KPR Institute of Engineering and Technology

Learn Beyond (Autonomous, NAAC "A")

Avinashi Road, Arasur, Coimbatore.

Great Place To Work. Certified MAR 2023-MAR 2024 INDIA

M.TECH. - DATA SCIENCES

Curriculum and Syllabi

(Regulations – 2023)

I. Vision and Mission of the Institute

Vision

To become a premier institute of academic excellence by imparting technical, intellectual and professional skills to students for meeting the diverse needs of the industry, society, the nation and the world at large.

Mission

- Commitment to offer value–based education and enhancement of practical skills
- Continuous assessment of teaching and learning processes through scholarly activities
- Enriching research and innovation activities in collaboration with industry and institutes of repute
- Ensuring the academic processes to uphold culture, ethics and social responsibilities

II. Vision and Mission of the Department

Vision

To foster the needs of students by providing learner centric teaching environment, continuous learning, research and development to become thriving professionals and entrepreneurs to excel in the field of computer science and contribute to the society.

Mission

- Providing value-based education and contented learning experience to the students.
- Educating the students with the state of art technologies and cultivating their proficiency in analytical and designing skills.
- Enabling the students to achieve a successful career in Data Sciences and Computer Science and Engineering or related fields to meet the changing needs of various stakeholders.
- Guiding the students in research by nurturing their interest in continuous learning towards serving the society and the country.

III. Program Educational Objectives (PEOs)

PEO1: To enable graduates to pursue research or take up successful career in academia or industries allied with Data Sciences or become entrepreneurs.

PEO2: To equip students with advanced techniques, tools and competency in applying technology to develop innovative and sustainable solutions.

PEO3: To empower students with critical analysis, leadership and decision-making skills guided by professional, ethical, and societal considerations to serve the nation.

IV. Program Outcomes (POs)

PO1: Demonstrate proficiency in the applied fields of Data Sciences and Computer Science.

PO2: Write and present a substantial technical report/document

PO3: Independently carry out research investigation and development work to solve practical problems

PO4: Perform self-learning and to keep oneself up-to-date in the field of Data Science and computer science engineering.

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

PO5: Develop creative, innovative solutions for real life problems.

PO6: Demonstrate team building, goal setting and leadership development skills to create successful entrepreneurs.

V. PEO/PO Mapping

Following three levels of correlation should be used:

1: Low

2: Medium

3: High

	PO1	PO2	PO3	PO4	PO5	PO6
PEO1	3	3	1	2	3	2
PEO2	3	2	3	3	3	3
PEO3	3	2	3	3	3	3

Dr. K. KARTHIKEYAN M.Sc.,M.Phil., Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

KPRIET

M.TECH DATA SCIENCES

REGULATIONS – 2023

CHOICE BASED CREDIT SYSTEM

CURRICULUM FOR I TO IV SEMESTERS

SEMESTER I

S.NO.	COURSE CODE	COURSE TITLE	CATEGORY	L	T	P	C
THEORY							
1	P23MA104	Probability and Statistics for Data Sciences	FC	3	0	0	3
2	P23CS101	Advanced Data Structures and Algorithms	PCC	2	0	2	3
3	P23DS101	Foundations of Data Science	PCC	3	0	0	3
4	P23CS103	Machine Learning	PCC	3	0	0	3
5	P23RMC01	Research Methodology and IPR	RMC	3	0	0	3
6		Professional Elective – I	PEC	3	0	0	3
PRACTICALS							
7	P23DS102	Data Science Laboratory	PCC	0	0	4	2
8	P23CS104	Machine Learning Laboratory	PCC	0	0	4	2
			TOTAL	17	0	10	22

SEMESTER II

S.NO.	COURSE CODE	COURSE TITLE	CATEGORY	L	T	P	C
THEORY							
1	P23MA201	Linear Algebra for Data Sciences	FC	3	0	0	3
2	P23DS201	Deep Learning	PCC	3	0	0	3
3	P23DS202	Big data Frameworks and Applications	PCC	3	0	0	3
4		Professional Elective – II	PEC	3	0	0	3
5		Professional Elective – III	PEC	3	0	0	3
PRACTICALS							
6	P23DS203	Big data Frameworks and Applications Laboratory	PCC	0	0	4	2
7	P23DS204	Technical Seminar	EEC	0	0	4	2
			TOTAL	15	0	8	19

SEMESTER III

S.NO.	COURSE CODE	COURSE TITLE	CATEGORY	L	T	P	C
THEORY							
1		Professional Elective IV	PEC	3	0	0	3
2		Professional Elective V	PEC	3	0	0	3
3		Professional Elective VI	PEC	3	0	0	3
PRACTICALS							
4	P23DS401	Project Work – Phase I	EEC	0	0	12	6
			TOTAL	9	0	12	15

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

SEMESTER IV

S.NO.	COURSE CODE	COURSE TITLE	CATEGORY	L	T	P	C
		PRACTICALS					
1	P23DS401	Project Work – Phase II	EEC	0	0	24	12
2	P23DSI01	Industrial Training / Internship	EEC	0	0	0	2
			TOTAL	0	0	24	14

LIST OF COURSES BASED ON ITS CATEGORY

FOUNDATION COURSES (FC)

S.NO.	COURSE CODE	COURSE TITLE	CATEGORY	L	T	P	C
1	P23MA104	Probability and Statistics for data sciences	FC	3	0	0	3
2	P23MA201	Linear Algebra for data sciences	FC	3	0	0	3

PROFESSIONAL CORE COURSES (PCC)

S.NO.	COURSE CODE	COURSE TITLE	CATEGORY	L	T	P	C
		THEORY					
1	P23CS101	Advanced Data Structures and Algorithms	PCC	2	0	2	3
2	P23DS101	Foundations of Data Science	PCC	3	0	0	3
3	P23CS103	Machine Learning	PCC	3	0	0	3
4	P23DS102	Data Science Laboratory	PCC	0	0	4	2
5	P23CS104	Machine Learning Laboratory	PCC	0	0	4	2
6	P23DS201	Deep Learning	PCC	3	0	0	3
7	P23DS202	Big data Frameworks and Applications	PCC	3	0	0	3
8	P23DS203	Big data Frameworks and Applications Laboratory	PCC	3	0	0	3

PROFESSIONAL ELECTIVES COURSES (PEC)

S.NO.	COURSE CODE	COURSE TITLE	CATEGORY	L	T	P	C
1	P23DSP01	Data Preparation and Analysis	PEC	3	0	0	3
2	P23DSP02	Web Analytics	PEC	3	0	0	3
3	P23DSP03	Information and Network Security	PEC	3	0	0	3

RESEARCH METHODOLOGY & IPR COURSES (RMC)

S.NO.	COURSE CODE	COURSE TITLE	CATEGORY	L	T	P	C
1	P23RMC01	Research Methodology and IPR	RMC	3	0	0	3

EMPLOYABILITY ENHANCEMENT COURSES (EEC)

S.NO.	COURSE CODE	COURSE TITLE	CATEGORY	L	T	P	C
1	P23DS204	Technical Seminar	EEC	0	0	4	2
2	P23DS301	Project Work – Phase I	EEC	0	0	12	6

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology

3	P23DS401	Project Work – Phase II	EEC	0	0	24	12
4	P23DSI01	Industrial Training / Internship	EEC	0	0	0	2

VIII. Scheme of Credit distribution – Summary

S. No	Stream	Credits/Semester				Credits	%	Suggested by AICTE
		I	II	III	IV			
1	Foundation Courses (FC)	3	3	-	-	6	8.57	-
2	Professional Core Courses (PCC)	13	8	-	-	21	30	32.35
3	Professional Elective Courses (PEC)	3	6	9	-	18	25.71	26.47
4	Research Methodology & IPR Courses (RMC)	3	-	-	-	3	4.28	2.94
5	Employability Enhancement Courses (EEC)	-	2	6	14	22	31.42	38.23
Total		**22**	**19**	**15**	**14**	**70**	**100**	**100**

