big Data	5 hours

Introduction – distributed file system – Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce.

Module:2	Map Reduce Applications	5 hours

MapReduce workflows – unit tests with MRUnit – test data and local tests – anatomy of MapReduce job run – classic Map-reduce – YARN – failures in classic Map-reduce and YARN – job scheduling – shuffle and sort – task execution – MapReduce types – input formats – output formats.

Module:3	Hadoop Architecture	6 hours

Analyzing data with Hadoop – scaling out – Hadoop streaming – Hadoop pipes – design of Hadoop distributed file system (HDFS) – HDFS Administering – Monitoring & Maintenance – Java interface – data flow – Hadoop I/O – data integrity – compression – serialization – Avro – file-based data structures - Cassandra – Hadoop integration.

Module:4	Hadoop Ecosystem and Yarn	4 hours

Hadoop ecosystem components - Schedulers - Fair and Capacity, Hadoop 2.0 New Features- Name Node High Availability, HDFS Federation, MRv2, YARN, Running MRv1 in YARN.

Module:5	Apache Pig	2 hours

Introduction, Parallel processing using Pig, Pig Architecture, Grunt, Pig Data Model-scalar, and complex types. Pig Latin- Input and output, Relational operators, User-defined functions. Working with scripts.

Module:6	**Hive, HiveQL, Hbase and Zookeeper**	**3 hours**
Hive Architecture and Installation, Comparison with Traditional Database, HiveQL - Querying Data - Sorting and Aggregating. HBase concepts- Advanced Usage, Schema Design, Advance Indexing - PIG, Zookeeper - how it helps in monitoring a cluster, HBase uses Zookeeper - Build Applications with Zookeeper.		
Module:7	**Real Time Applications**	**3 hours**
Apache Spark: Eco system, Components of the Spark unified stack-Spark SQL, Spark Streaming, Spark GraphX, Spark MLLib. Spark context, spark stage, spark executor. Spark Architecture, RDD and RDD Operations-RDD Features and limitations, RDDPersistence and Cashing mechanism, DAG, spark cluster management, performance tuning, DataFrames and Dataset – In-memory distributed processing using Apache Spark. Spark shell commands, Spark MLLib for Machine Learning.		
Module:8	**Contemporary Issues**	**2 hours**
	Total Lecture hours	**30 hours**

Text Book(s)	
1	Seema Acharya and Subhashini Chellappan, Big Data and Analytics, 2019, 2nd Edition, Wiley.
2	Michael Minelli, Michelle Chambers and Ambiga Dhiraj, Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, 2013, Wiley.
Reference Book(s)	
1	Judith Hurwitz, Alan Nugent,Fern Halper, Marcia Kaufman, Big data for dummies, 2013, 1st Edition, Wiley Publications.
2	E. Capriolo, D. Wampler, and J. Rutherglen, Programming Hive, 2012, O'Reilley.
3	Tom White, Hadoop: The Definitive Guide, 2015, O'Reilly Media Inc.
4	Alan Gates, Programming Pig, 2011, O'Reilley.
5	Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, Professional Hadoop Solutions, 2015, 1st Edition, Wiley.

Mode of Evaluation: CAT, Assignment, Quiz and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course code	Course Title	L	T	P	C
PMDS507P	Big-Data Analytics Lab	0	0	2	1
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To provide grounding in big data technology.
2. To learn map reduce analytics using Hadoop / Spark and related tools.
3. To work with map reduce applications.
4. To understand the working environment of Pig and Hive for processing the structured and unstructured data.
5. To understand the usage of Hadoop / Spark related tools for Big Data Analytics.

Course Outcome

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

1. Understand Big Data and its analytics in the real world.
2. Analyze the Big Data framework like Hadoop to efficiently store and process Big Data to generate analytics.
3. Design of Algorithms to solve Data Intensive Problems using Map Reduce Paradigm.
4. Design and Implementation of Big Data Analytics using pig and spark to solve data intensive problems and to generate analytics.
5. Implement Big Data Activities using Hive.

Indicative Experiments

1	Understanding different Hadoop modes. Startup scripts, Configuration files.
2	Hadoop / Spark Implementation of file management tasks, such as Adding files and directories, retrieving files and Deleting files.
3	Implement of Matrix Multiplication with Hadoop Map Reduce / Spark
4	Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm.
5	Implementation of K-means Clustering using MapReduce.
6	Generation of Frequent Itemset using MapReduce.
7	Hive: Architecture, Data modeling and data types
8	HBase: HMaster, Region Server and Zookeeper.
9	Application of Recommendation Systems using Hadoop/mahout libraries.
10	Mahout machine learning library to facilitate the knowledge build up in big data analysis.
	Total Laboratory Hours **30 hours**

Text Book(s)				
1	Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, Professional Hadoop Solutions, 2015, Wiley.			
2	Tom White, Hadoop: The Definitive Guide, 2015, O'Reilly Media Inc.			
Reference Book(s)				
1	Judith Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman, Big data for dummies, 2013, 1st Edition, Wiley Publications.			
2	Clinton Gormley, Zachary Tong, Elasticsearch – The Definitive Guide, 2015 1st Edition, O'Reilly Media.			
Mode of Evaluation: Assignment and FAT				
Recommended by Board of Studies		15.02.2024		
Approved by Academic Council		No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS508L	PYTHON PROGRAMMING	2	0	0	2
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To introduce the basic building blocks of algorithmic problem-solving.
2. To introduce core programming basics using Python language.
3. To introduce the data structures of Python and their applications.
4. To introduce the modules for data manipulation and visualization.

Course Outcomes

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

1. Classify various algorithmic approaches and categorize the appropriate data representation,
2. Build programs using control structures,
3. Develop solutions to problems using ordered and un-ordered collection of data types.
4. Utilize the in-built functions and modules and develop user defined functions and modules.
5. Demonstrate array operations, mathematical analysis and graphical representation of data.

Module:1	Algorithmic Problem Solving	2 hours

Building blocks of algorithms: Statements, state, control flow, functions, Developing an Algorithm, Flowchart and Pseudo code.

Module:2	Introduction to Python	3 hours

Introduction to Python - Indentation, variables, reserved words, basic data types: Integer, Floating point, Complex and Boolean; Operators and their precedence, Expressions, Mutability, Built-in Functions, and Importing from Packages.

Module:3	Control Structures	4 hours

Decision Making and Branching: if, if-else, nested if, multi-way if-elif statements; Looping: while-loop, for-loop, else clauses in loops, nested loops, break, continue and pass statements.

Module:4	Data Collections	4 hours

Strings: Comparison, Formatting, Slicing, Splitting, Stripping, Regular Expressions: Matching,Search and replace patterns; Lists, Tuples, Sets and Dictionaries – Operations, List Comprehension.

Module:5	Functions and Modules	5 hours

User-defined functions- parameters and arguments, namespaces and scope rules, Lambda function; Recursive functions, Generator Functions, Decorators. Built-in modules, User-Defined modules,

Module:6	Multidimensional Data Handling and Visualisation	5 hours

NumPy arrays – 1-d, multi-dimensional arrays and matrices. Difference between lists and arrays. Mathematical operations with arrays. Slicing arrays; Boolean masks; Broadcasting in NumPy. Python Plotting: matplotlib – Basic Plotting. Logarithmic Plots. Plots with multiple axes; interactive functions for 3d plotting.

Module:7	**Scientific Data Analysis**	**5 hours**
SciPy – Introduction, scipy.stats, scipy.integrate, scipy.optimize, scipy.interpolate. Pandas – Introduction. Series, DataFrame and Panel. Slicing the data. Reading and writing CSV, XLS and JSON files. Working with missing data, categorical data. Data visualization with Pandas.		
Module:8	**Contemporary Issues**	**2 hours**
	Total Lecture hours	**30 hours**

Text Book(s)	
1	Eric Matthes, Python Crash course: A Hands-On, Project-Based Introduction to Programming, 2023, 3rd edition, William Pollock.
Reference Book(s)	
1	Martic C Brown, Python: The Complete Reference, 2018, 4th Edition, McGraw Hill Publishers.
2	Wes McKinney, Python for Data Analysis, 2022, 3rd Edition, O'Reilly Media.

