Hindi Vidya Prachar Samiti's

Ramniranjan Jhunjhunwala College of Arts, Science & Commerce (Autonomous), Ghatkopar (W)

Affiliated to

University of Mumbai

Syllabus Framework as per LOCF

Program: M.Sc. Data Science & Artificial Intelligence

Program Code: RJSPGDSAI

Choice Based Credit System Syllabus

(With effect from the academic year 2020-21)

Course Structure

Semester I

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGDSAI101	Fundamentals of Data Science	CC	4	-	4
RJSPGDSAI102	Data Warehousing	CC	4	-	4
RJSPGDSAI103	Artificial Intelligence	CC	4	-	4
RJSPGDSAI1L1	PG Lab – I	PGL	-	2	2
RJSPGDSAI1L2	PG Lab – II	PGL	-	2	2
RJSPGDSAI1R1	Mini Project – I	MNP	-	2	2
RJSPGDSAI1S1	Seminar – I	SE	-	2	2
RJSPGDSAI1P1	Professional Elective – I	PE	3	-	3
RJSPGDSAI1C1	Career Advancement Course	CAC	1	-	1
	Total		**16**	**8**	**24**

Semester II

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGDSAI201	Machine Learning – I	CC	4	-	4
RJSPGDSAI202	Big Data Technology	CC	4	-	4
RJSPGDSAI203	Soft Computing	CC	4	-	4
RJSPGDSAI2L3	PG Lab – III	PGL	-	2	2
RJSPGDSAI2L4	PG Lab – IV	PGL	-	2	2
RJSPGDSAI2R2	Mini Project – II	MNP	-	2	2
RJSPGDSAI2S2	Seminar – II	SE	-	2	2
RJSPGDSAI2P2	Professional Elective – II	PE	3	-	3

RJSPGDSAI2C2	Career Advancement Course	CAC	1	-	1
	Total		**16**	**8**	**24**

Semester I

Course Code	**Course Name**	**Group**	**Teaching Scheme (Hrs/Week)**		**Credits: 4**
			Lectures	**Practical**	**Max. Lectures Required: 40**
RJSPGDSAI101	**Fundamentals of Data Science**	**CC**	**4**	**-**	

Course Objectives

1. To provide basic knowledge of data science.
2. To provide the foundation on topics in statistical methods and applied probability that forms the basis for data science.
3. To provide the foundation on topics of mathematics that forms the basis for data science.
4. To address the issues and the principals of estimation theory, testing hypothesis and regression and prediction.

Learning Outcomes:

Upon completion of this course, the student should be able to

1. Demonstrate the competency on topics like basics of data science, data transformation, statistical methods, applied probability etc.
2. Apply the various distribution methods to data.
3. Use statistical tests in testing hypothesis on data.
4. Demonstrate the competency on topics like unbiasedness of estimators, methods of Maximum Likelihood Estimation and Central Limit Theorem.
5. Perform exploratory analysis of multivariate data.

Unit	**Topics**	**Lectures**
Unit I	**Basics of Data Science** Decision Theory, Estimation Theory, Coordinate Systems, Matrices and Linear Algebra, Linear Transformations	10

	Data Collection, Modelling and Compilation, Data Analysis, Data Presentation and Visualization Data Science Software Tools, Programming Languages for Data Science, Applications of Data Science	
Unit II	**Data and Sampling Distributions** Random sampling and sample bias: Bias, Random selection, Selection Bias: Regression to mean, Sampling distributions of a statistic: Central limit theorem, Standard error, Bootstrap, Resampling, Confidence Intervals. **Distributions** Normal distribution: Standard normal and QQ plots, Long-tailed distributions, Student's t-distribution, Binomial distribution, Poisson distribution, Exponential distribution and Weibull distributions. **Significance Testing** A/B Testing, Hypothesis test: Null hypothesis, Alternative hypothesis, One-way and Two-way hypothesis test, Resampling.	10
Unit III	**Basic Probability and Terms** Events and their Probabilities, Rules of Probability, Conditional probability and independence, Permutations and combinations, Bayer's Theorem, Descriptive Statistics, Compound probability, Conditional probability. **Data Transformations and quality analysis** Merge, Rollup, Transpose and Append, Missing Analysis and Treatment, Outlier analysis and treatment.	10
Unit IV	**Hypothesis testing** Null hypothesis, Alternative hypothesis, One-way and Two-way hypothesis test, Permutation test, Exhaustive and bootstrap permutation test, P-values, t-Test, Multiple testing, Degree of Freedom, ANOVA: F-statistics and two-way ANOVA, Chi-square test, Fisher's exact test, Power and sample size. **Regression and Prediction**	10

	Linear regression, Multiple linear regression, Cross-validation, Model selection and stepwise selection, Weighted regression, Factor variables in regression, Interpreting the regression equation, Regression diagnostic, Polynomial and spline regression.	