TOTAL CREDITS: 70

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

Year	SEM	Subject	PO1	PO2	PO3	PO4	PO5	PO6
I Year	SEM I	Probability and Statistics for data sciences	✓	✓	–	✓	✓	–
		Advanced Data Structures and Algorithms	✓	–	✓	✓	✓	–
		Foundations of Data Science	✓	✓	✓	–	–	–
		Machine Learning	✓	–	–	✓	✓	–
		Research Methodology and IPR	–	–	–	✓	✓	✓
		Data Science Laboratory	✓	✓	✓	✓	✓	–
		Machine Learning Laboratory	✓	✓	✓	✓	✓	–
	SEM II	Linear Algebra for data sciences	✓	✓	✓	✓	–	✓
		Deep Learning	✓	✓	✓	✓	✓	–
		Big data Frameworks and Applications	✓	✓	✓	–	–	–
		Big data Frameworks and Applications Laboratory	✓	✓	✓	✓	✓	–
		Technical Seminar	–	–	–	–	✓	✓
II Year	SEM III	Project Work – Phase I						
		Data Preparation and Analysis	✓	✓	✓	✓	✓	–
		Web Analytics	✓	✓	–	–	✓	–
		Information and Network Security	✓	✓	–	–	✓	–

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

SEMESTER I

		Category: FC			
P23MA104	PROBABILITY AND STATISTICS FOR DATA SCIENCES	L	T	P	C
		3	0	0	3

COURSE OBJECTIVES

- To understand the concepts of probability, random variable and distributions that are applicable in the field of engineering
- To understand the concepts of testing of hypothesis for small and large samples which plays an important role in testing of industrial products
- To introduce the principles and techniques of parameter estimation using the method of moments and the maximum likelihood estimation (MLE) method, normal populations, with a focus on applications in data analysis and statistical inference.

UNIT I PROBABILITY AND RANDOM VARIABLES 9

Classical axiomatic definitions of probability – Conditional probability – Total probability – Bayes' Theorem and independence – Discrete, continuous and mixed random variables – probability mass probability density and cumulative distribution functions – Mathematical expectation – Moments, moment generating function – Foundational Probability Concepts for Data Science Applications

UNIT II SPECIAL DISTRIBUTIONS 9

Binomial – Poisson – Geometric distributions – Uniform – Exponential – Gamma – Normal distributions – Functions of a Random Variable.

UNIT III JOINT DISTRIBUTIONS 9

Joint marginal and conditional distributions – Product moments – Correlation – Independence of random variables – Bivariate normal distribution.

UNIT IV LARGE AND SMALL SAMPLE TESTS 9

Large sample test: Test of significance – Test for population mean, proportion – Test for equality of two means, proportions – Test of variance – Small sample test: Students' t-test, test for population mean, quality of two population means – F-test for equality of two population variances – Chi square test for goodness of fit and for independence of attributes

UNIT V ESTIMATION 9

The method of moments and the method of maximum likelihood estimation – Confidence intervals for the mean(s) and variance(s) of normal populations.

Contact Periods:

Lecture: 45 Periods Tutorial: – Periods Practical: – Periods Total: 45 Periods

REFERENCES:

1. Ronald E. Walpole, Raymond H. Meyers and Sharon L. Meyers, "Probability and Statistics for Engineers and Scientists", 9th edition, Pearson Education, 2013.
2. Jay L. Devore, "Probability and Statistics for Engineering and Sciences", 9th edition, Cengage India Pvt. Ltd., 2020.
3. Trivedi K.S., "Probability and Statistics with Reliability, Queuing and Computer Science Applications", 2nd edition, John Wiley & Sons, 2015.
4. Douglas C Montgomery and George C Runger, "Applied Statistics and Probability for Engineers", 6th edition, John Wiley & Sons, 2016.

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

KPRIET
Learn Beyond

COURSE OUTCOMES:

Upon completion of the course, the student will be able tc

Cos	Statements	K–Level
CO1	Probability axioms and the moments of dis: : e and continuous random variables	Apply
CO2	Discrete probability distributions including req.. ments, mean and variance for making decisions	Understand
CO3	Explain the correlation and linear regression \ th respect to two dimensional random variables	Understand
CO4	Analyze large and small sample tests and perfc r ı small sample tests based on Chi-square, t and F distributions	Apply
CO5	Apply the method of moments and maximum i elihood estimation techniques to estimate population parameters an : onstruct confidence intervals for the mean and variance of norn al / distributed data, supporting data-driven decision making in (al-world analytics scenarios	Apply

COURSE ARTICULATION MATRIX:

Pos / Cos	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	3	–	2	2	–
CO2	3	3	–	2	2	–
CO3	3	3	–	2	2	–
CO4	3	3	–	2	2	–
CO5	3	3	–	2	2	–
CO	3	3	–	2	2	–
Correlation levels:	1: Slight (Low)		2: Moc e' te (Med u ı)		3: Substantial (High)	

*Role Play / Group Discussions / Debates / Oral Presenta ı ; ıs / Poster Presentations / Technical presentations can also be provided. Course Designer can c h ; ıse any one / two components based on the nature of the course.

Dr. K. KARTHIKEYAN M.Sc.,M.Ph : ı.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Techr gy
Coimbatore - 641 407

KPRIET Learn Beyond

SEMESTER I

		Category: PCC			
P23CS101	ADVANCED DATA STRUCTURES AND ALGORITHMS	L	T	P	C
		2	0	2	3

COURSE OBJECTIVES:

- To acquire knowledge in role of algorithms and tree structures
- To gain knowledge on the graphs and linear programming
- To understand the concept of NP completeness and approximation algorithms

UNIT I ROLE OF ALGORITHMS IN COMPUTING 6

Algorithms – Algorithms as a Technology – Insertion Sort – Analyzing Algorithms – Designing Algorithms – Growth of Functions: Asymptotic Notation – Standard Notations and Common Functions – Recurrences: The Substitution Method – The Recursion – Tree Method

UNIT II HIERARCHICAL DATA STRUCTURES 6

Binary Search Trees: Basics – Querying a Binary search tree – Insertion and Deletion- Red-Black trees: Properties of Red-Black Trees – Rotations – Insertion – Deletion – B-Trees: Definition of Btrees – Basic operations on B-Trees – Deleting a key from a B-Tree – Fibonacci Heaps: structure – Mergeable – heap operations

UNIT III GRAPHS 6

Elementary Graph Algorithms: Representations of Graphs – Breadth-First Search – Depth – First Search – Topological Sort – Strongly Connected Components – Minimum Spanning Trees: Growing a Minimum Spanning Tree – Kruskal and Prim – Single – Source Shortest Paths: The Bellman-Ford algorithm – Single-Source Shortest paths in Directed Acyclic Graphs – Dijkstra's Algorithm

UNIT IV MULTITHREADED ALGORITHMS AND LINEAR PROGRAMMING 6

Dynamic multithreading – Multithreaded matrix multiplication – Multithreaded merge sort – Linear Programming – Standard and slack forms – Formulating problems as linear programs – Simplex algorithm – Duality – Feasible solution

UNIT V NP-COMPLETENESS AND APPROXIMATION ALGORITHMS 6

Polynomial time – verification – NP-completeness and reducibility – NP-completeness proofs – NP-complete problems – Approximation Algorithms – Vertex-cover problem – Traveling-salesman problem – Set-covering problem – Randomization and linear programming – Subset-sum problem

LIST OF EXPERIMENTS

1. Write a Python program that reads an infix expression and converts the expression to postfix form.
2. Write a Python program to perform the following operations:
   a) Construct a binary search tree of elements.
   b) Search for a key element in the above binary search tree.
   c) Delete an element from the above binary search tree.
3. Write a Python program to perform the following operations:
   a) Insertion into a B-tree
   b) Searching in a B-tree
4. Write a Python program to perform the following operation:

Dr. K. KARTHIKEYAN M.Sc.,M.Phil .,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

a) Construct a red black tree of elements.