Mode of Evaluation: CAT, Assignment, Quiz and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course code	Course Title	L	T	P	C
PMDS508P	PYTHON PROGRAMMING LAB	0	0	4	2
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. Explore problem-solving skills using Python programming and find solutions for real-time problems.
2. Acquire object-oriented programming skills in Python.

Course Outcomes

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

1. understand and comprehend the basic programming constructs of Python programming.
2. implement control statements for altering the sequential execution of programs in solving problems.
3. solve real-time problems using modular programming concepts.
4. develop programs for statistical processing of data using NumPy, Matplotlib, Scipy, and Pandas.

Indicative Experiments

1	Build applications using Operators and Expressions.
2	Build applications using Conditional IF-ELIF-ELSE statements).
3	Build applications using Looping (for, while loops).
4	Manipulations using Strings, Lists, Tuple, Sets and Dictionaries.
5	Create user-defined function Python scripts.
6	Create user-defined modules and import them into the programs.
7	Create data applications using array and matrix manipulations.
8	Build basic data visualizations using Matplotlib and interpret them.
9	Build programs to analyze the time series data using the SciPy module.
10	Build programs to manipulate the data and analyze it by Pandas module.
	Total Laboratory Hours **60 hours**

Text Book (s)

1	Reema Thareja, Python Programming using Problem Solving Approach, 2023, 2nd Edition, Oxford University Press.

Reference Book (s)

1	John Hunt, Advanced Guide to Python 3 Programming, 2023, 2nd Edition, Springer Cham.

Mode of evaluation: Assignment and FAT

Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Discipline Elective

Course Code	Course Title	L	T	P	C
PMDS601L	ARTIFICIAL INTELLIGENCE	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To sketch an overview of artificial intelligence (AI) principles and approaches.
2. To develop a basic understanding of the building blocks of AI as presented in terms of intelligent agents: Search, Knowledge representation, inference, logic, and learning.
3. To demonstrate the applications of AI techniques in intelligent agents, expert systems, artificial neural networks and other machine learning models.

Course Outcome

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

1. Gain knowledge of artificial intelligence principles and its foundations, representation and learning.
2. Illustrate the construction of learning and expert system.
3. Formalize a given problem in the language/framework of different AI methods.
4. Apply different search techniques for solving real world complex problems and select the most appropriate solution by comparative evaluation.
5. Attain the capability to represent various real life problem domains using logic-based techniques and use this to perform inference or planning.

Module:1	Introduction to AI	2 hours

Philosophy of artificial intelligence, Definitions - Evolution of AI - Applications of AI, Classification of AI- Intelligent Agents: Agents and Environment-Nature of Environment-Structure Environment

Module:2	Intelligent Agents	4 hours

Rational Agents, Mapping from Sequences to Actions, Properties of Environments, Structure of Intelligent Agents, Types of Agents: Simple Reflex Agents, Goal Based Agents, Utility Based Agents

Module:3	Searching Strategies	8 hours

Problem Solving Agent - Blind Search- Performance measures - Informed Search: Introduction to Heuristics-Variants of heuristic search-uniform cost, A*, Greedy - Overview of Hill Climbing – Simulated Annealing – Genetic Algorithms – Adversarial Search – Minimax, Alpha beta pruning - Constraint Satisfaction Problem.

Module:4	Knowledge Representation and Reasoning	8 hours

Logical Agents-Knowledge-Based Agents- The Wumpus World- Logic-Propositional Logic-Propositional Theorem Proving- First Order Logic- Syntax and Semantics of First-Order Logic, using First order logic, Knowledge Engineering in First-Order Logic. Inference in First Order Logic- Unification and Lifting,

Propositional vs. First order logic-Forward Chaining, Backward chaining, resolution.		
Module:5	**Uncertainty and Knowledge Reasoning**	**7 hours**
Probabilistic Reasoning - Representing Knowledge in an Uncertain Domain, The Semantics of Bayesian Networks, Efficient Representation of Conditional Distributions, Relational and First-Order Probability Models.		
Module:6	**Design of Expert System**	**9 hours**
Introduction to Expert system, Basic concepts, Structure of expert systems, the human element in expert systems, How expert systems works, Problem areas addressed by expert systems, Expert systems success factors, Types of expert systems, Expert systems and the internet interacts web.		
Module:7	**Applications of Artificial Intelligence**	**5 hours**
AI in Business - Health care – Robotics - Social media - Defence – Cyber security.		
Module:8	**Contemporary Issues**	**2 hours**
	Total Lecture hours	**45 hours**

Text Book(s)	
1	Elaine Rich, Kevin Knight, Artificial Intelligence, 2019, 3/Edition, Tata McGraw Hill.
2	Deepak Khemani, A First Course in Artificial Intelligence, 2017, 1/Edition, Tata McGraw Hill Education.
Reference Book (s)	
1	Stuart Russel and Peter Norvig, Artificial Intelligence, 2016, 3rd Edition, Pearson.
2	N.P. Padhy, Artificial Intelligence and Intelligent Systems, 2005, Oxford University Press.
3	Ivan Bratko, PROLOG Programming, 2020, 4th Edition, Pearson Education.

Mode of Evaluation: CAT, Assignment, Quiz and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course code	Course Title	L	T	P	C
PMDS601P	ARTIFICIAL INTELLIGENCE LAB	0	0	2	1
Pre-requisite	NIL	Syllabus Version			
		1.0			

Course Objectives

1. Understand the implementation procedures for the machine learning algorithms using Matlab /R/Python, Weka (Machine Learning software in JAVA).
2. Understand modern notions in data analysis-oriented computing and conduct experiments to design a component or a product applying all the relevant standards with realistic constraints.

Course Outcomes

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

1. Apply appropriate data sets to the Machine Learning algorithms.
2. Identify and apply Machine Learning algorithms to solve real world problems.

List of Challenging Experiments

1	Facts, objects, predicates and variables in PROLOG.
2	Rules and Unification in PROLOG.
3	Arithmetic operators, simple input/output and compound goals in PROLOG.
4	Recursion in PROLOG.
5	Lists in PROLOG.
6	String operations in PROLOG. Implement string operations like substring, string position, palindrome etc.
7	Write a prolog program to implement all set operations (Union, intersection, complement etc in PROLOG.
8	Solving Missionaries and cannibals problems and Water Jug Problem, 8-Queens Problem, Travelling Salesman Problem
9	Wampus Problem using Logic, Monkeys and Bananas Problem using Logic.
10	Development of Medical Expert system with a Recommendation system
	Total Laboratory hours **30 hours**

Text Book(s)

1	Daume, H., A Course in Machine Learning, 2015, Alanna Maldonado.
2	Elaine Rich and Kevin Knight, Artificial Intelligence, 2019, 3rd Edition, Tata McGraw Hill.

Reference Book(s)

1	Christopher Bishop, Pattern Recognition and Machine Learning, 2013, Springer.
2	Balas K Natarajan, Machine Learning, 2014, Elsevier Science.

3	Tom Mitchell, Machine Learning, 2010, McGraw-Hill Education.		
Mode of assessment	Digital Assignment and FAT		
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS602L	ADVANCED MACHINE LEARNING	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. Lay the foundation of machine learning and its practical applications and prepare students for real-time problem-solving in data science.
2. Develop self-learning algorithms using training data to classify or predict the outcome of future datasets.
3. Distinguish overtraining and techniques to avoid it such as cross-validation.

Course Outcomes

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

1. Understand the fundamentals of multivariate methods and non-parametric methods
2. Understand and perform an evaluation of Bayesian Belief Network algorithms.
3. Analyse the Monte-Carlo methods for reinforcement data.
4. Analyse the Eligibility trace and Function Approximation for the real world data.
5. Design and implement various advanced machine learning algorithms in a range of real-world applications.

Module: 1	Multivariate Methods	4 hours

Multivariate Data-Parameter Estimation-Estimation of Missing Values - Expectation-Maximization algorithm - Multivariate Normal Distribution- Multivariate Classification -Tuning Complexity - Discrete Features.

Module: 2	Nonparametric Methods	7 hours

Introduction - Nonparametric Density Estimation: Histogram Estimator - Kernel Estimator - K-Nearest Neighbour Estimator - Generalization to Multivariate Data - Nonparametric classification - Distance Based Classification - Outlier Detection.

Module:3	Bayesian Belief Network	7 hours

Bayesian methods for using prior knowledge and data - Bayesian inference - Bayesian Belief Networks and Graphical models - Probabilistic Latent Semantic Analysis - The Expectation-Maximisation (EM) algorithm - Gaussian Processes.