References

1. "Fundamentals of Data Science: Take the First Step to Become A Data Scientist", Samuel Burns, Amazon KDP Printing and Publishing.
2. "Practical Statistics for Data Science", Peter Bruce, Andrew Bruce, O'Reilly, 2017.
3. "Statistics for Data Science", James D. Miller, Packt, 2017.
4. "Probability and Statistics for Engineers", Dr. J. Ravichandran,2010.
5. "R for data Science: Import, Tidy, Transform, Visualize and Model Data", Hadley Wickham, Garrett Grolemund.
6. "Data Analysis with R", Tony Fischetti, 2015.
7. "Mastering Data Analysis with R", Gergely Daroczi, 2015.
8. "R Cookbook", Paul Teetor, O'Reilly, 2017.
9. "Practical Data Science Cookbook", Prabhanjan Tatter, Tony Ojeda, Sean Patrik Murphy, Benjamin Bengfort, Abhijit Dasgupta, 2nd Edition, Packt, 2014

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits: 4
			Lectures	Practical	Max. Lectures Required: 40
RJSPGDSAI102	Data Warehousing	CC	4	-	

Course Objectives

1. To learn the concepts of data warehouse and business intelligence.
2. To provide in-depth knowledge of dimension modelling.
3. To learn how to build and use data warehouse for various applications like Retail Sales, Order Management, Inventory, and Customer Relationship Management.

Learning Outcomes

Upon completion of this course, the student should be able to

1. Create the dimension model for any application.
2. Perform ETL process on source data and send it to data warehouse database.
3. Analyse the data for various applications.

Unit	Topics	Lectures
Unit I	**Data Warehousing, Business Intelligence, and Dimensional Modelling Primer** Different Worlds of Data Capture and Data Analysis, Goals of Data Warehousing and Business Intelligence, Dimensional Modelling Introduction, Kimball's DW/BI Architecture, Alternative DW/BI Architectures, Dimensional Modelling Myths.	10
Unit II	**Kimball Dimensional Modelling Techniques Overview** Fundamental Concepts, Basic Fact Table Techniques, Basic Dimension Table Techniques, Integration via Conformed Dimensions, Dealing with Slowly Changing Dimension Attributes, Dealing with Dimension Hierarchies.	10
Unit III	**Retail Sales** Four-Step Dimensional Design Process, Retail Case Study, Dimension Table Details, Retail Schema in Action, Retail Schema Extensibility, Fact less Fact Tables, Dimension and Fact Table Keys, Resisting Normalization Urges. **Order Management** Order Management Bus Matrix, Order Transactions, Invoice Transactions, Accumulating Snapshot for Order Fulfilment Pipeline.	10
Unit IV	**Inventory** Value Chain Introduction, Inventory Models, Fact Table Types, Value Chain Integration, Enterprise Data Warehouse Bus Architecture, Conformed Dimensions. **Customer Relationship Management** Overview, Customer Dimension Attributes, Bridge Tables for Multivalued Dimensions, Complex Customer Behavior, Customer	10

	Data Integration Approaches, Low Latency Reality Check.	

References

1. "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling", Ralph Kimball Margy Ross, Wiley.
2. "The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data", Ralph Kimball, Joe Caserta.
3. "Building the Data Warehouse", Fourth Edition, W. H. Inmon, Wiley.

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits: 4
			Lectures	Practical	Max. Lectures Required: 40
RJSPGDSAI103	Artificial Intelligence	CC	4	-	

Course Objectives

1. To gain a historical perspective of AI and its foundations.
2. To become familiar with basic principles of AI toward problem solving, inference, perception, knowledge representation, and learning.
3. To investigate applications of AI techniques in intelligent agents, expert systems, artificial neural networks and other machine learning models.
4. Experience AI development tools such as an 'AI language', expert system shell, and/or data mining tool.