5. Write a Python program for heap operations
6. Write a Python program for find single source shor t s bath for given graph using Dijktra's Algorithm
7. Write a Python program for traveling salesman probler ı ı ing any of the technique

Contact Periods:

Lecture: 30 Periods Tutorial: Periods Practic a : 30 Periods Total: 60 Periods

REFERENCES:

1. Thomas H. Cormen, Charles E. Leiserson, Ronalc . Rivest, Clifford Stein, Introduction toAlgorithms, Third Edition, Prentice-Hall, 2019
2. Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman, — ɔ ı a Structures and Algorithms, Pearson Education, Reprint 2016
3. S.Sridhar,Design and Analysis of Algorithms, 1st Edition. (ford University Press, 2014
4. Mark Allen Weiss, "Data Structures and Algorithm Analy s s in C", Pearson Education, 2nd Edition, 2012
5. E. Horowitz, S.Sahni and Dinesh Mehta, "Fundamental s : Data structures in C++", Universities Press, 2010

COURSE OUTCOMES:

Upon completion of the course, the student will be able to

COs	Statements	K–Level
CO1	Explain the role of algorithms in computing	Understand
CO2	Apply the importance of hierarchical data structur e s	Apply
CO3	Describe the elementary graph algorithms	Understand
CO4	Analyze the multithreaded algorithms and linear p r ramming	Analyze
CO5	Apply NP Completeness and approximation algor ı ıs	Apply

COURSE ARTICULATION MATRIX:

POs / COs	PO1	PO2	PO3	P O4	PO5	PO6
CO1	3	-	2	2	1	-
CO2	3	-	2	2	1	-
CO3	3	-	2	2	1	-
CO4	3	-	2	2	1	-
CO5	3	-	2	2	1	-
CO	3	-	2	2	1	-

Correlation levels:
1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)

*Role Play / Group Discussions / Debates / Oral Presentat o ; / Poster Presentations / Technical presentations can also be provided. Course Designer can ch o : se any one / two components based on the nature of the course.

Dr. K. KARTHIKEYAN M.S M.Phil P h D
Professor & Hea
Department of Mathe cs
KPR Institute of Engineering ar echnology

KPRIET
Learn Beyond

SEMESTER I

		Category: PCC			
		L	T	P	C
P23DS101	FOUNDATIONS OF DATA SCIENCE	3	0	0	3

COURSE OBJECTIVES

- To provide foundational data science knowledge and highlight the significance of statistics and optimization in performing mathematical operations within the field.
- To develop an understanding of handling diverse data types and utilizing visualization techniques for improved comprehension.
- To provide fundamental insights into various open-source data science tools and their practical applications in addressing industrial challenges.

UNIT I FUNDAMENTALS OF DATA SCIENCE 9

Introduction to data science – Typology of problems – Importance of linear algebra – Statistics and optimization – Structured thinking – Structured and unstructured data

UNIT II STATISTICAL FOUNDATIONS 9

Descriptive statistics – Statistical features – Summarizing data – Outlier analysis – Univariate – Bivariate and multivariate statistics – Dimensionality reduction – Over/under-sampling – Bayesian statistics and Statistical modelling

UNIT III ALGORITHMIC FOUNDATIONS 9

Linear algebra (matrices, eigenvalues, eigenvectors) – Distance measures – Projections – Hyperplanes – Spectral graph theory – Graph sampling – Random walks – MCMC algorithms – Learning separators – PAC learning

UNIT IV OPTIMIZATION 9

Unconstrained optimization – Gradient descent methods – Constrained optimization – KKT conditions – Least squares optimization – Introduction to non-gradient optimization techniques

UNIT V PROGRAMMING, DATA HANDLING, AND VISUALIZATION 9

Python basics (variables, loops, data structures) – Exploratory data analysis (EDA) – Data acquisition and preprocessing – Handling text data – Visualization (charts, workflows) and tools (SciPy, R, Weka)

Contact Periods:

Lecture: 45 Periods Tutorial: – Periods Practical: – Periods Total: 45 Periods

REFERENCES:

1. R. V. Hogg, J. W. McKean and A. Craig, Introduction to Mathematical Statistics, 8th Ed., Pearson Education India, 2019.
2. Avrim Blum, John Hopcroft, Ravindran Kannan, "Foundations of Data Science", Cambridge University Press, 2020.
3. Ani Adhikari and John DeNero, „Computational and Inferential Thinking: The Foundations of Data Science", GitBook, 2019.
4. Cathy O"Neil and Rachel Schutt, „Doing Data Science: Straight Talk from the Frontline", O"Reilly Media, 2013.
5. Hossein Pishro-Nik, "Introduction to Probability, Statistics, and Random Processes", Kappa Research, LLC, 2014.

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

KPRIET
Learn Beyond

COURSE OUTCOMES (CO)

Upon completion of the course, students will be able to

COs	Statements	K–Level
CO1	Obtain foundational knowledge of data science cc n : pts and principles.	Understand
CO2	Demonstrate expertise in conducting statistical an : / is on data.	Apply
CO3	Apply mathematical concepts and optimization t : l niques to execute data science operations.	Apply
CO4	Manage and visualize diverse data types us n ; programming for effective knowledge representation.	Apply
CO5	Utilize various open-source data science tools : c ıddress real-world challenges through industrial case studies.	Apply

COURSE ARTICULATION MATRIX:

POs / Cos	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	2	1	-	-	-
CO2	3	1	1	-	-	-
CO3	3	2	1	-	-	-
CO4	3	2	1	-	-	-
CO5	3	2	1	-	-	-
CO	3	2	1	-	-	-
Correlation levels:	1: Slight (Low)		2: Mod : e (Medi .		3: Substantial (High)	

*Role Play / Group Discussions / Debates / Oral Presenta i s / Poster Presentations / Technical presentations can also be provided. Course Designer can c h se any one / two components based on the nature of the course.

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

KPRIET
Learn Beyond

SEMESTER I

		Category: PCC			
P23CS103	MACHINE LEARNING	L	T	P	C
		3	0	0	3

COURSE OBJECTIVES:

- To acquire knowledge in learning methods to enhance the performance of learning
- To gain knowledge on the suitable machine learning techniques for data handling
- To evaluate the performance of algorithms, provide solution for various real-world applications.