Module:4	Monte-Carlo Methods and Temporal Difference Learning	6 hours

Monte-Carlo methods: policy evaluation, rollouts, on policy and off-policy learning, importance sampling - Temporal Difference learning: TD prediction, Optimality of TD (0), SARSA, Q-learning, Games and after states, Maximization Bias and Double Learning.

Module:5	Eligibility Traces	4 hours

Eligibility traces: n-step TD prediction, TD (lambda), forward and backward views, Q(lambda), SARSA (lambda), replacing traces and accumulating traces.

Module:6	Function Approximation	8 hours

Function Approximation: Value prediction, gradient descent methods, linear function approximation, Feature Construction for Linear Methods, Selecting Step-Size Parameters, Deep Q learning.

Module:7	**Policy Gradient Methods**	**7 hours**
Policy Approximation and its Advantages - REINFORCE algorithm - actor-critic methods - Policy Gradient for Continuing Problems - Policy Parameterization for Continuous Actions - Asynchronous Advantage Actor-Critic - Case studies: Samuel's checker player, TDgammon, Acrobot, AlphaGo.		
Module:8	**Contemporary Issues**	**2 hours**
	Total Lecture hours	**45 hours**

Text Book(s)	
1	R. S. Sutton and A. G. Barto., Reinforcement Learning - An Introduction, 2018, 2nd Edition, MIT Press.
2	Kevin P. Murphy, Probabilistic Machine Learning: Advanced Topics, 2023, MIT Press.
Reference Book(s)	
1	K. P. Murphy, Machine Learning: A Probabilistic Perspective, 2012, MIT Press.
2	E. Alpaydin, Introduction to Machine Learning, 2015, 3rd Edition, MIT Press.
3	Marc Peter Deisenroth, A. Aldo Faisal and Cheng Soon Ong., Mathematics for Machine Learning, 2020, Cambridge University Press.
4	Kevin P. Murphy. Probabilistic Machine Learning: An Introduction, 2022, MIT Press.

Mode of Evaluation: CAT, Assignment, Quiz and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course code	Course Title	L	T	P	C
PMDS602P	ADVANCED MACHINE LEARNING LAB	0	0	2	1
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. Understand the implementation procedures for the advanced machine learning algorithms.
2. Understand the modern notions in data analysis and applying the algorithms for analysing the real-world problems.

Course Outcomes

On the successful completion of the course, student will be able to:

1. Understand the fundamentals of multivariate methods and non-parametric methods
2. Understand and perform an evaluation of Bayesian Belief Network algorithms.
3. Analyse the Monte-Carlo methods for reinforcement data.
4. Analyse the Eligibility trace and Function Approximation for the real world data.
5. Analyse and implement various advanced machine learning algorithms in a range of real-world applications.

Indicative Experiments

1	Generate and estimate the multivariate normal distributed data
2	Outlier detection of multivariate data
3	Bayesian Neural Network
4	Multi-Armed Bandits – Application
5	Action selection using upper confidence bounds and Thompson sampling
6	Monte-Carlo methods: Monte-Carlo Prediction and Monte-Carlo Control
7	Temporal Difference Learning: One-Step TD Learning, n-Step TD Learning, TD Prediction
8	Deep Q-Learning, Deep Q-networks, Distributed Deep Q-Learning
9	Policy-Based Methods: The Vanilla policy gradient, Actor-Critic Methods
10	Application of Reinforcement Learning Algorithms in the Real-World data
	Total Laboratory Hours **30 hours**

Text Book(s)

1	Enes Bilgin., Mastering Reinforcement Learning with Python, 2020, Packt Publishing Ltd, Birmingham, UK.

Reference Book(s)

1	Taweh Beysolow II., Applied Reinforcement Learning with Python. 2019, Apress Berkeley, CA.

Mode of Evaluation: Assignment and FAT.

Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS603L	DEEP LEARNING	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To introduce the fundamentals of neural networks and types of neural networks.
2. To introduce Recurrent Neural Networks, Convolutional Neural Networks and its variants.
3. To develop and train deep neural networks.
4. To introduce complex learning models and deep learning models.
5. To introduce the internal structure of LSTM and GRU and the differences between them.

Course Outcomes

On completion of the course, students will be able to

1.Understand the fundamentals of deep learning and build deep learning models.
2. Apply the most appropriate deep learning method in any given situation.
3. Analyse neural network models in data-intensive real-time problems.
4. Create efficient generative models.
5. Learn and apply convolutional and recurrent neural network techniques.

Module:1	Neural Networks	7 hours

Introduction to neural networks, biological neuron, Idea of Computational Units, McCulloch-Pitts Unit and Thresholding Logic, Linear Perceptron, Perceptron Learning Algorithm, Convergence theorem for Perceptron Learning Algorithm, Linear separability, feed-forward network, input, hidden and output layers, organization and architecture of neural networks, linear and nonlinear networks.

Module:2	Training Algorithms for Feed Forward Networks	7 hours

Cost functions, Back-propagation algorithms, Learning the weights, gradient descent algorithm, Unit saturation, heuristics to avoid local optima, accelerated algorithms, Multilayer Perceptron, Empirical Risk Minimization, regularization methods.

Module:3	Convolutional Neural Networks	6 hours

Architectures, Properties of CNN representations: invertibility, stability, invariance, convolution, pooling of layers, CNN and Tensor Flow, Difficulty of training deep neural networks, Greedy layer-wise training, LeNet and AlexNet Architectures.

Module:4	Optimization Methods for Neural Networks	6 hours

Adagrad, Adadelta, RMS Prop, ADAM, NAG, Second order methods for training, Saddle point problem in Neural Networks, Dropout, Drop Connect, Batch Normalization.

Module:5	**Recurrent Neural Networks**	**6 hours**
RNN, LSTM, GRU, Encoder-Decoder Architectures, Auto encoders, Variational autoencoders, Bidirectional LSTMs, Bidirectional RNNs.		
Module:6	**Deep Generative Learning Models**	**8 hours**
Dynamic Memory Models, Reinforcement learning, Restrictive Boltzmann Machines (RBMs), Introduction to MCMC and Gibbs Sampling, gradient computation in RBMs, Deep Boltzmann Machine, Deep Belief Networks, Generative adversarial networks.		
Module:7	**Advanced Deep Neural Networks**	**3 hours**
Variational autoencoders, multitask deep learning, multi-view deep learning.		
Module:8	**Contemporary Issues**	**2 hours**
	Total Lecture hours	**45 hours**

Text Book(s)	
1	Bengio, Yoshua, Ian Goodfellow and Aaron Courville, Deep Learning, 2016, MIT Press.
2	Aston Zhang, Zachary C. Lipton, Mu Li and Alexander J. Smola, Dive into Deep Learning, December 2023, 1st Edition, Cambridge University Press.
Reference Book(s)	
1	Raúl Rojas, Neural Networks: A Systematic Introduction, 1996, 2nd Edition, Springer.
2	Nikhil Buduma and Nikhil Lacascio, Fundamentals of Deep Learning, 2017, O'Reilly Publishers.

Mode of Evaluation: CAT, Assignment, Quiz and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course code	Course Title	L	T	P	C
PMDS603P	DEEP LEARNING LAB	0	0	2	1
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To enable the students for having experimental knowledge of implementing neural network algorithms using python programming.
2. To make students capable to do classification of images using deep learning algorithms.
3. To enable the students to implement various deep learning networks like Adaline, Madalin.

Course Outcomes

At the end of the course the student should be able to:

1. Do Feature Extraction from Image and Video Data
2. Implement Image segmentation and Instance segmentation in Images.
3. Implement image recognition and image classification using a pretrained network.
4. Implement Analysis on Traffic Information Dataset
5. Do classification and feature extraction using autoencoders.

Indicative Experiments

1	Implementation of different activation functions to train Neural Network.
2	Implementation of different Learning Rules.
3	Implementation of Perceptron Algorithm.
4	Implementation of various neural networks.
5	Implementation of Optimization Methods for Neural Networks
6	Implementation of Image recognition and Image classification using a pretrained network.
7	Implementation of autoencoders.
	Total Laboratory Hours **30 hours**

Text Book (s)

1	Josh Patterson and Adam Gibson, Deep Learning: A Practitioner's Approach, 2017, O'Reilly Media.
2	Vinita Silaparasetty, Deep Learning Projects using Tensor Flow 2, 2018, Apress.