Learning Outcomes

Upon completion of this course, the student should be able to

1. Demonstrate fundamental understanding of the history of artificial intelligence (AI) and its foundations.
2. Apply basic principles of AI in solutions that require problem solving, inference, perception, knowledge representation, and learning.
3. Demonstrate awareness and a fundamental understanding of various applications of AI techniques in intelligent agents, expert systems, artificial neural networks and other machine learning models.
4. Demonstrate proficiency developing applications in an 'AI language'.

Unit	Topics	Lectures
Unit I	**Introduction and Problem Solving** **Introduction** What is AI? Foundation of AI, History of AI **Intelligent Agents:** Agents and Environment, concept of Rationality, Nature of Environments, Structure of Agents. **Problem Solving** Problem Solving Agents, Example Problems, searching for solutions, Uninformed search strategies – (Breadth First, Uniform cost, Depth First, Depth Limited, Iterative deepening depth first, bidirectional), informed search strategies – (Greedy best first, A*, Optimality of A*, Memory bounded), Heuristic Functions. **Beyond Classical Search** Local search algorithms and optimization problems, local search in continuous spaces, searching with non-deterministic actions, searching with partial observations, online search agents and unknown environments. **Adversarial Search** Games, Optimal decision in games, Alpha--Beta Pruning, Imperfect Real-Time Decisions, Stochastic Games, Partially Observable Games, Defining Constraint Satisfaction Problems, Constraint Propagation: Inference in CSPs, Backtracking Search for CSPs.	10
Unit II	**Knowledge, Reasoning and Planning** **Logical Agents** Knowledge-Based Agents, Propositional Logic, Propositional Theorem Proving, Horn clauses and definite clauses, Forward and backward chaining, Effective Propositional Model Checking **First Order Logic** Syntax and Semantics of First-Order Logic, using First order logic, Knowledge Engineering in First-Order Logic	10

	Inference in First Order Logic Unification and Lifting, Forward Chaining, Backward chaining, resolution **Classical Planning and Acting** Definition, Algorithms for Planning as State-Space Search, planning graphs, analysis of planning approaches.	
Unit III	**Uncertain Planning and Reasoning** **Quantifying Uncertainty** Acting under Uncertainty, Basic Probability Notation, Inference Using Full Joint Distributions, Independence, Bayes' Rule and Its Use. **Probabilistic Reasoning** Representing Knowledge in an Uncertain Domain, The Semantics of Bayesian Networks, Efficient Representation of Conditional Distributions, Relational and First-Order Probability Models. **Making Simple Decisions** The Basis of Utility Theory, Utility Functions, Multiattribute Utility Functions, Decision Networks, The Value of Information. **Making Complex Decisions** Sequential Decision Problems, Value Iteration, Policy Iteration, Partially Observable MDPs, Decisions with Multiple Agents: Game Theory, Mechanism Design.	10
Unit IV	**Learning in Artificial Intelligence** Forms of Examples, Supervised Learning, Learning Decision Trees, Evaluating and Choosing the Best Hypothesis, The Theory of Learning, Regression and Classification with Linear Models, Artificial Neural Networks, A Logical formulation of Learning, Knowledge in Learning, Learning using Relevance Information, Inductive Logic Programming, Reinforcement Learning, Passive Reinforcement Learning, Active Reinforcement Learning, Generalization in Reinforcement Learning, Policy Search, Applications of Reinforcement Learning	10

	Applications of Artificial Intelligence Natural Language Processing, Expert Systems, Neural Networks.	

References

1. "Artificial Intelligence: A Modern Approach", 3rd Edition, Stuart Russell and Peter Norvig, Hawkins, J. and Blakeslee, S. On Intelligence. Times Books, 2004.
2. "Artificial Intelligence theory and practice", Dean, T., Allen, J. and Aloimonos, Y., New York: Benjamin Cummings,1995.
3. "Essentials of Artificial Intelligence ", Ginsberg, M., Palo Alto, CA: Morgan Kaufmann, 1993.
4. "The Description Logic Handbook: Theory, Implementation and Applications ", Baader, F., Calvanese, D., McGuinness, D., Nardi, D., & Patel-Schneider, P., Cambridge University Press, 2003.
5. "Knowledge Representation", Brachman, R. J. & Levesque, H. J., New York: Elsevier,2004.
6. "Expert Systems and Probabilistic Network Models", Castillo, E., Gutierrez, J. M., Hadi, A. S., Berlin: Springer, 1996.
7. "Neural Networks for Pattern Recognition", Bishop, C. M., New York: Oxford University Press, 1995.