UNITI SUPERVISED LEARNING ALGORITHMS 9

Machine Learning – Applications – Supervised Learning – Learning a Class – Vapnik-Chervonenkis Dimension – Probably Approximately correct learning – Noise – Learning Multiple Classes – Regression – Model selection and Generalization – Dimensions – Bayesian Decision Theory – Classification – Discriminant Functions – Association Rules – Parametric Methods

UNITII CLUSTERING AND NON-PARAMETRIC METHODS 9

Introduction to clustering – Expectation Maximization Algorithm – Mixtures of Latent Variable Models– Spectral Clustering – Spectral Clustering – Choosing the Number of Clusters – Nonparametric Density Estimation – Generalization to Multivariate Data – Nonparametric Classification – Condensed Nearest Neighbor – Nonparametric Regression: Smoothing Models

UNITIII MULTILAYER PERCEPTRONS AND LOCAL MODELS 9

Introduction – Training a Perceptron – Learning Boolean Functions – MLP as a Universal Approximator – Backpropagation Algorithm – Training Procedures – Bayesian View of Learning – Learning Time – Deep Learning – Competitive Learning – The Mixture of Experts – Hierarchical Mixture of Experts

UNITIV KERNEL MACHINES, GRAPHICAL AND HIDDEN MARKOV MODELS 9

Optimal Separating Hyperplane – Kernel Trick – Vectorial Kernels – Multiple Kernel Learning – Kernel Machines for Regression and Ranking – Canonical Cases for Conditional Independence – Belief Propagation – Undirected Graphs: Markov Random Fields – Three Basic Problems of HMMs – Learning Model Parameters – Model Selection in HMMs

UNITV DESIGN AND ANALYSIS OF MACHINE LEARNING EXPERIMENTS 9

Introduction – Factors, Response, and Strategy of Experimentation – Response Surface Design – Randomization, Replication, and Blocking – Cross-Validation and Resampling Methods – Measuring Classifier Performance – Interval Estimation – Hypothesis Testing – Comparing Two Classification Algorithms – Comparison over Multiple Datasets – Multivariate Tests

Contact Periods:

Lecture: 45 Periods Tutorial: –Periods Practical: –Periods Total: 45 Periods

REFERENCES:

1. Ethem Alpaydin,"Introduction to Machine Learning", MIT Press, Prentice Hall of India, 3rd Edition 2017
2. Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, "Foundations of Machine Learning", MIT Press, 2018
3. Tom Mitchell, "Machine Learning", McGraw Hill, 3rd Edition,2008
4. Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong, "Mathematics for Machine Learning",Cambridge University Press, 2019
5. Stephen Marsland, "Machine Learning – An Algorithmic Perspective", 2nd Edition, CRC Press,2015

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

KPRIET
Learn Beyond

COURSE OUTCOMES:

Upon completion of the course, the student will be able to

COs	Statements	K–Level
CO1	Explain the concept of supervised learning	Understand
CO2	Use of clustering and non-parametric methods to real world problems	Apply
CO3	Describe the multi-layer perceptrons and local models	Understand
CO4	Analyze the kernel machines, graphical and hi n Markov models	Analyze
CO5	Design and analyze the machine learning experiments using various testing methods	Analyze

COURSE ARTICULATION MATRIX:

POs / COs	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	-	-	2	-	-
CO2	3	-	2	2	3	-
CO3	3	-	-	2	-	-
CO4	3	-	2	2	-	-
CO5	3	-	3	3	3	-
CO	3	-	2	2	3	-
Correlation levels:	1: Slight (Low)	2: Moderate (Medium)		3: Substantial (High)		

*Role Play / Group Discussions / Debates / Oral Presentations / Poster Presentations / Technical presentations can also be provided. Course Designer can choose any one / two components based on the nature of the course.

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,P
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

SEMESTER I

		Category: RMC			
		L	T	P	C
P23RMC01	RESEARCH METHODOLOGY AND IPR	3	0	0	3

COURSE OBJECTIVES:

- To impart knowledge in problem formulation, analysis and solutions
- To impart skills required for technical paper writing / presentation without violating professional ethics.
- To familiarize knowledge on patent drafting and filing patents.

UNITI RESEARCH PROBLEM FORMULATION 9

Meaning of research problem – Sources of research problem – Criteria characteristics of a good research problem – Errors in selecting a research problem – Scope and objectives of research problem. Approaches of investigation of solutions for research problem – Data collection – Analysis – Interpretation– Necessary instrumentations

UNITII LITERATURE REVIEW AND DATA COLLECTION 9

Effective literature studies approaches – Analysis – Plagiarism and research ethics. Method of data collection – Types of data – Primary Data – Scales of measurement – Source and collection of data observation method – Secondary data

UNITIII TECHNICALWRITING / PRESENTATION 9

Effective technical writing: How to write report – Paper – Developing a research proposal – Format of research proposal – Presentation and assessment by a review committee

UNITIV INTRODUCTION TO INTELLECTUALPROPERTY RIGHTS (IPR) 9

Nature of Intellectual Property: Patents – Designs – Trade and Copyright. Process of Patenting and Development – technological research – Innovation – Patenting – Development – International Scenario – International cooperation on Intellectual Property – Procedure for grants of patents – Patenting under PCT

UNITV INTELLECTUAL PROPERTYRIGHTS(IPR) 9

Patent Rights: Scope of Patent Rights – Licensing and transfer of technology – Patent information and databases – Geographical Indications – New Developments in IPR – Administration of Patent System – IPR of Biological Systems – Computer Software etc

Contact Periods:

Lecture: 45 Periods Tutorial: –Periods Practical: –Periods Total: 45 Periods

REFERENCES:

1. Ranjit Kumar, "Research Methodology: A Step by Step Guide for beginners", 3rd edition, Sage Publications,2010
2. Cooper, DR and Schindler, P S., "Business Research Methods", 9th edition, Tata McGraw Hill, 2014
3. Robert P.Merges, Peter S, Menell, Mark A.Lemley, "Intellectual Property in New Technological age", Volume I, Clause 8 Publishing,2022

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

KPRIET
Learn Beyond

COURSE OUTCOMES:

Upon completion of the course, the student will be able tc

COs	Statements	K–Level
CO1	Formulate research problem	Apply
CO2	Carryout research analysis	Analyze
CO3	Develop research proposal	Evaluate
CO4	Draft process of patenting	Apply
CO5	File and publish patents in R&D	Evaluate

COURSE ARTICULATION MATRIX:

POs / COs	PO1	PO2	PO3	PO4	PO5	PO6
CO1	–	–	–	1	3	1
CO2	–	–	–	1	3	1
CO3	–	–	–	1	3	1
CO4	–	–	–	1	3	1
CO5	–	–	–	1	3	1
CO	–	–	–	1	3	1
Correlation levels:	1: Slight (Low)	2: Moderate (M e	ım)	3: Substantial (High)		

*Role Play / Group Discussions / Debates / Oral Presenta : ıs / Poster Presentations / Technical presentations can also be provided. Course Designer can : : ose any one / two components based on the nature of the course.

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,)
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Techno
Coimbatore - 641 407

SEMESTER I

		Category: PCC			
P23DS102	DATA SCIENCE LABORATORY	L	T	P	C
		0	0	4	2

COURSE OBJECTIVES:

- To provide hands-on experience with foundational techniques in data science, including data analysis, dimensionality reduction, and statistical modelling.
- To enable students to implement mathematical and optimization algorithms in Python for solving real-world data science problems.

List of Experiments

1. Implement descriptive statistics (mean, median, mode, variance, standard deviation) of a dataset.
2. Implement outlier detection algorithms (Z-score, IQR) and handle missing values.
3. Implement univariate statistical analysis and visualize data using bar plots and histograms.
4. Implement bivariate analysis using scatter plots and calculate correlation coefficients.
5. Implement Principal Component Analysis (PCA) for dimensionality reduction and visualize the results.
6. Implement t-SNE for non-linear dimensionality reduction.
7. Implement basic linear algebra operations (matrix multiplication, determinant, inversion) and compute eigenvalues/eigenvectors.
8. Implement distance measures like Euclidean, Manhattan, and Cosine distances.
9. Implement the gradient descent algorithm for minimizing a quadratic function and visualize its convergence.
10. Implement constrained optimization problems using KKT conditions and visualize the process.

