AC 20.08.22 ITEM NO: 1.1.3

Deccan Education Society's

Kirti M. Doongursee College of Arts, Science and Commerce (AUTONOMOUS)

Affiliated to

UNIVERSITY OF MUMBAI

Syllabus for
Program: Master of Science
Course: M.Sc.
Subject: Computer Science with Specialization Data Science

Choice Based Credit System (CBCS)
with effect from
Academic Year 2022-2023

PROGRAMME OUTCOME

PROGRAMME SPECIFIC OUTCOMES (PSOs)

On completion of M.Sc. Data Science programme, students will be able:

PO_01: To become a skilled Data Scientist in industry, academia, or government.

PO_02: To use specialised software tools for data storage, analysis and visualization.

PO_03: To independently carry out research/investigation to solve practical problems.

PO_04: To gain problem-solving ability- to assess social issues (ethical, financial, management, analytical and scientific analysis) and engineering problems.

PO_05: To have a clear understanding of professional and ethical responsibility.

PO_06: To collaboratevirtually.

PO_07: To have critical thinking and innovative skills.

PO_08: To translate vast data into abstract concepts and to understand database reasoning.

PROGRAMME STRUCTURE

Semester – I		
Course Code	Course Title	Credits
KPSDS22101	Programming Paradigms	4
KPSDS22102	Database Technologies	4
KPSDS22103	Fundamentals of Data Science	4
KPSDS22104	Statistical Methods for Data Science	4
KPSDS221P1	Programming Paradigms Practical	2
KPSDS221P2	Database Technologies Practical	2
KPSDS221P3	Fundamentals of Data Science Practical	2
KPSDS221P4	Statistical Methods for Data Science Practical	2
	Total Credits	24

Semester – II		
Course Code	Course Title	Credits
KPSDS22201	Artificial Intelligence and Machine Learning	4
KPSDS22202	Soft Computing	4
KPSDS22203	Algorithms for Data Science	4
KPSDS22204	Optimization Techniques	4
KPSDS222P1	Artificial Intelligence and Machine Learning Practical	2
KPSDS222P2	Soft Computing Practical	2
KPSDS222P3	Algorithms for Data Science Practical	2
KPSDS222P4	Optimization Techniques Practical	2
	Total Credits	24

DETAILED SYLLABUS FOR SEMESTER - I & SEMESTER - II

Semester – 1

Programming Paradigms

M.Sc (Data Science)		Semester – I	
Course Name: **Programming Paradigms**		Course Code: KPSDS22101	
Periods per week (1 Period is 60 minutes)		4	
Credits		4	
		Hours	Marks
Evaluation System	Theory Examination	2½	60
	Internal	--	40

Course Objectives:

- To understand the basic building blocks of programming Languages.
- To Learn and understand various programming paradigms.

Learning Outcomes:

After successful completion of this course, students would be able to

- To independently carry out research/investigation to solve practical problems through various programming languages.

Unit	Details	Lectures
I	Foundations-Language design, why to study programming language, compilation and interpretation, programming environments. Programming language syntax – Specifying syntax: regular expressions and Context-Free grammar(Token and Regular expressions, Context Free grammar, Derivations and parse trees), Scanning(Generating Finite automation, Scanner code, Table-driven scanning, Lexical errors, pragmas), Parsing(Recursive Descent, Writing L1 grammar, Table driven top down parsing, Bottom up parsing, Syntax errors)	12
II	OBJECT ORIENTATION Basic concepts: objects, classes, methods, overloading methods, messages inheritance: overriding methods,	12

	single inheritance, multiple inheritance Interfaces, encapsulation, polymorphism.	
III	FUNCTIONAL PROGRAMMING Definition of a function: domain and range, total and partial functions, strict functions. Recursion, Referential transparency, Side effects of functions	12
IV	LOGIC PROGRAMMING Basic constructs, Facts: queries, existential queries, conjunctive queries and rules. Definition and semantics of a logic program, Recursive programming: Computational model of logic programming, Goal reduction, Negation in logic programming	12
V	SCRIPTING LANGUAGE What is scripting language, Problem domain(Shell languages, Text processing and report generation, Mathematics and statistics, General purpose scripting, Extension languages), Scripting the world wide web(CGI scripts, Embedded server side script, client side script, Java Applets, XSLT)	12

Books and References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	Programming Language Pragmatics	Michael Scott	Morgan Kaufmann	4th Edition	2015
2.	The Craft of Functional Programming	Thompson, Simon. Haskell:	Addison-Wesley Professional	2nd Editon	2011
3.	“Foundations of Programming Languages Design & Implementation”	RoostaSeyed	Cenage learning	3rdEditon	2003
4.	Programming Languages: Concepts and Constructs	Sethi Ravi	Pearson Education	3rdEditon	2000

Programming Paradigms Practical

<table>
<tr><td colspan="2">M. Sc. (Data Science)</td><td colspan="2">Semester – I</td></tr>
<tr><td colspan="2">Course Name: Programming Paradigms Practical</td><td colspan="2">Course Code: KPSDS221P1</td></tr>
<tr><td colspan="2">Periods per week (1 Period is 60 minutes)</td><td colspan="2">4</td></tr>
<tr><td colspan="2">Credits</td><td colspan="2">2</td></tr>
<tr><td colspan="2"></td><td>Hours</td><td>Marks</td></tr>
<tr><td rowspan="2">Evaluation System</td><td>Practical Examination</td><td>2</td><td>50</td></tr>
<tr><td>Internal</td><td>--</td><td>--</td></tr>
</table>

Practical:

Perform minimum ten practical based on the basic concepts of each programming paradigm covering the entire syllabus.