PG Labs

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits: 2
			Lectures	Practical	Max. Lectures Required: 20
RJSPGDSAI1L1	PG Lab – I Fundamentals of Data Science	PGL	-	2	

Practical List:

1. Data Collection, Modelling and Compilation.
2. Data Visualization.
3. Exploratory data analysis.
4. Exploring Binary and categorical data.
5. Data and sampling distributions.
6. Significance testing.
7. Data transformations and quality analysis.
8. Hypothesis testing.
9. Regression and prediction.
10. Logistic Regression

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits: 2
			Lectures	Practical	Max. Lectures Required: 20
RJSPGDSAI1L2	PG Lab – II Data Warehousing	PGL	-	2	

Practical List:

1. Creating the database using various constraints.
2. Using DDL, DML, DCL and TCL statements.
3. Introduction to ER model and Relational Model.

4. Creating Dimension Model for a Datawarehouse.
5. Loading data into the dimension and fact tables.
6. Validating data while loading into a warehouse.
7. ETL - Staging process.
8. Creation of Cube.
9. Using data analysis services for data mining.
10. Creating Reports and charts.

Professional Electives

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits: 4
			Lectures	Practical	Max.
RJSPGDSAI1P1A	Professional Elective – I Advanced Data Structures	PE	3	-	Lectures Required: 40

Course Objectives

1. To enhance the students' knowledge of algorithms, data structures, algorithmic analysis and algorithm design techniques.
2. To learn a variety of useful algorithms, techniques.
3. To derive from them in order to apply those algorithms and techniques to solve problems.

Learning Outcomes

Upon completion of this course, the student should be able to

1. Have a basic ability to analyse algorithms and to determine algorithm time efficiency and correctness.
2. Use a variety of advanced data structures and their implementations.
3. Apply the learnt algorithm design techniques and data structures to solve problems.

Unit	Topics	Lectures
Unit I	**Fundamentals** Mathematical Proof Techniques: Induction, proof by contradiction, direct proofs Asymptotic Notations, Properties of Big-oh Notation, Conditional Asymptotic Notation, Algorithm Analysis, Amortized Analysis, Introduction to NP-Completeness, NP-Hard, Recurrence Equations, Solving Recurrence Equations, Time-Space Trade-off.	10
Unit II	**HEAP STRUCTURES** Min heap, Max heaps, Deaps, Leftist Heaps, Binomial Heaps, Fibonacci Heaps, Skew Heaps, Lazy-Binomial Heaps.	10

	SEARCH STRUCTURES Binary Search Trees, AVL Trees, Red-Black trees, Multi-way Search Trees, B- Trees, Splay Trees, Tries.	
Unit III	**GEOMETRIC ALGORITHMS** Segment Trees, 1-Dimensional Range Searching, k-d Trees, Line Segment Intersection, Convex Hulls, Computing the Overlay of Two Subdivisions, Range Trees, Voronoi Diagram.	10
Unit IV	**PARELLEL ALGORITHMS** Flynn's Classifications, List Ranking, Prefix computation, Array Max, Sorting on EREW PRAM, Sorting on Mesh and Butterfly-Prefix sum on Mesh and Butterfly, Sum on mesh and butterfly, Matrix Multiplication, Data Distribution on EREW, Mesh and Butterfly.	10

References

1. "Fundamentals of Data Structures in C", Ellis Horowitz, Sartaj Sahni, Dinesh Mehta, Silicon Pr, 2007.
2. "Computational Geometry Algorithms and Applications", Mark de Berg, Otfried Cheong, Marc van Kreveld, Mark Overmars, Third Edition, 2008.
3. "Introduction to Algorithms", Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein, MIT Press, 2009.

Course Code	**Course Name**	**Group**	**Teaching Scheme (Hrs/Week)**		**Credits: 4**
			Lectures	**Practical**	**Max. Lectures Required: 40**
RJSPGDSAI1P1B	**Professional Elective – I Image Processing**	**PE**	**3**	**-**	

Course Objectives

1. To study the image fundamentals and mathematical transforms necessary for image processing.
2. To study the image enhancement techniques.
3. To study image restoration procedures.
4. To study the image compression procedures.

Learning Outcomes

Upon completion of this course, the student should be able to

1. Analyse images in the frequency domain using various transforms.
2. Evaluate the techniques for image enhancement and image restoration.
3. Categorize various compression techniques.
4. Interpret Image compression standards.
5. Interpret image segmentation and representation techniques.