Contact Periods:

Lecture: – Periods Tutorial: – Periods Practical: 60 Periods Total: 60 Periods

COURSE OUTCOMES:

Upon completion of the course, the student will be able to

Cos	Statements	K-Level
CO1	Demonstrate the ability to implement statistical techniques such as descriptive statistics and outlier detection on datasets.	Apply
CO2	Apply dimensionality reduction methods like PCA and t-SNE to simplify high-dimensional data and visualize patterns.	Apply
CO3	Implement linear algebra operations and distance measures essential for data representation and analysis.	Apply
CO4	Develop optimization algorithms like gradient descent and constrained optimization using KKT conditions.	Apply
CO5	Demonstrate expertise in using programming and visualization tools to address real-world data science challenges.	Apply

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology

KPRIET
Learn Beyond

COURSE ARTICULATION MATRIX:

POs / Cos	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	1	2	1	1	-
CO2	3	1	2	1	1	-
CO3	3	1	2	1	1	-
CO4	3	1	2	1	1	-
CO5	3	1	2	1	1	-
CO	3	1	2	1	1	-
Correlation levels: 1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)						

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

SEMESTER I

		Category: PCC			
P23CS104	MACHINE LEARNING LABORATORY	L	T	P	C
		0	0	4	2

COURSE OBJECTIVES:

- To acquire knowledge in learning methods to enhance the performance of learning
- To implement the suitable machine learning techniques for data handling
- To evaluate the performance of algorithms and to provide solution for various real-world applications

List of Experiments

1. Implement Decision Tree learning
2. Implement Logistic Regression
3. Implement classification using Multilayer perceptron
4. Implement classification using SVM
5. Implement Bagging using Random Forests
6. Implement k-nearest Neighbours algorithm
7. Implement K-means, K-Modes Clustering to Find Natural Patterns in Data
8. Implement Hierarchical clustering
9. Implement Gaussian Mixture Model Using the Expectation Maximization
10. Implement Principal Component Analysis for Dimensionality Reduction

Contact Periods:

Lecture: – Periods Tutorial: – Periods Practical: 60 Periods Total: 60 Periods

COURSE OUTCOMES:

Upon completion of the course, the student will be able to

COs	Statements	K-Level
CO1	Implement the concept of supervised learning	Apply
CO2	Use of clustering and non-parametric methods to real world problems	Apply
CO3	Implement the multi-layer perceptrons and local models using classification techniques	Apply
CO4	Analyze the K-means, K-Modes Clustering to Find Natural Patterns in Data	Apply
CO5	Implement Principal Component Analysis for Dimensionality Reduction	Apply

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

KPRIET
Learn Beyond

COURSE ARTICULATION MATRIX:

POs / COs	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	1	2	1	1	-
CO2	3	1	2	1	1	-
CO3	3	1	2	1	1	-
CO4	3	1	2	1	1	-
CO5	3	1	2	1	1	-
CO	3	1	2	1	1	-
Correlation levels: 1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)						

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

KPRIET
Learn Beyond

PROFESSIONAL ELECTIVE – I

P23DSP01	DATA PREPARATION AND ANALYSIS	Category: PEC			
		L	T	P	C
		3	0	0	3

COURSE OBJECTIVES:

- To prepare and preprocess diverse datasets using appropriate data gathering, cleaning, and transformation techniques
- To apply exploratory data analysis methods for identifying patterns, relationships, and generating hypotheses
- To develop meaningful and interactive data visualizations using modern tools and technologies

UNIT I DATA GATHERING AND PREPARATION 9

Data Formats – Structured – Semi-structured – Unstructured data (CSV, JSON, XML, SQL tables, log files) – Parsing and Transformation – Extracting raw data – Parsing techniques for text – web scraping – APIs-Data normalization – Aggregation – Feature extraction – Scalability & Real-Time Issues – Big data challenges (volume, velocity, variety) – Batch vs. real-time data processing (Hadoop, Spark, Kafka)

UNIT II DATA CLEANING 9

Consistency Checking – Removing duplicates – Fixing typos – Ensuring integrity – Heterogeneous & Missing Data-Handling multiple sources / formats – Missing values: deletion – Imputation – Interpolation – Data Transformation & Segmentation – Scaling – Encoding – Normalization – Splitting datasets (train/test, clusters, segments)

UNIT III EXPLORATORY ANALYSIS 9

Descriptive & Comparative Statistics – Mean – Median – Variance – Standard deviation – Correlations – Comparative measures: t-test – Chi-square – Clustering & Association – K-means – Hierarchical clustering – DBSCAN – Market basket analysis – Association rules (Apriori, FP-Growth) – Hypothesis Generation – Formulating testable assumptions – Statistical testing – Significance

UNIT IV VISUALIZATION 9

Designing Visualizations – Principles: clarity – Minimalism – Storytelling – Time Series – Trend – Seasonality – Forecasting plots – Geo located Data – Heat maps – Choropleth maps – GPS – based visualizations – Correlations & Connections – Scatter plots – Pair plots – Network graphs – Hierarchies & Networks – Tree maps – Dendrograms – Graph visualizations – Interactivity – Dashboards – Interactive plots

UNIT V VISUALIZATIONS USING R OR PYTHON 9

R Libraries – Ggplot2 – Plotly – Shiny – Python Libraries – Matplotlib – Sea born – Plotly – Bokeh – Dash – Case Studies – Real-world datasets – Building interactive dashboards

Contact Periods:

Lecture: 45 Periods Tutorial: – Periods Practical: – Periods Total 45 Periods

REFERENCES:

1. Anil Maheshwari – Data Analytics Made Accessible, 2nd Edition, Kindle Direct Publishing, 2017.
2. Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar – Introduction to Data Mining, 2nd Edition, Pearson, 2019.
3. Trevor Hastie, Robert Tibshirani, Jerome Friedman – The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition, Springer, 2017.

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology

4. Alex Gutman, Jordan Goldmeier – Becoming a Data He: How to Think, Speak, and Understand Data Science, Statistics, and Machine Learning, Wiley, 2 21.
5. Glenn J. Myatt, "Making sense of Data: A practical Gu to Exploratory Data Analysis and Data Mining", 2nd Edition, John Wiley Publishers, 2014.

COURSE OUTCOMES:

Upon completion of the course, the student will be able to

COs	Statements	K-Level
CO1	Analyze structures and unstructured datasets	Apply
CO2	Data cleaning and transformation methods to h idle inconsistencies, missing values, and heterogeneous data sources	Apply
CO3	Statistical, clustering, and association techniques t explore datasets and generate testable hypotheses	Apply
CO4	Statistical, clustering, and association techniques t explore datasets and generate testable hypotheses	Apply
CO5	R and Python visualization libraries to build inte tive dashboards and analyze real-world datasets	Apply

COURSE ARTICULATION MATRIX:

POs / COs	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	2	2	2	3	-
CO2	3	3	2	2	2	-
CO3	3	3	2	3	2	-
CO4	2	2	3	2	3	-
CO5	2	2	3	2	3	-
CO	3	2	2	2	3	-
Correlation levels: 1: Slight (Low) 2: Moderate (Medium) ubstantial (High)						

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

KPRIET
Learn Beyond

SEMESTER II

		Category: FC			
P23MA201	**LINEAR ALGEBRA FOR DATA SCIENCES**	**L**	**T**	**P**	**C**
		3	**0**	**0**	**3**

COURSE OBJECTIVES

- To understand the concepts of decomposition of matrices
- To understand the concepts of independence, basis and dimensions in vector spaces
- To understand the concepts of inner product spaces and orthogonality

UNIT I MATRICES 9

Vectors and Linear combinations – Rank of a matrix – Solution of system of linear equations by Gaussian elimination, Gauss Jordan and LU decomposition methods

UNIT II VECTOR SPACE 9

Vector spaces – Subspaces – Linear combinations and linear system of equations – Linear dependence and independence – Bases and dimensions

UNIT III INNER PRODUCT SPACE 9

Introduction to linear transformation – Inner product – Norm – Angle - Orthogonality: definition and simple problems – Projections – Gram Schmidt orthogonalization and QR decomposition

UNIT IV EIGENVALUE PROBLEMS 9

Linear transformations – Range, kernel and problems – Eigenvalues and eigenvectors – Hermitian and unitary matrices (simple problems)

UNIT V PRINCIPAL COMPONENT ANALYSIS 9

Positive definite matrices – Cayley-Hamilton theorem – Singular value decomposition and principal component analysis using the covariance method – Introduction to their applications in image processing and machine learning (problems not included)

Contact Periods:

Lecture: 45 Periods Tutorial: - Periods Practical: – Periods Total 45 Periods

REFERENCES:

1. Howard Anton and Chris Rorres, "Elementary Linear Algebra – Applications version", 9th edition, John Wiley & Sons, 2005.
2. David C. Lay, "Linear Algebra and its Applications", 5th edition, Pearson College Division, 2014.
3. Steven J. Leon, "Linear Algebra with Applications", 9th edition, Pearson College Division, 2014.
4. Gilbert Strang, "Introduction to Linear Algebra", 5th edition, Wellesley Publishers, 2016.
5. Gonzalez R C and Woods R E, "Digital Image Processing", 4th edition, Pearson Education, 2018.