Course Outcomes:

- To explore a range of modern programming languages and programming techniques.
- To select appropriate software development tools for given application environments.

Database Technologies

<table>
<tr><td colspan="2">M.Sc (Data Science)</td><td colspan="2">Semester – I</td></tr>
<tr><td colspan="2">Course Name: Database Technologies</td><td colspan="2">Course Code: KPSDS22102</td></tr>
<tr><td colspan="2">Periods per week (1 Period is 60 minutes)</td><td colspan="2">4</td></tr>
<tr><td colspan="2">Credits</td><td colspan="2">4</td></tr>
<tr><td colspan="2"></td><td>Hours</td><td>Marks</td></tr>
<tr><td rowspan="2">Evaluation System</td><td>Theory Examination</td><td>2½</td><td>60</td></tr>
<tr><td>Internal</td><td>--</td><td>40</td></tr>
</table>

Course Objectives:

- The objective of the course is to present an introduction to database management systems, with an emphasis on how to organize, maintain and retrieve - efficiently, and effectively - information from a DBMS.

Learning Outcomes:

After successful completion of this course, students would be able to

- Students will develop the ability to build and assess Databased models.

Unit	Details	Lectures
I	Database Concepts: Why Databases?, Data versus Information, Introducing the Database, Why Database Design Is Important, Evolution of File System Data Processing, Problems with File System Data Processing, Database Systems Data Models:Data Modeling and Data Models, The Importance of Data Models, Data Model Basic Building Blocks, Business Rules, The Evolution of Data Models, Degrees of Data Abstraction The Relational Database Model:A Logical View of Data, Keys, Integrity Rules, Relational Algebra, The Data Dictionary and the System Catalog, Relationships within the Relational Database, Data Redundancy Revisited Entity Relationship (ER) Modeling: The Entity Relationship Model, Developing an ER Diagram, Database Design Challenges: Conflicting Goals	12
II	Advanced Data Modelling:The Extended Entity Relationship Model, Entity Clustering, Design Cases: Learning Flexible Database Design Normalization of Database Tables:Database Tables and Normalization, The Need for Normalization, The Normalization Process, Improving the Design Introduction to Structured Query Language (SQL):Introduction to SQL, Basic SELECT Queries, SELECT Statement Options, FROM Clause Options, ORDER BY Clause Options, WHERE Clause Options, Aggregate Processing, Subqueries, SQL Functions, Relational Set Operators, Crafting SELECT Queries Advanced SQL:Data Definition Commands, Creating Table Structures, Altering Table Structures, Data Manipulation Commands, Virtual Tables: Creating a View, Sequences, Procedural SQL, Embedded SQL Transaction Management and Concurrency Control:What Is a Transaction?, Concurrency Control, Concurrency Control with Locking Methods, Concurrency Control with Time Stamping Methods,	12

	Concurrency Control with Optimistic Methods, ANSI Levels of Transaction Isolation, Database Recovery Management	
III	Three Database Revolutions: Early Database Systems, The First Database Revolution, The Second Database Revolution, The Third Database Revolution Google, Big Data, and Hadoop: The Big Data Revolution, Google: Pioneer of Big Data, Hadoop: Open-Source Google Stack Sharding, Amazon, and the Birth of NoSQL: Scaling Web 2.0, Amazon's Dynamo Document Databases: XML and XML Databases, JSON Document Databases	12
IV	Tables are Not Your Friends: Graph Databases: What is a Graph?, RDBMS Patterns for Graphs, RDF and SPARQL, Property Graphs and Neo4j, Gremlin, Graph Database Internals, Graph Compute Engines Column Databases: Data Warehousing Schemas, The Columnar Alternative, Sybase IQ, C-Store, and Vertica, Column Database Architectures The End of Disk? SSD and In-Memory Databases: The End of Disk?, In-Memory Databases, Berkeley Analytics Data Stack and Spark Distributed Database Patterns: Distributed Relational Databases, Nonrelational Distributed Databases, MongoDB Sharding and Replication, HBase, Cassandra Consistency Models: Types of Consistency, Consistency in MongoDB, HBase Consistency, Cassandra Consistency	12
V	Data Models and Storage: Data Models, Storage Languages and Programming Interfaces: SQL, NoSQL APIs, The Return of SQL Databases of the Future: The Revolution Revisited, Counterrevolutionaries, Can We have it All?, Meanwhile, Back at Oracle HQ, Other Convergent Databases, Disruptive Database Technologies	12

Books and References:					
Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Database System	Carlos	Cengage	13th	2018

	designs, Implementation & Management	Coronel, Steven Morris			
2	Next Generation Databases	Guy Harrison	Apress	1st	2015
3	Advanced Database Technology and Design	Mario Piattini, Oscar Díaz	Artech House	1st	2000

Database Technologies Practical

M. Sc. (Data Science)		Semester – I	
Course Name: **Database Technologies Practical**		Course Code: KPSDS221P2	
Periods per week (1 Period is 60 minutes)		4	
Credits		2	
		Hours	Marks
Evaluation System	Practical Examination	2	50
	Internal	--	--

Practical:

Perform minimum ten practical based on the basic concepts of each Database Technologies covering the entire syllabus.