Reference Book (s)

1	Francois Chollet, Deep Learning with Python, 2017, Manning Press.
2	Jojo Mollayil, Learn Keras for Deep Neural Networks, 2018, Apress.

Mode of Evaluation: Assignment and FAT.

Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course code	Course Title	L	T	P	C
PMDS604L	Exploratory Data Analysis	2	0	0	2
Pre-requisite	Nil	Syllabus version			
		1.0			

Course Objectives

1. The course introduces the methods for data preparation and data understanding.
2. It covers essential exploratory techniques for understanding multivariate data summarizing it through statistical and graphical methods.
3. Supports to summarize use of predictive analytics, data science and data visualization.

Course Outcomes

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

1. Understand various data formats, sources and storage mechanisms.
2. Prepare the missing data and manage data wrangling and manipulation
3. Explain and present the findings in the Data Sets, after the Analysis is complete
4. Demonstrate the data visualization and make interpretations
5. Construct the data story using various software tools.

Module:1	Introduction to Exploratory Data Analysis	4 hours

Data Analysis - Exploratory Data Analysis and Data Science Process - Responsibilities of a Data Analyst - Data Analytics vs. Data Analysis - Types of Data - Understanding Different Types of File Formats - Languages for Data Professionals - Overview of Data Repositories - Data Marts, Data Lakes, ETL, and Data Pipelines - Foundations of Big Data - Identifying Data for Analysis

Module 2	Data Wrangling	4 hours

Data Sources – Data Loading, Storage and File Formats - Reading and Writing Data in Text Format, Web Scraping, Binary Data Formats, interacting with Web APIs, Interacting with Databases – Data Wrangling - Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting - Tools for Data Wrangling - Data Cleaning and Preparation - Handling Missing Data, Data Transformation, String Manipulation

Module:3	Data Analysis	4 hours

Statistical summary measures, data elaboration, 1-D Statistical data analysis, 2-D Statistical data Analysis, contingency tables, n-D Statistical data analysis.

Module:4	Outlier Analysis	4 hours

Outliers and Outlier Analysis - Outlier Detection Methods - Proximity-Based Approaches – distance metrics, Mahalanobis distance, Outlier Detection in High-Dimensional Data.

Module:5	Data Visualization	4 hours

Intro to data visualization - Visualization Tools - Getting started with Tableau Desktop – Connecting to the dataset - Creating charts – Creating common visualizations (bar charts, line charts etc.) - Filtering and sorting data - Adding Titles, Labels, and descriptions - Publish your work to Tableau Cloud - Interactivity with text and visual tooltips - Interactivity with actions (filter, highlight, URL) – Assembling dashboards from multiple charts

Module:6	**Exploratory Visualization Techniques**	**4 hours**
Introduction to data Visualization libraries – Customizing plots for effective communication-Interactive visualization tools - Geographic visualization - Text and sentiment Analysis.		
Module:7	**Insights of Data Visualization**	**4 hours**
Introduction to Power BI - Understanding Desktop - Understanding Power BI Report Designer - Report Canvas, Report Pages: Creation, Renames - Report Visuals, Fields and UI Options - Experimenting Visual Interactions, Advantages - Reports with Multiple Pages and Advantages - Pages with Multiple Visualizations - PUBLISH Options and Report Verification in Cloud - Adding Report Titles-Report Format Options.		
Module:8	**Contemporary Issues**	**2 hours**
	Total Lecture hours	**30 hours**

Text Book(s)	
1	McKinney, W., Python for Data Analysis: Data Wrangling with Pandas, NumPy and IPython, 2017, 2nd Edition, O"Reilly Media.
2	Suresh Kumar Mukhiya and Usman Ahmed, Hands-On Exploratory Data Analysis with Python, 2020, Packt Publishing.
Reference Book(s)	
1	O"Neil, C., and Schutt, R., Doing Data Science: Straight Talk from the Frontline by, 2013, O"Reilly Media.
2	Alberto Ferrari and Marco Russo, Introducing Microsoft Power BI, 2016, Microsoft Press, Washington.
3	Steve Wexler, Jeffrey Shaffer, Andy Cotgreave, The Big Book of Dashboards, 2017, John Wiley & Sons.
4	Ryan Sleeper, Practical Tableau, 2018, O" Reilly Media.
5	Roger F Silva, Business Intelligence Clinic: Create and Learn, 2018, Create and Learn

Mode of Evaluation: CAT, Assignment, Quiz and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course code	Course Title	L	T	P	C
PMDS604P	Exploratory Data Analysis Lab	0	0	2	1
Pre-requisite	Nil	Syllabus version			
		1.0			

Course Objectives

1. Emphasize the importance of programming in EDA.
2. Familiarize the student with programming for various tasks.
3. Explore data sets and file processing facilities using software tools
4. Explore diverse datasets to tackle real-world data challenges and uncover hidden insights.

Course Outcomes

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

1. Handle missing data and manage data wrangling and manipulation
2. Create data visualization and report making using various software tools
3. Demonstrate the visualizations and make interpretations
4. Create a data story using various software tools.

Indicative Experiments:

Note: - The following set of practical's could be implemented in Python/ R/ Power BI/ Tableau or any other suitable software.

1	Implement Data Loading, Storage and File Formats. Read data and store them in text format.
2	Implement the code to interact with Web APIs and to perform web scrapping.
3	Demonstrate Data Cleaning and Preparation.
4	Implement Data wrangling on a data set.
5	Demonstrate the handling of missing data and string manipulation.
6	Create common charts with titles, labels, and descriptions using Tableau.
7	Perform sorting and filtering using Tableau, create visualizations, and publish it on Tableau Cloud.
8	Perform data visualization using Power BI.
9	Create reports using Power BI.
10	Create a data story in Tableau or Power BI.

Total Laboratory hours	**30 hours**

Text Book(s)

1	Suresh Kumar Mukhiya, Usman Ahmed, Hands-On Exploratory Data Analysis with Python, 2020, Packt Publishing.

Reference Book(s)

1	Alberto Ferrari and Marco Russo, Introducing Microsoft Power BI, 2016, Microsoft Press, Washington.
2	Steve Wexler, Jeffrey Shaffer, Andy Cotgreave, The Big Book of Dashboards, 2017, John Wiley & Sons.
3	Ryan Sleeper, Practical Tableau, 2018, O" Reilly Media.

Mode of assessment: Assignment and FAT.			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS605L	Data Structure and Algorithms	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To provide knowledge on various data structures and their real time applications
2. To familiar in the design and performance evaluation of data structure and algorithms
3. To familiar in advanced techniques with industrial development

Course Outcomes

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

1. Understand the foundation of data structure, compute the complexity and notations, design and implement Array ADT.
2. Identify suitable algorithm for the abstract data structure Stack, Queue and List.
3. Classify various Tree data structures and its applications.
4. Select the suitable algorithm for Sorting and Searching.
5. Develop suitable data structure for Graph and its Applications.

Module:1	Foundation of Data Structure	7 hours

Importance of Data Structure - Asymptotic Notations (Big O, Small O, Theta, Omega) Performance of Algorithm and Analysis - Time and Space Complexity - ADT- Arrays: One dimension and Two dimension, Structure and Union - Pointers, Storage Allocation: Static and Dynamic Allocation

Module:2	Stacks and Queue	6 hours

Stack: Definition, Operations, Implementations, Applications: Recursion, Infix to Postfix and Evaluation of Postfix, Queue: Definition, Operations, Implementations, Applications: Circular Queue - Multiple Stack and Queues.

Module:3	Lists	6 hours

Linked List: Definition, Operations (INSERT, DELETE, TRAVERSE- DISPLAY) – Implementation: Singly Linked Lists, Doubly Linked Lists, Circular Linked Lists- Application : Polynomial Addition using Linked List

Module:4	Trees	6 hours

Definition, Terminology, Binary Tree: Binary Tree Representation, Binary Search Tree, Binary Tree Traversal: In order, Pre order, Post Order and Level order traversal. Heap Data Structure- Min Heap and Max Heap Tree construction.