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

COURSE OUTCOMES:

Upon completion of the course, the student will be able to

COs	Statements	K-Level
CO1	Implement the various matrix techniques in solv ; the system of linear equations	Understand
CO2	Use the concept of vector spaces to predict an c ›normal basis	Understand
CO3	Attribute a set of vectors in an inner produ ;pace using Gram-Schmidt orthogonalisation and decompose a n matrix using QR decomposition	Understand
CO4	Discuss the Eigen values and Eigen v ›rs of the linear transformations for the simple real-life problems	Understand
CO5	Apply the Singular value decomposition anc rincipal component analysis technique to real world datasets or performing the dimensional reduction on the given data	Apply

COURSE ARTICULATION MATRIX:

POs / COs	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	2	1	1	-	1
CO2	3	2	1	1	-	1
CO3	3	2	1	1	-	1
CO4	3	2	2	1	-	1
CO5	3	2	2	1	-	1
CO	3	2	1	1	-	1
Correlation levels:	1: Slight (Low)		2: Moder (Medium)		3: Substantial (High)	

*Role Play / Group Discussions / Debates / Oral Preser t ›ns / Poster Presentations / Technical presentations can also be provided. Course Designer car ›ose any one / two components based on the nature of the course.

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

SEMESTER II

		Category: PCC			
P23DS201	DEEP LEARNING	L	T	P	C
		3	0	0	3

COURSE OBJECTIVES:

- To explain the basic concepts of neural networks and deep networks.
- To discuss the major architectures of deep networks.
- To demonstrate the applications of deep learning

UNIT I BASICS OF NEURAL NETWORKS 9

Neural Network Basics–Binary Classification, Logistic Regression, Gradient Descent, Derivatives, Computation graph, Vectorization, Training Neural Networks – Activation Functions – Loss Functions – Hyper parameters.

UNIT II FUNDAMENTALS OF DEEP NETWORKS 9

Defining Deep Learning – Common Architectural Principles of Deep Networks – Building Blocks of Deep Networks.

UNIT III MAJOR ARCHITECTURES OF DEEP NETWORKS 9

Unsupervised Pre–Trained Networks – Convolutional Neural Networks – Recurrent Neural Networks – Recursive Neural Networks –Tuning Deep Networks.

UNIT IV TUNING SPECIFIC DEEP NETWORK ARCHITECTURES 9

Convolution Neural Networks (CNNs)– Recurrent Neural Networks– Restricted Boltzmann Machines– DBNs.

UNIT V APPLICATIONS 9

Large-Scale deep learning – Computer Vision – Speech Recognition – Natural Language Processing– Recommender systems. Case Study– Applications of Deep Learning in Health care, Deep learning tools-TensorFlow, Keras, MatConvNet.

Contact Periods:

Lecture: 45 Periods Tutorial: - Periods Practical: – Periods Total: 45 Periods

REFERENCES:

1. Adam Gibson, Josh Patterson, "Deep Learning, A Practitioner's Approach", O'Reilly Media, 2017.
2. Ian Goodfellow, YoshuaBengio and Aaron Courville, "Deep Learning", MIT Press, 2016.
3. Python Machine Learning by Example, Yuxi (Hayden) Liu, First Edition, 2017.
4. Daniel Graupe, "Deep Learning Neural Networks: Design and Case Studies", World Scientific Publishing , 2016.
5. Yu and Li Deng, "Deep Learning: Methods and Applications", Now Publishers Inc 2014.

Dr. K. KARTHIKEYAN M.Sc.,M.Phil Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

KPRIET
Learn Beyond

COURSE OUTCOMES:

Upon completion of the course, the student will be able to

COs	Statements	K-Level
CO1	Distinguish neural and deep networks	Understand
CO2	Select the appropriate deep network architecture	Understand
CO3	Analyze the performance of a deep learning network.	Analyze
CO4	Apply deep learning for solving real world problems	Apply
CO5	Develop new deep network models	Apply

COURSE ARTICULATION MATRIX:

POs / COs	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	2	2	3	3	-
CO2	3	3	2	2	3	-
CO3	3	3	3	3	3	-
CO4	3	3	3	3	3	-
CO5	3	3	3	3	3	-
CO	3	2	2	2	3	-

Correlation levels:
1: Slight (Low) 2: Moderate (Medium). 3: Substantial (High)

Dr. K. KARTHIKEYAN M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

SEMESTER II

		Category: PCC			
P23DS202	BIGDATA FRAMEWORKS AND APPLICATIONS	L	T	P	C
		3	0	0	3

COURSE OBJECTIVES:

- To provide knowledge of functional programming and big data frameworks for effective data science applications.
- To familiarize students with the production-level challenges of ML models and impart knowledge on MLOps principles, tools, and platforms.
- To develop skills for designing and deploying scalable machine learning systems and handling real-time data problems in enterprise settings.

UNIT I FOUNDATIONS OF BIG DATA 9

Understanding Big Data: Concepts and terminology – Big Data Characteristics – Types of Data – Identifying Data Characteristics – Big Data Architecture – Big Data Storage – File System vs Distributed File System – NoSQL Databases – Sharding – Replication – ACID and BASE Properties

UNIT II HADOOP ECOSYSTEM AND MAPREDUCE 9

Hadoop Architecture – Hadoop Distributed File System (HDFS) – YARN Overview – Hadoop I/O – MapReduce Framework – Developing MapReduce Applications – Workflow – Types and Formats – Features of MapReduce – Sorting – Joins and Pipelining

UNIT III HADOOP TECHNOLOGIES - PIG AND HIVE 9

Introduction to Pig: Pig Architecture – Data Model – Grunt – Pig Latin (Input, Output, Relational Operators, User-Defined Functions) – Working with Scripts – Introduction to Hive – Modules – Data Types – File Formats – HiveQL (Data Definition and Manipulation) – Views – Queries – Scripts – Indexes – Bucketing vs Partitioning

UNIT IV SPARK FRAMEWORK AND PROGRAMMING 9

Overview of Spark – Comparison of Hadoop vs Spark – Cluster Design and Management – Spark Context – Resilient Distributed Datasets (RDD) – RDD Operations – Lazy Evaluation – Spark Jobs – Writing Spark Applications – Programming in Scala – Python – R and Java – Application Execution

UNIT V ADVANCED SPARK: SQL AND GRAPHX 9

Spark SQL – SQL Context – Data Import / Export – Data Frames – Using SQL Queries – GraphX Overview – Creating Graphs – Graph Algorithms for Big Data Analysis – Real-world Applications of Spark SQL and GraphX

Contact Periods:

Lecture: 45 Periods Tutorial: – Periods Practical: – Periods Total: 45 Periods

REFERENCES:

1. Tanvir Habib Sardar and Bishwajeet Kumar Pandey, Big Data Computing: Advances in Technologies, Methodologies, and Applications, 2024.
2. homas Erl, WajidKhattak, and Paul Buhler, Big Data Fundamentals: Concepts, Drivers & Techniques, Pearson India Education Service Pvt. Ltd., First Edition, 2016.
3. Zaigham Mahmood (Editor), Data Science and Big Data Computing: Frameworks and Methodologies, Springer, 2016.
4. Big Data Analytics: Theory, Techniques, Platforms, and Applications, 2023 Edition.
5. Technologies and Applications for Big Data Value, 2021 Edition.