Course Outcomes:

Upon successful completion of this course, students should be able to:

- Describe the fundamental elements of relational database management systems
- Explain the basic concepts of relational data model, entity-relationship model, relational database design, relational algebra and SQL
- Design ER-models to represent simple database application scenarios
- Convert the ER-model to relational tables, populate relational database and formulate SQL queries on data.
- Improve the database design by normalization.

<table>
<tr><td colspan="2">M.Sc (Data Science)</td><td colspan="2">Semester – I</td></tr>
<tr><td colspan="2">Course Name: Fundamentals of Data Science</td><td colspan="2">Course Code: KPSDS22103</td></tr>
<tr><td colspan="2">Periods per week (1 Period is 60 minutes)</td><td colspan="2">4</td></tr>
<tr><td colspan="2">Credits</td><td colspan="2">4</td></tr>
<tr><td colspan="2"></td><td>Hours</td><td>Marks</td></tr>
<tr><td rowspan="2">Evaluation System</td><td>Theory Examination</td><td>2½</td><td>60</td></tr>
<tr><td>Internal</td><td>--</td><td>40</td></tr>
</table>

Course Objectives:

- To provide strong foundation for data science and application in area related to it and understand the underlying core concepts and emerging technologies in data science.

Learning Outcomes:

After successful completion of this course, students would be able to

- To use specialized software tools for data storage, analysis and visualization for data science

<table>
<tr><th>Unit</th><th>Details</th><th>Lectures</th></tr>
<tr><td>I</td><td>Introduction to Data Science:
What is Data? Kinds of data: e.g. static, spatial, temporal, text, media,
Introduction to high level programming language + Integrated Development
Environment (IDE)
Describing data: Exploratory Data Analysis (EDA) + Data Visualization - Summaries, aggregation, smoothing, distributions
Data sources: e.g. relational databases, web/API, streaming,
Data collection: e.g. sampling, design (observational vs experimental) and its impact on visualization, modeling and generalizability of results</td><td>12</td></tr>
<tr><td>II</td><td>Data analysis/modeling:
Question/problem formation along with EDA
Introduction to estimation and inference (testing and confidence intervals) including simulation and resampling
Scope of inference
Assessment and selection e.g. training and testing sets</td><td>12</td></tr>
</table>

	Data Curation, Management and Organization-I
	▪ Query languages and operations to specify and transform data (e.g. projection, selection, join, aggregate/group, summarize)
	▪ Structured/schema based systems as users and acquirers of data
	   o Relational (SQL) databases, APIs and programmatic access, indexing
	   o XML and XPath, APIs for accessing and querying structured data contained therein	

III	Data Curation, Management and Organization-I
	▪ Semi-structured systems as users and acquirers of data
	   o Access through APIs yielding JSON to be parsed and structured
	▪ Unstructured systems in the acquisition and structuring of data
	   o Web Scraping
	   o Text/string parsing/processing to give structure
	
	Data Curation, Management and Organization-II
	▪ Security and ethical considerations in relation to authenticating and authorizing access to data on remote systems
	▪ Software development tools (e.g. github, version control)
		12

IV	Data Curation, Management and Organization-II
	▪ Large scale data systems
	   o Paradigms for distributed data storage
	   o Practical access to example systems (e.g. MongoDB, HBase, NoSQL systems)
	   o Amazon Web Services (AWS) provides public data sets in Landsat, genomics, multimedia
	Introduction to Statistical Models
	▪ Simple Linear Regression
	▪ Multiple Linear Regression
	▪ Logistic Regression
	▪ Review of hypothesis testing, confidence intervals, etc.
	▪ Estimation e.g. likelihood principle, Bayes,
		12

V	Introduction to Statistical Models	12

	▪ Linear models o Regression theory i.e. least-squares: Introduction to estimation principles o Multiple regression ▪ Transformations, model selection ▪ Interactions, indicator variables, ANOVA o Generalized linear models e.g. logistic, etc. ▪ Alternatives to classical regression e.g. trees, smoothing/splines ▪ Introduction to model selection o Regularization, bias/variance tradeoff e.g. parsimony, AIC, BIC o Cross validation Ridge regressions and penalized regression e.g. LASSO	

Books and References:					
Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Hands-On Programming with R	Garrett Grolemund	O'Reilly	1st	2014
2	Doing Data Science	Rachel Schutt, Cathy O'Neil	O'Reilly Media	1st	2013
3	An Introduction to Statistical Learning with Applications in R	Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani:	Springer US	2nd	2021
4	Applied Predictive Modelling	M. Kuhn, K. Johnson	Springer New York	3rd	2019
5	Mastering Machine Learning with R	Cory Lesmeister	Packt Publishing	2nd	2015

Fundamentals of Data Science Practical

M. Sc (Data Science)		Semester – I	
Course Name: **Fundamentals of Data Science Practical**		Course Code: KPSDS221P3	
Periods per week 1 Period is 60 minutes	Lectures	4	
	Credits	2	
		Hours	Marks
Evaluation System	Practical Examination	2	50

Practical:

Perform minimum ten practical based on the basic concepts of each unit of Fundamentals of Data Science Practical covering the entire syllabus.