Module:5	Advanced Trees	6 hours

Balanced Trees - AVL trees: Terminology, basic operations (rotation, insertion and deletion), 2-3 Trees, 2-3-4 Trees, B Trees, B+ Trees

Module:6	Graphs	6 hours

Graph ADT, Elementary Graph Operation, Minimum cost spanning tree Algorithms, Shortest Path -Single Source and All Pair Algorithms.

Module:7	Search and Sort	6 hours

Search - Linear and Binary Search - Applications; Sorting: Bubble Sort, Insertion Sort, Selection Sort Quick, Merge Sort and Heap Sort.

Module:8	Contemporary Issues	2 hours
	Total Lecture hours	45 hours

Text Book(s)	
1	Thomas H. Cormen, Charles E. Leiserson, Ronald L.Rivest and Clifford Stein, Introduction to Algorithms, 2022, 4th Edition, McGraw Hill/ MIT Press.
2	Langsam, Augenstein and Tanenbaum, Data Structures Using C and C++, 2015, 2nd Edition, Pearson.
Reference Book(s)	
1	Ellis Horowitz, Sartaj Sahni and Susan Anderson-Freed, Fundamentals of Data Structures in C, 2008, 2nd Edition, University Press.
2	R.C.T Lee, S.S Tseng, R.C Chang and Y.T Tsai, Introduction to the Design and Analysis of Algorithms, 2012, Tata McGraw-Hill.
3	Ellis Horowitz and Sartaj Sahni, Fundamental of Computer Algorithms, 1985, Galgotia.
4	Thomas H Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein, Introduction to Algorithms, 2010, 3rd Edition, Prentice Hall.

Mode of Evaluation: CAT, Assignment, Quiz and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS605P	Data Structure and Algorithms LAB	0	0	2	1
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To provide knowledge on various data structures and its real time applications
2. To familiar in design and performance evaluation of data structure and algorithms
3. To familiar in advanced techniques with industrial development

Course Outcomes

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

1. Understand the foundation of data structure, compute the complexity and notations, design and implement Array ADT.
2. Identify suitable algorithm for the abstract data structure Stack, Queue and List
3. Classify various Tree data structures and its applications.
4. Recognize suitable algorithm for Sorting and Searching.
5. Develop suitable data structure for Graph and its applications.

Indicative Experiments

1	Evaluation of time and space complexity for the recursive and iterative Algorithms.
2	Implementation of Array Data Structures.
3	Implementation of Stack and Queue and its Applications.
4	Implementation of Linked List and its Applications.
5	Binary Tree and Binary search tree Implementation by performing Insert and Delete Operations.
6	Implement Min Heap and Max Heap data structures and AVL Trees.
7	Implementation of AVL Tree.
8	Implementation of Graph algorithm Minimal Spanning Trees and Single Source Shortest path algorithm.
9	Implementation of Bubble Sort, Insertion Sort and Selection Sort algorithm.
10	Implementation of Quick Sort and Merge Sort.
	Total Lecture hours **30 hours**

Text Book(s)

1	Thomas H. Cormen, Charles E. Leiserson, Ronald L.Rivest and Clifford Stein, Introduction to Algorithms, 2022, 4th Edition, Mcgraw Hill/MIT Press.
2	Langsam, Augenstein and Tanenbaum, Data Structures Using C and C++, 2015, 2nd Edition, Pearson.

Reference Book(s)

1	Ellis Horowitz, Sartaj Sahni and Susan Anderson-Freed, Fundamentals of Data Structures in C, 2008, 2nd Edition, University Press.
2	R.C.T Lee, S.S Tseng, R.C Chang and Y.T Tsai, Introduction to the Design and Analysis of Algorithms, 2012, Tata McGraw-Hill Edition.

Mode of assessment: Assignment and FAT.

Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS606L	Natural Language Processing	3	0	0	3
Pre-requisite	Nil	Syllabus version			
		1.0			

Course Objectives

1. To introduce the fundamental concepts and techniques of Natural language Processing for analyzing words based on Morphology and CORPUS.
2. To examine the NLP models and interpret algorithms for classification of NLP sentences by using both the traditional, symbolic and the more recent statistical approach.
3. To get acquainted with the algorithmic description of the main language levels that includes morphology, syntax, semantics, and pragmatics for information retrieval and machine translation applications.

Course Outcomes

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

1. Understand the fundamental concepts of natural language processing.
2. Understand the text pre-processing and corpora.
3. Analyze the words and perform POS tagging.
4. Distinguish between the syntactic and semantic correctness of the natural language.
5. Develop simple language models using NLTK.

Module:1	Introduction to NLP	5 hours

Introduction to various levels (stages) of natural language processing, Ambiguities, varieties and computational challenges in processing natural languages. Introduction to Real life applications of NLP such as spell and grammar checkers, information extraction, information retrieval, question answering, and machine translation.

Module:2	Text processing	6 hours

Text pre-processing, challenges, tokenization, sentence segmentation, regular expressions, words, text normalization, minimum edit distance, introduction to corpora, corpora analysis.

Module:3	Language modelling	6 hours

The role of language models. N-gram models. Estimating parameters and smoothing. Evaluating language models.

Module:4	Morphological analysis and POS tagging	7 hours

Parts of speech and morphology, Inflectional and Derivation Morphology, Morphological Analysis, FSA and Generation using finite state transducers. Introduction to POS tagging, HMM, Viterbi decoding for HMM.

Module:5	Syntactic analysis	6 hours

Ambiguities in syntactic parsing, context free grammar, CYK parsing, shallow parsing and chunking, dependency parsing, statistical parsing and PCFG

Module:6	Semantic analysis	7 hours

Semantics, Lexical Semantics, Word senses, Relations between senses, Word

Sense Disambiguation, Word similarity, WordNet, Thesaurus based word similarity, Thematic Roles, Semantic Role Labelling with CRFs.		
Module:7	**NLTK with Python**	**6 hours**
Tokenizing Text and WordNet Basics- Replacing and Correcting Words- Part-of Speech Tagging- Extracting Chunks- Text Classification – Named Entity Recognition.		
Module:8	**Contemporary Issues**	**2 hours**
	Total Lecture hours	**45 hours**

Text Book(s)	
1	Daniel Jurafsky and James H. Martin, Speech and Language Processing, 2017, 3rd edition, Prentice Hall.
2	Chris Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, 2016, MIT Press.
Reference Book(s)	
1	James Allen “Natural Language Understanding, 2012, 8th Edition, Pearson Publication.
2	Vajjala, Sowmya, Bodhisattwa Majumder, Anuj Gupta and Harshit Surana. Practical natural language processing: A comprehensive guide to building real-world NLP systems, 2020, O'Reilly Media.

Mode of Evaluation: CAT, Assignment, Quiz and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS607L	OPTIMIZATION TECHNIQUES	3	0	0	3
Pre-Requisite	NIL	Syllabus Version			
		1.0			

Course Objectives

1. To familiarize the students with some basic concepts of optimization techniques and approaches.
2. To formulate a real-world problem as a mathematical programming model.
3. To develop the model formulation and applications are used in solving decision problems.
4. To solve specialized linear programming problems like the transportation and assignment Problems.

Course Outcomes

At the end of the course, the student will be able to :

1. Understand the operations research techniques like linear programming problem.
2. Apply the linear programming problem in industrial optimization problems.
3. Solve allocation problems using various operations research methods.
4. Understand the characteristics of different types of decision-making environment and the appropriate decision-making approaches and tools to be used in each type.
5. Recognize competitive forces in the marketplace and develop appropriate reactions based on existing constraints and resources.

Module: 1	Introduction to Operations Research and Linear Programming Problems	6 hours

Introduction to Operation Research, Mathematical Models, Scope and applications of Operation Research - Mathematical formulation of Linear Programming - Limitations or constraints, Problem (LPP): Methods for solution, Graphical analysis, LPP, Graphical LPP Maximization Problems, and Minimization problems.

Module: 2	Basic LPP Problem	6 hours

Linear programming to standard form, Simplex Method: Basics of Simplex Method, Simplex Method with two variables, Simplex Method with more than two variables and Big M Method.

Module: 3	Dual Linear Programming Problem	7 hours

Introduction to Primal and Dual problems - Duality theorem - Dual problem properties - Solution techniques of Dual Problem - Dual Simplex Method - Relationship between direct and dual problems - Economic interpretation of Duality.