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

KPRIET
Learn Beyond

COURSE OUTCOMES:

Upon completion of the course, the student will be able to

COs	Statements	K-Level
CO1	Understand the need of new frame work to deal w t uge amounts of Data.	Understand
CO2	Demonstrate the Hadoop framework Hadoop Dis uted File System and MapReduce.	Apply
CO3	Demonstrate the Pig architecture and evaluation o ; scripts.	Apply
CO4	Describe the Hive architecture and execute SQL eries on sample data sets.	Apply
CO5	Demonstrate spark programming with different prc mming languages and graph algorithms.	Apply

COURSE ARTICULATION MATRIX:

POs / COs	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	2	1	-	-	-
CO2	3	1	1	-	-	-
CO3	3	2	1	-	-	-
CO4	3	2	1	-	-	-
CO5	3	2	1	-	-	-
CO	3	2	1	-	-	-
Correlation levels: 1: Slight (Low) 2: Moderate (dium) 3: Substantial (High)						

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

SEMESTER II

P23DS203	BIG DATA FRAMEWORKS AND APPLICATION LABORATORY	Category: PCC			
		L	T	P	C
		0	0	2	2

COURSE OBJECTIVES:

- To gain practical experience with Big Data frameworks and tools for processing, analyzing, and managing large datasets.
- To develop skills in building, deploying, and optimizing machine learning models in big data environments.

List of Experiments:

1. Set up a multi-node Hadoop cluster and perform advanced HDFS operations like distcp and snapshots.
2. Develop a MapReduce program for calculating word frequency, demonstrating the use of combiners and partitioners.
3. Implement a MapReduce program with Avro file format and evaluate its performance for structured data.
4. Set up Hive, create tables using partitioning and bucketing, and write HiveQL queries for data analysis.
5. Write Pig scripts to process large-scale log data, extract meaningful patterns, and store the results in HDFS.
6. Use Distributed Cache to perform a Map-side join for relational datasets and compare performance with Reduce-side join.
7. Build a Spark application to filter, group, and aggregate large-scale data using Data Frames.
8. Perform complex RDD transformations, including joins, cartesian products, and custom partitioning in Spark.
9. Create a real-time data streaming application using Spark Streaming to process live data from a socket or file source.
10. Use Spark SQL to query large-scale structured data, integrate it with a NoSQL database, and visualize the results.

Contact Periods:

Lecture: – Periods Tutorial: – Periods Practical: 30 Periods Total: 30 Periods

COURSE OUTCOMES:

Upon completion of the course, the student will be able to

COs	Statements	K-Level
CO1	Analyze and optimize big data workflows using Hadoop, Spark, and related tools for processing large datasets.	Apply
CO2	Develop and implement data processing applications with MapReduce, Hive, and Spark for structured and real-time data.	Apply
CO3	Deploy machine learning models in production, focusing on handling real-time data streams.	Apply
CO4	Address challenges in data processing, such as data drift, using Spark and other big data tools.	Apply
CO5	Demonstrate how to integrate Big Data tools to solve real-world problems efficiently.	Apply

KPRIET
Learn Beyond

COURSE ARTICULATION MATRIX:

POs / COs	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	2		2	3	-
CO2	3	3	2	2	3	-
CO3	3	2	1	1	-	-
CO4	3	3	1	2	-	-
CO5	3	3	2	2	3	-
CO	3	2	1	1	3	-
Correlation levels: 1: Slight (Low) 2: Moderate (Medium) 3: ubstantial (High)						

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

PROFESSIONAL ELECTIVE – II

P23DSP02	WEB ANALYTICS	Category: PEC			
		L	T	P	C
		3	0	0	3

COURSE OBJECTIVES:

- To understand the importance of qualitative data and apply appropriate techniques for gaining insights
- To develop a customer-centric approach in analyzing and interpreting data for decision-making
- To apply principles, tools, and methods of web intelligence and analytics in real-world business situations.

UNIT I WEB ANALYTICS 9

Web Analytics – Basics – Traditional Ways – Expectations – Data Collection – Clickstream Data Weblogs – Beacons – JavaScript Tags – Packet Sniffing – Outcomes data – Competitive data – Search Engine Data

UNIT II QUALITATIVE ANALYSIS 9

Qualitative Analysis – Customer Centricity –Site Visits – Surveys – Questionnaires – Website surveys – Post visits – Creating and Running – Benefits of surveys – Critical components of successful strategy

UNIT III WEB ANALYTIC CONCEPTS 9

Web Analytic concepts – URLS – Cookies – Time on site – Page views – Understand standard reports – Website content quality – Navigation reports (top pages, top destinations, site overlay) – Search Analytics – Internal search – SEO and PPC – Measuring Email and Multichannel Marketing – Competitive intelligence and Web 2.0 Analytics – Segmentation – Connectable reports

UNIT IV SEARCH ENGINE ANALYTICS 9

Search Engine Analytics – Analytics – Cookies – Accounts vs Property – Tracking Code – Tracking Unique Visitors - Demographics - Page Views & Bounce Rate Acquisitions – Custom Reporting

UNIT V GOALS & FUNNELS 9

Goals & Funnels – Filters – Ecommerce Tracking – Real Time Reports – Customer Data Alert – Ad Words Linking – Ad Sense Linking – Attribution Modelling – Segmentation – Campaign Tracking – Multi-Channel Attribution

Contact Periods:

Lecture: 45 Periods Tutorial: – Periods Practical: Periods Total 45 Periods

REFERENCES:

1. Avinash Kaushik, "Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity", 1st Edition, Sybex, 2009.
2. Michael Beasley, "Practical Web Analytics for User Experience: How Analytics can help you Understand your Users", 1st Edition, Morgan Kaufmann, 2013..
3. Magy Seif El-Nasr, Anders Drachen, Alessandro Canossa, "Game Analytics: Maximizing the Value of Player Data", 1st Edition, Springer, 2013.
4. Bing Liu, "Web Data Mining: Exploring Hyperlinks, Content, and Usage Data", 2nd Edition, Springer, 2011.
5. Justin Cutroni, "Google Analytics", 1st Edition, O'Reilly, 2010. Eric Fettman, Shiraz Asif, Feras Alhlou, "Google Analytics Breakthrough", John Wiley & sons, 2016.

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

COURSE OUTCOMES:

Upon completion of the course, the student will be able to

COs	Statements	K-Level
CO1	Explain traditional and modern web data collection methods such as clickstream, logs, beacons, and tags	Understand
CO2	Design and implement effective website surveys to gather customer-centric insights	Apply
CO3	Apply search analytics techniques (internal search, SEO, PPC) for improving website visibility and performance.	Apply
CO4	Develop custom reports to measure visitor acquisition, page engagement, and bounce	Apply
CO5	Apply e-commerce tracking and integration tools (Ad Words, Ad Sense, real-time reports to evaluate marketing effectiveness	Apply

COURSE ARTICULATION MATRIX:

POs / COs	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	2	-	-	2	-
CO2	2	2	3	2	2	2
CO3	2	2	3	2	2	2
CO4	2	3	3	2	3	-
CO5	2	3	3	2	3	-
CO	3	2	-	-	2	-
Correlation levels: 1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)						

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

KPRIET
Learn Beyond

PROFESSIONAL ELECTIVE – III

		Category: PEC			
P23DSP03	**INFORMATION AND NETWORK SECURITY**	**L**	**T**	**P**	**C**
		3	0	0	3

COURSE OBJECTIVES:

- To understand the concepts of information security, various types of attacks, and the role of encryption and authentication in securing systems
- To learn the principles and mechanisms of intrusion prevention, detection, and firewall technologies.
- To develop the ability to analyze, detect, and apply preventive measures against security threats and attacks.