Course Outcomes:

- The students will be able to independently carry out research/investigation to solve practical problems
- The students should be able to understand & comprehend the problem; and should be able to define suitable statistical method to be adopted.

Statistical Methods for Data Science

M. Sc (Data Science)		Semester – I	
Course Name: **Statistical Methods for Data Science**		Course Code:KPSDS22104	
Periods per week 1 Period is 60 minutes	Lectures	4	
	Credits	4	
		Hours	Marks
Evaluation System	Theory Examination	2½	60
	Theory Internal	--	40

Pre requisites :

- Knowledge of statistics and mathematical concepts

Course Objectives:

- To present the mathematical, statistical and computational challenges

of building neural networks • To study the concepts of deep learning • To enable the students to know deep learning techniques to support real-time applications

Unit	Details	Lectures
I	Introduction to Applied Statistics: The Nature of Statistics and Inference, What is “Big Data”?, Statistical Modelling, Statistical Significance Testing and Error Rates, Simple Example of Inference Using a Coin,Statistics Is for Messy Situations, Type I versus Type II Errors, Point Estimates and Confidence Intervals, Variable Types, Sample Size, Statistical Power, and Statistical Significance, The Verdict on Significance Testing, Training versus Test Data.	12
II	Computational Statistics:Vectors and Matrices, The Inverse of a Matrix, Eigenvalues and Eigenvectors Means, Correlations, Counts: Drawing Inferences: Computing z and Related Scores, Statistical Tests, Plotting Normal Distributions, Correlation Coefficients, Evaluating Pearson’s r for Statistical Significance, Spearman’s Rho: A Nonparametric Alternative to Pearson, Tests of Mean Differences, t-Tests for One Sample, Two-Sample t-Test, Paired-Samples t-Test, Categorical Data, Binomial Test, Categorical Data Having More Than Two Possibilities.	12
III	Power Analysis and Sample Size Estimation:Power for t-Tests, Power for One-Way ANOVA, Power for Correlations. Analysis of Variance: Fixed Effects, Random Effects, Mixed Models, Introducing the Analysis of Variance (ANOVA), Performing the ANOVA, Random Effects ANOVA and Mixed Models, One-Way Random Effects ANOVA, Simple and Multiple Linear Regression,Simple Linear Regression, Multiple Regression Analysis, Hierarchical Regression, How Forward Regression Works,	12
IV	Logistic Regression and the Generalized Linear Model:Logistic Regression, Logistic Regression, Predicting Probabilities, Multiple Logistic Regression, Training Error Rate Versus Test Error Rate. Multivariate Analysis of Variance (MANOVA) and Discriminant Analysis:Multivariate Tests of Significance, Example of MANOVA, Outliers, Homogeneity of	12

	Covariance Matrices, Linear Discriminant Function Analysis, Theory of Discriminant Analysis, Predicting Group Membership, Visualizing Separation	
V	Principal Component Analysis: Principal Component Analysis Versus Factor Analysis, Properties of Principal Components, Component Scores, How Many Components to Keep?, Exploratory Factor Analysis, Common Factor Analysis Model, Factor Analysis Versus Principal Component Analysis on the Same, Initial Eigenvalues in Factor Analysis, Rotation in Exploratory Factor Analysis, Estimation in Factor Analysis Cluster Analysis:k-Means Cluster Analysis, Minimizing Criteria, Example of k-Means Clustering, Hierarchical Cluster Analysis, Why Clustering Is Inherently Subjective, Nonparametric Tests, Mann–Whitney U Test, Kruskal–Wallis Test, Nonparametric Test for Paired Comparisons and Repeated	12

,

Books and References:					
Sr. No.	Title	Author/s	Publisher	Edition	Year
01	Univariate, Bivariate, and Multivariate Statistics Using R	Daniel J. Denis	Wiley	1st	2020
02	Practical Data Science	Andreas François Vermeulen	APress	1st	2018
03	Data Science from Scratch first Principle in python	Joel Grus	Shroff Publishers	1st	2017
04	Experimental Design in Data science with Least Resources	N C Das	Shroff Publishers	1st	2018

Statistical Methods for Data Science Practical

M. Sc (Data Science)		Semester – I
Course Name: Statistical Methods for Data Science Practical		Course Code: KPSDS221P4
Periods per week 1 Period is 60 minutes	Lectures	4

	Credits	2	
		Hours	Marks
Evaluation System	Practical Examination	2	40

Practical:

Perform minimum ten practical based on the basic concepts of each Statistical Methods for Data Science covering the entire syllabus.