Module: 4	Transportation and Assignment Problem	6 hours

Introduction: Transportation Problem-Balanced-Unbalanced-Methods of basic feasible solution - Optimal solution-MODI method. Assignment problem-Hungarian Method.		
Module: 5	**Non-Linear programming Problems**	**6 hours**
Method of Lagrange multipliers, Kuhn-Tucker theory, convex optimization, Quadratic optimization		
Module: 6	**Network Analysis**	**6 hours**
Basic concepts-Construction of Network-Rules and precautions- Critical Path Method (CPM) and Program Evaluation and Review Technique (PERT) Networks-Crashing a Network as LPP - Probability and cost consideration.		
Module: 7	**Game Theory**	**6 hours**
Introduction to Two Person Zero-Sum Game-Solution of games with saddle points and without saddle points - 2×2 games-dominance principal m×2 and 2×n games - Graphical method.		
Module: 8	**Contemporary Issues**	**2 Hours**
	Total Lecture Hours	**45 Hours**

Text Book(s)	
1	Hamdy Taha, Operations Research, 10th edition, Prentice Hall India, 2019
2	P. K. Gupta and D. S. Hira, Operations Research, 2018, S. Chand & co.
Reference Book(s)	
1	S.D. Sharma , Operations Research, 2000, Nath & Co., Meerut.
2	Maurice Solient, Arthur Yaspen, Lawrence Fridman, OR methods and Problems, 2003, New Age International Edition.
3	J K Sharma, Operations Research Theory & Applications, 2007, 3rd Edition, Macmillan India Ltd.

Mode of Evaluation**:** CAT, Assignment, Quiz and FAT.			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS608L	HEALTHCARE ANALYTICS	2	0	0	2
Pre-Requisite	NIL	Syllabus Version			
		1.0			

Course Objectives

1. To acquire the knowledge about the healthcare data sources and basic analytics.
2. To provide a keen information on biomedical signal and image analysis.
3. To use the latest computer technologies in healthcare sectors effectively.
4. To explore the clinical trials, epidemiological measures, clinical prediction models and survival analysis in the medical science, public health and others.
5. To analyze the practical systems and applications of healthcare data analytics to the society.

Course Outcomes

At the end of the course, the student will be able to:

1. Remember the concepts of healthcare data sources and basic analytics.
2. Understand the biomedical signal and image analysis for medical problems.
3. Apply natural language processing, data mining, social media analytics and other modern technologies for healthcare data.
4. Utilize the clinical trials, epidemiological measures, clinical prediction models, survival models, and ROC curve analysis for public health studies.
5. Analyze the practical systems and applications of healthcare data analytics to the society.

Module: 1	Healthcare Data Sources and Basic Analytics	4 Hours

Electronic Health Records (HER) – Components of HER – Coding Systems – Benefits of HER – Barriers to Adopting HER – Challenges of Using EHR Data – Phenotyping Algorithms.

Module: 2	Biomedical Signal Analysis	4 Hours

Types of Biomedical Signals – ECG, EEG & EMG Signals – Denoising of Signals – Multivariate Biomedical Signal Analysis – Cross-Correlation Analysis – Genomic Data Analysis – Genomic Data Generation.

Module: 3	Biomedical Image Analysis	4 Hours

Biomedical Imaging Modalities – X-Ray, CT, PET, MRI, Ultrasound, Microscopy Images – Object Detection – Image Segmentation – Image Registration – Feature Extraction – Sensor Data Analysis – Mining of Sensor Data in Medical Informatics.

Module: 4	NLP, Data Mining and Social Media Analytics	4 Hours

Natural Language Processing (NLP) – Report Analyzer – Text Analyzer – Core NLP Components – Clinical Data Mining – Information Extraction – Current Methodologies – Clinical Text Corpora and Evaluation Metrics – Social Media Analysis – Detection & Tracking of Infectious Disease and Public Health Research.

Module: 5	**Clinical Trials and Epidemiological Measures**	**4 Hours**
Clinical Trial and its Phases – Blinding in Clinical Trials – Bioassays and its Types – Measures of Disease Frequency – Incidence – Prevalence – Relative Risk – Epidemiological Study Designs – Concept of Bias – Sensitivity – Specificity – ROC Curve – Properties of ROC Curve.		
Module: 6	**Clinical Prediction and Survival Analysis**	**4 Hours**
Clinical Prediction Models – Survival Models – Survival Data and Censoring – Survival and Hazard Functions – Kaplan–Meier Curve – Clinical Life Table – Mantel–Haenszel Test – Cox Proportional Hazards Model – Survival Trees – Temporal Data Pattern Mining.		
Module: 7	**Practical Systems for Healthcare**	**4 Hours**
Medical Data Visualization – Visual Analytics for Clinical Workflow, Clinicians & Patients – Fraud Detection in Healthcare – Identifying Healthcare Fraud from Data – Clinical Decision Support Systems (CDSS) – Types of CDSS – Knowledge-Based CDSS – Nonknowledge-Based CDSS – Diagnostic Decision Support.		
Module: 8	**Contemporary Issues**	**2 Hours**
	Total Lecture Hours	**30 Hours**

Text Book(s)	
1	C.K. Reddy and C.C. Aggarwal, Healthcare Data Analytics, 2020, CRC Press, New York.
Reference Book(s)	
1	N.P. Jewell, Statistics for Epidemiology, 2004, CRC Press, New York.
2	W.J. Krzanowski and D.J. Hand, ROC Curves for Continuous Data, 2009, CRC Press, New York.
3	H. Yang and E.K. Lee, Healthcare Analytics: From Data to Knowledge to Healthcare Improvement, 2016, John Wiley & Sons, New Jersey.

Mode of Evaluation**:** CAT, Assignment, Quiz and FAT.			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS608P	HEALTHCARE ANALYTICS LAB	0	0	2	1
Pre-Requisite	NIL	Syllabus Version			
		1.0			

Course Objectives

1. To use the latest computer technologies in healthcare sectors effectively.
2. To utilize the data analysis, biomedical signal & image processing and data mining techniques for the healthcare data.
3. To explore the epidemiological measures, ROC curve analysis, survival models, visual analytics and fraud detection methods for the medical data.

Course Outcomes

At the end of the course, the student should be able to:

1. Understand the data analysis, biomedical signal & image processing and data mining techniques for the healthcare data.
2. Apply the epidemiological measures, ROC curve analysis, survival models, visual analytics and fraud detection methods for the clinical data.

Indicative Experiments

1	Data Analytics for Electronic Health Records (EHR)
2	Denoising and Cross-Correlation Analysis of Biomedical Signals
3	Object Detection and Segmentation for Biomedical Images
4	Clinical Data Mining Algorithms
5	Computation of Incidence, Prevalence and Relative Risk Measures
6	ROC Curve Analysis for Medical Data
7	Clinical Life Table Construction
8	Kaplan-Meier and Cox Regression Analysis
9	Visual Analytics for Clinical Data
10	Fraud Detection Methods for Healthcare Data
	Total Lecture Hours	**30 Hours**

Text Book(s)

1	C.K. Reddy and C.C. Aggarwal, Healthcare Data Analytics, 2020, CRC Press, New York.

Reference Book(s)

1	E. Harrison and R. Pius, R for Health Data Science, 2020, CRC Press, New York.
2	H. Yang and E.K. Lee, Healthcare Analytics: From Data to Knowledge to Healthcare Improvement, 2016, John Wiley & Sons, New Jersey.

Mode of Evaluation: Assignment and FAT.

Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS609L	BLOCKCHAIN TECHNOLOGY	2	0	0	2
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. Understanding the concepts of decentralization and blockchain.
2. Explore the concepts related to security and privacy.
3. Apply cryptography concepts of blockchain for real world problems.
4. Examine the working of cryptocurrency and blockchain platform.

Course Outcome

At the end of the course, the student should be able:

1. Gain knowledge on the fundamentals of decentralization and blockchain.
2. Understand the concepts of cryptography.
3. Apply cryptography concepts for real world problems.
4. Familiarise with cryptocurrency.
5. Recognize the use cases of blockchain technology.