UNIT I FUNDAMENTALS OF INFORMATION SECURITY 9

Critical characteristics of Information – NSTISSC Security Model – Components of information System – SDLC – Information assurance – Security Threats and vulnerabilities – Overview of Security threats – Security Standards

UNIT II CRYPTOGRAPHY AND ACCESS CONTROL 9

Classical Cryptography – Symmetric Cryptography – Asymmetric Cryptography – Modern Cryptography – Access Control – DRM – Steganography –Biometrics

UNIT III NETWORK AND E-COMMERCE SECURITY 9

Network Security – Intrusion Prevention – Detection and Management – Firewall – Ecommerce Security – Computer Forensics – Security for VPN and Next Generation Networks

UNIT IV HOST & APPLICATION SECURITY 9

Host and Application security – Control hijacking – Software architecture and a simple buffer overflow – Common exploitable application bugs – Shellcode – Buffer Overflow – Side-channel attacks –Timing attacks – Power analysis – Cold-boot attacks – Defences – Malware – Viruses and worms spyware, key loggers, and botnets; defences auditing, policy – Defending weak applications – Isolation, sandboxing, virtual machines

UNIT V MOBILE & WIRELESS SECURITY 9

Mobile, GSM and Wireless LAN security – Protection measures – Business risk analysis – Information Warfare and Surveillance – Case study on Attack prevention, detection and response

Contact Periods:

Lecture: 45 Periods Tutorial: – Periods Practical: – Periods Total 45 Periods

REFERENCES:

1. William Stallings, "Cryptography and Network Security: Principles and Practice", 6th Edition, PHI,2014.
2. Michael E. Whitman and Herbert J Mattord, "Principles of Information Security", 6th Edition, Vikas Publishing House, 2017.
3. Bill Nelson, Amelia Phillips, F. Enfinger and Christopher Stuart, "Guide to Computer Forensics and Investigations", 4th Edition, Thomson Course Technology, 2010.
4. Matt Bishop, "Computer Security: Art and Science", 1st Edition, Addison-Wesley Professional, 2015.

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407

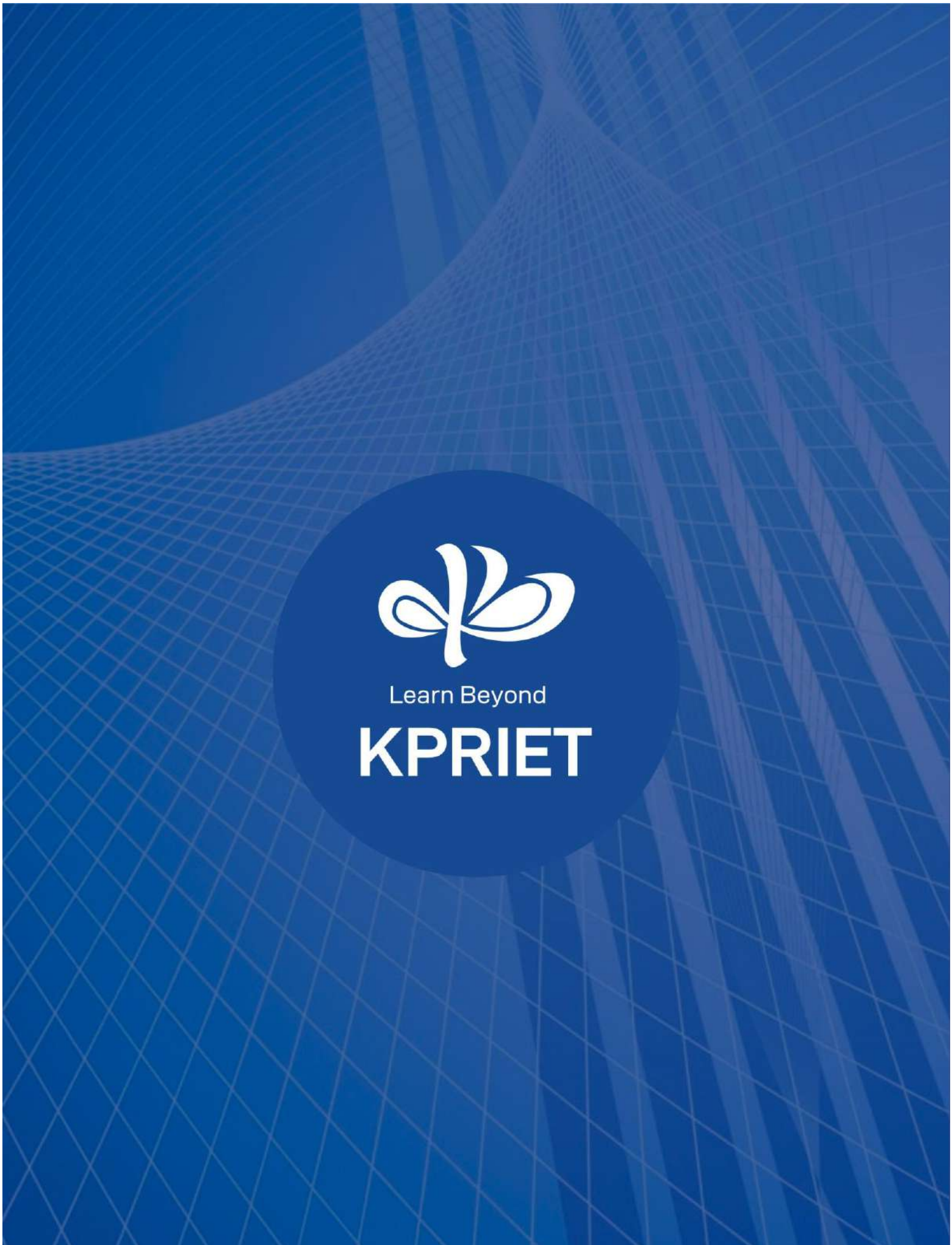
Upon completion of the course, the student will be able to

COs	Statements	K-Level
CO1	Explain the critical characteristics of information security models, and standards, along with threats and vulnerabilities	Understand
CO2	Apply the System Development Life Cycle (SDLC) and information assurance principles in designing secure systems	Apply
CO3	Describe the principles of classical, symmetric, and asymmetric cryptography, as well as modern techniques such as biometrics and DRM	Apply
CO4	Explain the optical properties of materials	Apply
CO5	Explain the fundamentals of intrusion detection, firewalls, VPNs, and e-commerce security.(	Apply

COURSE ARTICULATION MATRIX:

POs / COs	PO1	PO2	PO3	PO4	PO5	PO6
CO1	3	2	-	-	2	-
CO2	2	2	3	2	2	2
CO3	2	2	3	2	2	2
CO4	2	3	3	2	3	-
CO5	2	3	3	2	3	-
CO	3	2	-	-	2	-
Correlation levels: 1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)						

Dr. K. KARTHIKEYAN M.Sc.,M.Phil.,Ph.D
Professor & Head
Department of Mathematics
KPR Institute of Engineering and Technology
Coimbatore - 641 407


Learn Beyond
KPRIET