Course Outcomes:

At the end of successful completion of the course the student will be able to:

- Describe basics of mathematical foundation that will help the learner to understand the concepts of Deep Learning.
- Understand and describe model of deep learning
- Design and implement various deep supervised learning architectures for text & image data.
- Design and implement various deep learning models and architectures.
- Apply various deep learning techniques to design efficient algorithms for real-world applications.

SEMESTER-II

Artificial Intelligence and Machine Learning

M. Sc (Data Science)		Semester – II	
Course Name: **Artificial Intelligence and Machine Learning**		Course Code: KPSDS22201	
Periods per week 1 Period is 60 minutes	Lectures	4	
	Credits	4	
		Hours	Marks
Evaluation System	Theory Examination	2½	60
	Theory Internal	--	40

Pre requisites : • Knowledge of Algorithms and mathematical foundation
Course Objectives: • To provide the foundations for AI problem-solving techniques and knowledge representation formalisms • Understanding Human learning aspects. • Understanding primitives in learning process by computer. • Understanding nature of problems solved with Machine Learning

Unit	Details	Lectures
I	Introduction to AI: The AI problems, AI technique, philosophy and development of Artificial intelligence. Minimax algorithm, alpha-beta pruning, stochastic games, Constraint- satisfaction problems. Knowledge and Reasoning: Logical agents, Propositional logic, First-order logic, Inference in FoL: forward chaining, backward chaining, resolution, Knowledge representation: Frames, Ontologies, Semantic web and RDF.	12
II	Introduction to PROLOG: Facts and predicates, data types, goal finding, backtracking, simple object, compound objects, use of cut and fail predicates, recursion, lists, simple input/output, dynamic database. Machine Learning: Machine learning, Examples of Machine Learning Problems, Structure of Learning, learning versus Designing, Training versus Testing, Characteristics of Machine learning tasks, Predictive and descriptive tasks, Machine learning Models: Geometric Models, Logical Models, Probabilistic Models. Features: Feature types, Feature	12

	Construction and Transformation, Feature Selection	
III	Classification and Regression: Classification: Binary Classification- Assessing Classification performance, Class probability Estimation Assessing class probability Estimates, Multiclass Classification. Regression: Assessing performance of Regression- Error measures, Overfitting- Catalysts for Overfitting, Case study of Polynomial Regression. Theory of Generalization: Effective number of hypothesis, Bounding the Growth function, VC Dimensions, Regularization theory.	12
IV	Linear Models: Least Squares method, Multivariate Linear Regression, Regularized Regression, Using Least Square regression for Classification. Perceptron, Support Vector Machines, Soft Margin SVM, Obtaining probabilities from Linear classifiers, Kernel methods for non-Linearity. Logic Based and Algebraic Model: Distance Based Models:Neighbours and Examples, Nearest Neighbours Classification, Distance based clustering-K means Algorithm, Hierarchical clustering,	12
V	Rule Based Models: Rule learning for subgroup discovery, Association rule mining. Tree Based Models: Decision Trees, Ranking and Probability estimation Trees, Regression trees, Clustering Trees. Probabilistic Model: Normal Distribution and Its Geometric Interpretations, Naïve Bayes Classifier, Discriminative learning with Maximum likelihood, Probabilistic Models with Hidden variables: Estimation-Maximization Methods, Gaussian Mixtures, and Compression based Models.	12

,

Books and References:					
Sr. No.	Title	Author/s	Publisher	Edition	Year
01	Artificial Intelligence	Elaine Rich, Kevin Knight	Tata McGraw Hill	3rd	2017
02	Machine Learning: The Art and Science of Algorithms that Make Sense of Data	Peter Flach	Cambridge University Press	1st	2012
03	Introduction to Statistical Machine Learning with	Hastie, Tibshirani, Friedman	Springer	2nd	2012

	Applications in R				
04	Introduction to Machine Learning	EthemAlpaydin	PHI	2nd	2013

Artificial Intelligence and Machine Learning Practical

<table>
<tr><td colspan="2">M. Sc (Data Science)</td><td colspan="2">Semester – II</td></tr>
<tr><td colspan="2">Course Name:Artificial Intelligence and Machine Learning Practical</td><td colspan="2">Course Code: KPSDS222P1</td></tr>
<tr><td colspan="2">Periods per week (1 Period is 60 minutes)</td><td colspan="2">4</td></tr>
<tr><td colspan="2">Credits</td><td colspan="2">2</td></tr>
<tr><td colspan="2"></td><td>Hours</td><td>Marks</td></tr>
<tr><td rowspan="2">Evaluation System</td><td>Practical Examination</td><td>2</td><td>50</td></tr>
<tr><td>Internal</td><td>--</td><td>-</td></tr>
</table>

Practical:

Perform minimum ten practical based on the basic concepts of each Artificial Intelligence and Machine Learning covering the entire syllabus.