Module:1	**Introduction to Blockchain**	**4 Hours**
The evolution of blockchain, Generic elements of a blockchain, Features of a blockchain, Types of blockchain, Consensus mechanism.		
Module:2	**Decentralization**	**4 Hours**
Decentralization using blockchain, Methods of Decentralization, Blockchain and Full Ecosystem Decentralization, Decentralized Organizations, Platforms for Decentralization.		
Module:3	**Symmetric Cryptography**	**4 Hours**
Cryptography, Principles of cryptography, Cryptographic primitives: Stream ciphers, Block ciphers, Data Encryption Standard, Advanced Encryption Standard.		
Module:4	**Asymmetric Cryptography**	**4 Hours**
Public and Private keys, Encryption and Decryption, Discrete logarithm problem, Hashing Function, Elliptic Curve Digital Signature Algorithm.		
Module:5	**Cryptocurrency: Bitcoin**	**4 hours**
Bitcoin, Digital keys and addresses, Transactions, Structure of a Block, Mining, Bitcoin Network and Payments, Bitcoin clients.		
Module:6	**Blockchain Platform: Ethereum**	**4 Hours**
Components of Ethereum, Transactions, Transaction Substate, State Storage, Transaction receipts, Gas and Fee schedule, Nodes and Miners, Tokens, Smart Contracts.		
Module:7	**Use cases of Blockchain Technology**	**4 Hours**
Financial and Non-Financial Use cases: Global Payments, Financial Markets & Trading, Croud funding, Voting system, Event Registration, Document Verification, Electronic Health Record System.		
Module:8	**Contemporary Issues**	**2 hours**
Alternate cryptocurrency and Blockchain Platforms, Development tools and Frameworks.		
	Total Lecture hours	**30 hours**

Text Book(s)	
1	Imran Bashir, Mastering Blockchain, 2023, 4th Edition, United Kingdom, Packt Publisher.
Reference Book(s)	
1	Julien Riposo, Some Fundamentals of Mathematics of Blockchain, 2023, 1st Edition, Springer Nature, Switzerland.
2	Alexander Lipton & Adrien Treccani, Blockchain and Distributed Ledgers: Mathematics, Technology, and Economics , 2021, 1st Edition, World Scientific.

Mode of Evaluation: CAT, Assignment, Quiz and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course code	Course Title	L	T	P	C
PMDS610L	FINANCIAL ANALYTICS	2	0	0	2
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To learn to model financial time series using linear ARMA type time series.
2. To study and analyze to test and model heteroscedastic effects using ARCH /GARCH type time series.
3. To learn how to test for unit root and construct ARMA models.

Course Outcomes

At the end of the course, the student will be able to:

1. Understand the deep knowledge of financial data and properties
2. Understand and apply the financial time series analysis
3. Apply and Analyse the Volatility models to financial data.
4. Perform cross-validation of various financial models developed.
5. Forecast future observations on financial data.

Module:1	Financial data and their properties	5 hours

Asset Returns – Bond Yields and Prices – Implied Volatility – Examples and Visualization offinancial data – Multivariate returns.

Module:2	Linear models for financial time series	5 hours

Simple autoregressive models – Simple moving average models – Simple ARMA models – Unit Root non-stationarity – Exponential smoothing.

Module:3	Seasonal and Long memory models	4 hours

Seasonal models – Regression models with time series errors – Long memory models.

Module:4	Asset Volatility and Volatility models	4 hours

Characteristics of Volatility – Structure of a model – Testing for ARCH Effect – ARCH Model– GARCH Model – GARCH-M Model – Exponential GARCH Model – Threshold GARCHmodel – Stochastic volatility model – alternative approaches.

Module:5	Applications of Volatility Models	4 hours

GARCH Volatility Term structure – Option pricing and hedging – Time-Varying Correlationsand Betas – Minimum Variance Portfolios – Prediction.

Module:6	High-Frequency Financial Data	3 hours

Nonsynchronous trading – Bid-ask spread of trading prices – Empirical characteristics oftrading data – Models for price changes.

Module:7	Value at Risk	3 hours

Risk measure and Coherence – Risk metrics –Extreme value approach to Value at Risk –Peak over thresholds.

Module:8	Contemporary Issues	2 hours
	Total Lecture hours	30 hours

Text Book(s)			
1	Sinem Derindere Köseoğlu, Financial Data Analytics Theory and Application, 2022, Springer.		
Reference Book(s)			
1	Statistical Analysis of Financial Data in S Plus, by R Carmona, April 2004, Springer.		
2	Ruey S. Tsay, An Introduction to Analysis of Financial Data with R, 2013, Wiley.		
3	Ruey S. Tsay, Analysis of Financial Time Series, 2010, 3rd edition, Wiley.		
Mode of Evaluation: CAT, Assignment, Quiz and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course code	Course Title	L	T	P	C
PMDS610P	Financial Analytics Lab	0	0	2	1
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. To enable the students for having experimental knowledge of basic concepts of financial analytics.
2. To make students capable of using important models for analyzing financial data.
3. To enable students learn about different time series model.

Course Outcomes

At the end of the course the student should be able to:

1. Utilize important models to analyse market data.
2. Utilize time series models for forecasting.
3. Utilize models to predict volatility.
4. Measure risk assessment.

Indicative Experiments

1	Calculation of Bond values, Bond yields
2	Visualization of financial data- charts, graphs, maps, info-graphics, diagrams and virtual dashboards
3	Executing Moving Average Models (MA models): 1st and 2nd order
4	Executing ARMA model for weakly stationary stochastic time series
5	Single Exponential Smoothing, Double Exponential Smoothing
6	Regression Models with Time Series Errors- regression model with ARIMA errors
7	Engle's ARCH Test- to assess the significance of ARCH effects
8	GARCH model, GARCH M model- to predict the volatility of returns on financial assets
9	Minimum Variance Portfolio
10	Coherent risk measure- Back-testing, POT approach
	Total Lecture Hours **30 Hours**

Text Book (s)

1	Mark J. Bennett & Dirk L. Hugen, Financial Analytics with R, 2016, Cambridge University Press.
2	Yves Hilpisch, Python for Finance- Mastering Data-Driven Finance, 2019, O'Reilly Media.

Reference Book(s)

1	R Carmona, Statistical Analysis of Financial Data in S Plus, April 2004, Springer.
2	Ruey S. Tsay An Introduction to Analysis of Financial Data with R, 2013, Wiley.

Mode of assessment: Assignment and FAT

Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS611L	MULTIVARIATE DATA ANALYSIS	3	0	0	3
Pre-requisite	NIL	Syllabus version			
		1.0			

Course Objectives

1. Inculcate deep knowledge on various multivariate distribution and multivariate techniques
2. Develop clear idea on when and where to use dependence and interdependence multivariate methods.
3. Bridge the relation between multivariate analysis and machine learning and strengthen the applications in diversified spectrum of fields.

Course Outcomes

At the end of the course, the student will be able to:

1. Understand the multivariate normal distribution and properties
2. Analyse the multivariate data using multivariate statistical linear model.
3. Compare the strengths and weaknesses of many popular dimensionality reduction approaches.
4. Apply the multivariate techniques for dimensionality reduction of multivariate data.
5. Design and implement various multivariate statistical models in a range of real-world applications.

Module:1	Introduction to Multivariate Data Analysis	4 hours

Multivariate data and their diagrammatic representation. Exploratory multivariate data analysis, sample mean vector, sample dispersion matrix, sample correlation matrix, graphical representation, means, variances, co-variances, correlations of linear transforms, six step approach to multivariate model building. Introduction to multivariate linear regression, logistic regression, principal component analysis, factor analysis, cluster analysis, canonical analysis and canonical variables.

Module:2	Multivariate Normal Distribution	7 hours

Introduction to multivariate normal distribution, probability density function and moment generating function of multivariate normal distribution, singular and non-singular normal distributions, distribution of linear and quadratic form of normal variables, marginal and conditional distributions. Wishart matrix-its distribution and properties.

Module:3	Multivariate Linear Models	7 hours

Maximum likelihood estimation of parameters, tests of linear hypothesis, distribution of partial and multiple correlation coefficients and regression coefficients. Multivariate linear regression, multivariate analysis of variance of one way and two way classification data (only LR test). Multivariate analysis of covariance. Hoteling T^2 and Mahalanobis D^2 applications in testing and confidence set construction.

Module:4	Discriminant Analysis	7 hours

Discriminant model and analysis: a two group discriminant analysis, a three group discriminant analysis - Logistic Regression model and analysis: regression with a

binary dependent variable, estimating the logistic regression model, assessing the goodness of fit of the estimation model, testing for significance of the coefficients, interpreting the coefficients.		
Module:5	**Principal Component and Factor Analysis**	**7 hours**
Population and sample principal components, their uses and applications, large sample inferences, graphical representation of principal components, Biplots, the orthogonal factor model, dimension reduction, estimation of factor loading and factor scores, interpretation of factor analysis.		
Module:6	**Cluster Analysis**	**6 hours**
Concepts of cluster analysis and multidimensional scaling, similarity measures, hierarchical clustering methods, Ward's hierarchical clustering method, non-hierarchical clustering methods, K-means methods. Clustering based on statistical models, multidimensional scaling and correspondence analysis, perceptual mapping.		
Module:7	**Canonical Correlation**	**5 hours**
Concept of canonical correlation analysis, conjoint analysis, multidimensional scaling.		
Module:8	**Contemporary Issues**	**2 hours**
	Total Lecture hours	**45 hours**

Text Book(s)	
1	B.K. Tripathy, Anveshrithaa Sundareswaran, Shrusti Ghela. Unsupervised Learning Approaches for Dimensionality Reduction and Data Visualization, 2021, CRC Press.
2	Frederic Ros; Rabia Riad. Feature and Dimensionality Reduction for Clustering with Deep Learning. 2023, Springer.
Reference Book(s)	
1	Rao, C. R. and Rao, M. M., Multivariate Statistics and Probability, 2014, Elsevier & Academic Press.
2	Joseph F. Hair, Jr., William C. Black, Barry J. Babin, Rolph E. Anderson, and Ronald L. Tatham, Multivariate Data Analysis, 2014, 7th Edition, Pearson Education, India.
3	Hardly W.K. and Simor L., Applied Multivariate Statistical Analysis, 2015, 4th Edition, Springer- Verlag.
4	Richard A. Johnson and Dean W. Wichern, Applied Multivariate Statistical Analysis, 2017, 7th Edition, Prentice Hall India.

Mode of Evaluation: CAT, Assignment, Quiz and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS611P	MULTIVARIATE DATA ANALYSIS LAB	0	0	2	1
Pre-Requisite	NIL	Syllabus Version			
		1.0			

Course Objectives

1. Understand the implementation procedures for the multivariate data analysis using R/Python.
2. Analyse the multivariate data using the multivariate statistical techniques.

Course Outcome

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

1. Analyse and perform an evaluation of algorithms and model selection.
2. Perform the dimensionality reduction techniques.
3. Appreciate the underlying mathematical relationships within and across multivariate Algorithms.
4. Design and implement various multivariate algorithms in a range of real-world applications.

Indicative Experiments

1	MLE of mean vector and variance-covariance matrix from the normal population. Generating random numbers from a multivariate normal distribution.	2 hours
2	Hoteling T^2 and Mahalanobis D^2	
3	Computation of principal components and conducting factor analysis	
4	Fitting a multivariate linear regression model and its interpretation.	
5	Error analysis, outliers' detection and related tests	
6	Estimation, fitting and validating a logistic regression model.	
7	Classification between two normal populations using discriminant analysis.	
8	Cluster analysis	
9	Computation of canonical variables and correlation	
10	Structural Equation Model and related computations	
	Total Lecture Hours	**30 Hours**

Text Book(s)

1	B.K. Tripathy, Anveshrithaa Sundareswaran, Shrusti Ghela. Unsupervised Learning Approaches for Dimensionality Reduction and Data Visualization, 2021, CRC Press.

Reference Book(s)

2	Frederic Ros and Rabia Riad, Feature and Dimensionality Reduction for Clustering with Deep Learning, 2023, Springer.

Mode of Evaluation: Assignment and FAT.

Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024

Course Code	Course Title	L	T	P	C
PMDS612L	STATISTICS FOR MANAGERS	3	0	0	3
Pre-requisite	NIL	Syllabus Version			
		1.0			

Course Objectives

1. To equip the students the ability to solve and deal with the business analytics technique and quantitative decision making skills required to make smart managerial decision.
2. To amalgamate the intellectual facts and figures related to managerial statistics especially suited to business and econometric analysis.
3. To search for the real time feasible solution of modern managerial challenges and policy making.

Course Outcomes

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

1. Understand the fundamentals of managerial computing through their comprehension of data and data distribution using exploratory, descriptive, inferential statistics and industrial statistics.
2. Learn to apply and conduct the sample survey for data acquisition, processing, investigation and real time problem identification.
3. Learn to demonstrate the problem solving skill using statistical tools and methods for best decision making through minimal estimated risk.
4. Able to present, examine, analyse, evaluate, validate and drawing inference from data at hand and computed results.
5. Learn to build a model and implement it for future forecast and use it as a tool to solve a real time problem objectively to arrive at correct decision.

Module:1	Exploratory and Descriptive Statistics	6 hours

Data, data scale, data sources, tabular and graphical presentation, summarising qualitative and quantitative data, stem and leaf diagram, cross tabulation, scatter plot and trend line. Absolute and relative measures. Measures of location, measures of variability and association, measures of skewness and kurtosis, box plot, five summary statistics, z-score and detecting outliers.

Module:2	Random variable and Probability Distributions	6 hours

Classical, axiomatic and limiting definition of probability. Concept and definition of random variable, joint, conditional and marginal probability, Bayes theorem. Concept and computations of expectation. Moment generating function. Basic statistical constants Standard discrete distributions (Binomial, Poisson, Negative Binomial, Uniform), standard continuous distributions (normal, exponential, t-distribution, F-distribution, chi-square distribution, gamma and beta distribution).

Module: 3	Statistical Inference-Estimation and Testing	6 hours

Concept, definitions and computations of point and interval estimation, confidence intervals, properties of a good estimator, determining the sample size in estimation, concepts and procedure of hypothesis testing, measuring the power of a hypothesis test, testing of hypothesis for one sample as well as two sample tests, testing the difference between the means and difference between the

proportions. Non parametric tests (Chi square test, the sign test for paired data, rank sum test, The Mann-Whitney U test and the Krukskal- Wallis test and the Kolmogorov-Smirnov test).		
Module: 4	**Sample surveys and Design**	**6 hours**
Introduction to survey and sampling methods, basic terminology, types of survey and sampling methods, survey errors-sampling and non-sampling errors, random sampling, non-random sampling, simple random sampling, Stratified simple random sampling, cluster sampling and systematic sampling. Fundamentals of design of experiment.		
Module:5	**Correlation and Regression analysis**	**6 hours**
Correlation analysis, Pearson correlation coefficient, Spearman rank correlation, correlation ratio, concept of partial correlation, simple and multiple linear regression, making inference about population parameters, modelling techniques, checking the violations of assumptions of regression model and conduct of validation of regression model.		
Module:6	**Time Series and Forecasting**	**5 hours**
Introduction and basic concepts of time series data, decomposition of time series components, additive and multiplicative model, solving a problem involving all four components of time series, de- seasonalization using additive and multiplicative model. Single exponential smoothing method. Concept and definition of stationary and non-stationary time series.		
Module:7	**Decision Theory and Industrial Statistics**	**8 hours**
The decision environment, expected profit under uncertainty, marginal analysis using continuous distributions, expected value of perfect information, expected value of sample information, decision tree analysis. Index number, unweighted aggregate index, weighted aggregate index, average of relative methods, quantity and value indices, issues in constructing and using index numbers. Statistical process control, control charts, Interpretation of Control chart. Introduction to reliability and its measurements.		
Module:8	**Contemporary Issues**	**2 hours**
	Total Lecture hours	**45 hours**

Text Book(s)	
1	Ken Black, Business Statistics for Contemporary Decision Making, 2019, 10th Edition, Wiley.
Reference Book(s)	
1	Richard I, David M. Levin, David S. Rubin, Sanjay Rastogi, and Masood Husain Siddiqui, Statistics for Management, 2017, 7th Edition, Pearson.
2	David R. Anderson, Dennis J. Sweeney, Thomson A. Williams, Jim Freeman, and Eddie Shoesmith. Statistics for Business and Economics, 2014, 3rd Edition, Cengage Learning.

Mode of Evaluation: CAT, Digital Assignment, Quiz and FAT			
Recommended by Board of Studies	15.02.2024		
Approved by Academic Council	No. 73	Date	14.03.2024