Course Outcomes:

- Understand the key issues and concepts in Artificial Intelligence.
- Acquire the knowledge about classification and regression techniques where a learner will be able to explore his skill to generate data base knowledge using the prescribed techniques.
- Understand and implement the techniques for extracting the knowledge using machine learning methods.
- Achieve adequate perspectives of big data analytics in various applications like recommender systems, social media applications etc.
- Understand the statistical approach related to machine learning.He will also apply the algorithms to a real-world problem, optimize the models learned and report on the expected accuracy that can be achieved by applying the mode

Soft Computing

M. Sc (Data Science)		Semester – II	
Course Name: **Soft Computing**		Course Code:KPSDS22202	
Periods per week 1 Period is 60 minutes	Lectures	4	
	Credits	4	
		Hours	Marks
Evaluation System	Theory Examination	2½	60
	Theory Internal	--	40

Course Objectives:

- Soft computing concepts like fuzzy logic, neural networks and genetic algorithm, where Artificial Intelligence is mother branch of all.
- All these techniques will be more effective to solve the problem efficiently

Unit	Details	Lectures
I	Artificial Neural Network:Fundamental concepts,Evolution of neural network, basic model of Artificial Neural Network, Important terminologies, McCulloch Pits neuron, linear separability, Hebb network Supervised Learning Network: Perceptron networks, Adaline, MAdaline, Backpropogation network, Radial Basis Function, Time Delay Network, Functional Link Networks, Tree Neural Network.	12
II	UnSupervised Learning Networks: Fixed weight competitive nets, Kohonen self-organizing feature maps, learning vectors quantization, counter propogation networks, adaptive resonance theory networks. Associative Memory Networks: Training algorithm for pattern Association, Autoassociative memory network, hetroassociative memory network, bi-directional associative memory, Hopfield networks, iterative autoassociative memory networks, temporal associative	12

	memory networks.	
III	Special Networks: Simulated annealing, Boltzman machine, Gaussian Machine, Cauchy Machine, Probabilistic neural net, cascade correlation network, cognition network, neo-cognition network, cellular neural network, optical neural network Third Generation Neural Networks: Spiking Neural networks, convolutional neural networks, deep learning neural networks, extreme learning machine model.	12
IV	Introduction to Fuzzy Logic, Classical sets, Fuzzy sets, Classical Relations and Fuzzy Relations: Cartesian Product of relation, classical relation, fuzzy relations, tolerance and equivalence relations, non-iterative fuzzy sets. Membership Function: features of the membership functions, fuzzification and methods of membership value assignments. Defuzzification: Lambda-cuts for fuzzy sets, Lambda-cuts for fuzzy relations, Defuzzification methods. Fuzzy Arithmetic and Fuzzy measures: fuzzy arithmetic, fuzzy measures, measures of fuzziness, fuzzy integrals.	12
V	Genetic Algorithm: Biological Background, Traditional optimization and search techniques, genetic algorithm and search space, genetic algorithm vs. traditional algorithms, basic terminologies, simple genetic algorithm, general genetic algorithm, operators in genetic algorithm, stopping condition for genetic algorithm flow, constraints in genetic algorithm, problem solving using genetic algorithm, the schema theorem, classification of genetic algorithm, Holland classifier systems, genetic programming, advantages and limitations and applications of genetic algorithm	12

Books and References:					
Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	Artificial Intelligence and Soft Computing	Anandita Das Battacharya	SPD	3rd	2018
2.	Principles of Soft computing	S.N.Sivanandam S.N.Deepa	Wiley	3rd	2019

3.	Neuro-Fuzzy and Soft Computing	J.S.R.Jang, C.T.Sun and E.Mizutani	Prentice Hall of India	1st	2004
4.	Neural Networks, Fuzzy Logic and Genetic Algorithms: Synthesis & Applications	S.Rajasekaran, G. A. Vijayalakshami	Prentice Hall of India	1st	2004
5.	Fuzzy Logic with Engineering Applications	Timothy J.Ross	McGraw-Hill	1st	1997
6.	Genetic Algorithms: Search, Optimization and Machine Learning	Davis E.Goldberg	Addison Wesley	1st	1989
7.	Introduction to AI and Expert System	Dan W. Patterson	Prentice Hall of India	2nd	2009

Soft Computing Practical

M. Sc (Data Science)		Semester – II	
Course Name:Soft Computing Practical		Course Code:KPSDS222P2	
Periods per week (1 Period is 60 minutes)		4	
Credits		2	
		Hours	Marks
Evaluation System	Practical Examination	2	50
	Internal	--	-

Practical:

Perform minimum ten practical based on the basic concepts of each soft computing covering the entire syllabus.

Course Outcome:

- Identify and describe soft computing techniques and their roles in building intelligent machines
- Recognize the feasibility of applying a soft computing methodology for a particular problem
- Apply fuzzy logic and reasoning to handle uncertainty and solve engineering problems and also Apply neural networks for classification and regression problems

- Apply genetic algorithms to combinatorial optimization problems
- Evaluate and compare solutions by various soft computing approaches for a given problem.

Algorithms for Data Science

<table>
<tr><td colspan="2">M. Sc (Data Science)</td><td colspan="2">Semester – II</td></tr>
<tr><td colspan="2">Course Name: Algorithms for Data Science</td><td colspan="2">Course Code: KPSDS22203</td></tr>
<tr><td colspan="2">Periods per week (1 Period is 60 minutes)</td><td colspan="2">4</td></tr>
<tr><td colspan="2">Credits</td><td colspan="2">4</td></tr>
<tr><td colspan="2"></td><td>Hours</td><td>Marks</td></tr>
<tr><td rowspan="2">Evaluation System</td><td>Theory Examination</td><td>2½</td><td>60</td></tr>
<tr><td>Internal</td><td>--</td><td>40</td></tr>
</table>

Course Objectives:

The course is aimed at:

- Focussing on the principles of data reduction and core algorithms for analysing the data of data science
- Providing many opportunities to develop and improve programming skills
- Applying algorithms to real world data set
- Imparting design thinking capability to build big-data

Unit	Details	Lectures
I	Introduction: What Is Data Science?, Diabetes in America, Authors of the Federalist Papers, Forecasting NASDAQ Stock Prices, Algorithms, Python, R, Terminology and Notation Data Mapping and Data Dictionaries: Data Reduction, Political Contributions, Dictionaries, Tutorial: Big Contributors, Data Reduction, Election Cycle Contributions, Similarity Measures, Computing Similarity Scalable Algorithms and Associative Statistics: Introduction, Associative Statistics, Univariate Observations, Functions, Histogram Construction, Multivariate Data, Computing the Correlation Matrix, Linear Regression, Computing β	12

II	Hadoop and MapReduce: Introduction, The Hadoop Ecosystem, Medicare Payments, The Command Line Environment, Programming a MapReduce Algorithm, Using Amazon Web Services Data Visualization: Introduction, Principles of Data Visualization, Making Good Choices, Harnessing the Machine	12
III	Linear Regression Methods: Introduction, The Linear Regression Model, Introduction to R, Large Data Sets and R, Factors, Analysis of Residuals Healthcare Analytics: Introduction, The Behavioral Risk Factor Surveillance System, Diabetes Prevalence and Incidence, Predicting At-Risk Individuals, Identifying At-Risk Individuals, Unusual Demographic Attribute Vectors, Building Neighborhood Sets	12
IV	Cluster Analysis: Introduction, Hierarchical Agglomerative Clustering, Comparison of States, Hierarchical Clustering of States, The k-Means Algorithm k-Nearest Neighbor Prediction Functions: Introduction, Notation and Terminology, Distance Metrics, The k-Nearest Neighbor Prediction Function, Exponentially Weighted k-Nearest Neighbors, Digit Recognition, Accuracy Assessment, k-Nearest Neighbor Regression, Forecasting the S&P 500, Forecasting by Pattern Recognition, Cross-Validation The Multinomial Naïve Bayes Prediction Function: Introduction, The Federalist Papers, The Multinomial Naïve Bayes Prediction Function, Reducing the Federalist Papers, Predicting Authorship of the Disputed Federalist Papers, Customer Segmentation	12
V	Forecasting: Introduction, Working with Time, Analytical Methods, Computing $\rho\tau$, Drift and Forecasting, Holt-Winters Exponential Forecasting, Regression-Based Forecasting of Stock Prices, Time-Varying Regression Estimators Real-time Analytics: Introduction, Forecasting with a NASDAQ Quotation Stream, Forecasting the Apple Inc. Stream, The Twitter Streaming API, Sentiment Analysis, Sentiment Analysis of Hashtag Groups	12

Books and References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Algorithms for Data Science	Brian Steele, John Chandler,Swarna	Springer	1st	2016

		Reddy			
2	Data Science Algorithms in a Week	David Natingga	Packt Publishing	1st	2017
3	Data Science: Theories, models, Algorithms and Analytics	SanjivRanjan Das	S.R. Das	1st	2017

Algorithms for Data Science Practical

M. Sc (Data Science)		Semester II	
Course Name: Algorithms for Data Science Practical		Course Code: KPSDS222P3	
Periods per week (1 Period is 60 minutes)		4	
Credits		2	
		Hours	Marks
Evaluation System	Practical Examination	2	50
	Internal	--	

Practical:

Perform minimum ten practical based on the basic concepts of each algorithm for data science covering the entire syllabus.

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

- Understand fundamentals of data science
- Apply data visualisation in big-data analytics
- Apply Hadoop and map-reduce algorithm to big data
- Apply different algorithms to data sets
- Perform real-time analytics

Optimization Techniques

M. Sc (Data Science)		Semester – II	
Course Name: **Optimization Techniques**		Course Code: KPSDS22204	
Periods per week (1 Period is 60 minutes)		4	
Credits		4	
		Hours	Marks
Evaluation System	Theory Examination	2½	60
	Internal	--	40

Pre requisites :

- Knowledge of Algorithms and mathematical foundation

Course Objectives:

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

Unit	Details	Lectures
I	Mathematical Foundations: Functions and Continuity, Review of Calculus, Vectors, Matrix Algebra, Eigenvalues and Eigenvectors, Optimization and Optimality, General Formulation of Optimization Problems Algorithms, Complexity, and Convexity: What Is an Algorithm?, Order Notations, Convergence Rate, Computational Complexity, Convexity, Stochastic Nature in Algorithms	12
II	Optimization: Unconstrained Optimization, Gradient-Based Methods, Gradient-Free Nelder–MeadMethod Constrained Optimization: Mathematical Formulation, Lagrange Multipliers, Slack Variables, Generalized Reduced GradientMethod, KKT Conditions, PenaltyMethod Optimization Techniques: Approximation Methods: BFGS Method, Trust-Region Method, Sequential Quadratic Programming, Convex Optimization, Equality Constrained Optimization, Barrier Functions, Interior-PointMethods, Stochastic and Robust Optimization	12

III	Linear Programming: Introduction, SimplexMethod, Worked Example by Simplex Method, Interior-PointMethod for LP Integer Programming: Integer Linear Programming, LP Relaxation, Branch and Bound, Mixed Integer Programming, Applications of LP, IP, and MIP Regression and Regularization: SampleMean and Variance, Regression Analysis, Nonlinear Least Squares, Over-fitting and Information Criteria, Regularization and Lasso Method, Logistic Regression, Principal Component Analysis	12
IV	Machine Learning Algorithms: Data Mining, Data Mining for Big Data, Artificial Neural Networks, Support Vector Machines, Deep Learning Queueing Theory and Simulation: Introduction, Arrival Model, Service Model, Basic QueueingModel, Little's Law, Queue Management and Optimization Multiobjective Optimization: Introduction, Pareto Front and Pareto Optimality, Choice and Challenges, Transformation to Single Objective Optimization, The ϵConstraint Method, Evolutionary Approaches	12
V	Constraint-Handling Techniques: Introduction and Overview, Method of Lagrange Multipliers, Barrier Function Method, PenaltyMethod, Equality Constraints via Tolerance, Feasibility Criteria, Stochastic Ranking, Multiobjective Constraint-Handling and Ranking Evolutionary Algorithms: Evolutionary Computation, Evolutionary Strategy, Genetic Algorithms, Simulated Annealing, Differential Evolution Nature-Inspired Algorithms: Introduction to SI, Ant and Bee Algorithms, Particle Swarm Optimization, Firefly Algorithm, Cuckoo Search, Bat Algorithm, Flower Pollination Algorithm, Other Algorithms	12

Books and References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Optimization Techniques and Applications with Examples	Xin-She Yang	Wiley	3rd	2018
2	Optimization	A.K. Malik,	I.K.	1st	2012

	Techniques	S.K. Yadav, S.R. Yadav	International Publishing House		
3	Optimization methods: from theory to design	Marco Cavazzuti	Springer	1st	2012
4	Optimization Techniques	Chander Mohan, Kusum Deep	New Age International	1st	2009

Optimization Techniques Practical

M. Sc (Data Science)		Semester II	
Course Name: Optimization Techniques Practical		Course Code: KPSDS222P4	
Periods per week (1 Period is 60 minutes)		4	
Credits		2	
		Hours	Marks
Evaluation System	Practical Examination	2	50
	Internal	--	

Practical:

Perform minimum ten practical based on the basic concepts of each optimization technique covering the entire syllabus.

Course Outcomes:
Learner will be able to

- Apply operations research techniques like linear programming problem in industrialoptimization problems.
- Solve allocation problems using various OR methods.
- Understand the characteristics of different types of decisionmaking environment and theappropriate decision making approaches and tools to be used in each type.
- Recognize competitive forces in the marketplace and develop appropriate reactions based onexisting constraints and resources.

Evaluation Scheme for First Year (PG) under AUTONOMY

I. Internal Evaluation for Theory Courses – 40 Marks

Continuous Internal Assessment 1 (Seminar Presentations) – 40 Marks

II. External Examination for Theory Courses – 60 Marks

Duration: 2 Hours

Theory question paper pattern:

All questions are compulsory.

Question	**Based on**	**Options**	**Marks**
Q.1	Unit I	*Any 2 out of 4*	12
Q.2	Unit II	*Any 2 out of 4*	12
Q.3	Unit III	*Any 2 out of 4*	12
Q.4	Unit IV	*Any 2 out of 4*	12
Q.5	Unit I, II, III, IV	*Any 2 out of 4*	12

- All questions shall be compulsory with internal choice within the questions.
- Each Question may be sub-divided into sub questions as a, b, c, d, etc. & the allocation of Marks depends on the weightage of the topic.

III. Practical Examination

- Each core subject carries 50 Marks (30 marks External + 20 marks Internal)

Sr. No.	**Postgraduate Practical Internal Evaluation:**	**Marks**
1	Short Experiment/Field Trip/Excursion/Industrial Visit Report	15
2	Journal	5

Sr. No.	**Postgraduate Practical External Evaluation:**	**Marks**
1	Experiment/s	25

2	Viva	5

- Duration: 2 Hours for each practical course.
- Minimum 80% practical from each core subjects are required to be completed.
- Certified Journal is compulsory for appearing at the time of Practical Exam