Unit	Topics	Lectures
Unit I	**Introduction to Image Processing** Example of fields that uses image processing, Steps of image processing, Components, Applications, Image sensors and Image formats. **Visual Preliminaries** Brightness adaptation and contrast, Acuity and contour, Texture and pattern discrimination, Shape detection and recognition, perception of color, Computational model of perceptual processing, Image sampling and quantization, Basic relationship between pixels.	10
Unit II	**Intensity transformations** Introduction, Basic intensity transformation functions, Histogram equalization, Local histogram processing, and Using histogram statistics for image enhancement. **Spatial filtering** Fundamentals of spatial filtering, Smoothing and sharpening spatial filters, combining spatial enhancement methods, Using fuzzy techniques for intensity transformations and spatial filtering.	10
Unit III	**Colour Image Processing** Colour fundamentals, Colour models, Pseudo color image processing, Basic of full, color image processing, colour transformations, Smoothing and sharpening, Image segmentation bases on colour, Noise in colour images, Colour image compression.	10

	Image compression Fundamentals, Basic methods, Digital image watermarking, Full motion video compression.	
Unit IV	**Morphological Image Processing** Introduction, Erosion and Dilation, Opening and closing, History, Miss transformation, Basic morphological algorithms, Gray scale morphology. **Segmentation** Fundamentals, Point, Line and Edge detection, Thresholding, Region based segmentation, Segmentation using morphological watersheds, Use of motion in segmentation – Spatial techniques.	10

References

1. "Digital Image Processing", Gonzalez and Woods, 3rd Edition, Pearson Education.
2. "Digital Image Processing and Analysis", Bhabatosh Chanda, Dwijesh Dutta Majumder, 2nd Edition, PHI.
3. "Fundamentals of Digital Image Processing", Anil K Jain, 1st Edition, PHI.

Career Advancement Course

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits: 1 Max. Lectures Required: 40
			Lectures	Practical	
RJSPGDSAI1C1	Career Advancement Course – I Research in Computing	CAC	2	-	

Course Objectives

1. To be able to conduct business research with an understanding of all the latest theories.
2. To develop the ability to explore research techniques used for solving any real world or innovate problem.

Learning Outcomes

Upon completion of this course, the student should be able to

1. Use various stages of research process.
2. Apply research methods.
3. Use various methods for data collection.
4. Apply the methods of measurement and sampling.

Unit	Topics	Lectures
Unit I	**Introduction:** Role of Business Research, Information Systems and Knowledge Management, Theory Building, Organization ethics and Issues	10
Unit II	**Beginning Stages of Research Process:** Problem definition, Qualitative research tools, Secondary data research	10
Unit III	**Research Methods and Data Collection:** Survey research, communicating with respondents, Observation methods, Experimental research	10

Unit IV	**Measurement Concepts, Sampling and Field work:** Levels of Scale measurement, attitude measurement, questionnaire design, sampling designs and procedures, determination of sample size.	10

References

1. "Business Research Methods", William G. Zikmund, B J Babin, J.C. Carr, Atanu Adhikari, M. griffin, Cengage,8e, 2016.
2. "Business Analytics", Albright Winston, Cengage 5e, 2015.

Semester II

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits: 4 Max. Lectures Required: 40
			Lectures	Practical	
RJSPGDSAI201	Machine Learning – I	CC	4	-	

Course Objectives

1. To introduce various statistical and machine learning concepts and methods.
2. To introduce machine learning solutions to regression, classification and clustering problems.
3. To evaluate and interpret the results of algorithm.

Learning Outcomes

Upon completion of this course, the student should be able to

1. Perform end-to-end process of investigating data through a machine learning lens.
2. Extract and identify best features of data.
3. Evaluate the performance of machine learning algorithms.

Unit	Topics	Lectures
Unit I	**Data pre-processing** Vectors, Matrices and Arrays, Loading data, Data handling Handling numerical data and categorical data, Handling text, dates and time, Handling images. **Statistical Learning** What is statistical learning, assessing model accuracy.	10
Unit II	**Linear Regression** Linear Regression, Multiple Linear Regression, Other Considerations in Regression Model, The Marketing Plan, Comparison of Linear Regression with K-Nearest.	10
Unit III	**Classification**	10

	An overview of classification, why not linear regression, logistic regression, linear discriminant analysis, a comparison of classification methods.	
Unit IV	**Unsupervised Learning** The challenge of unsupervised learning, principal components analysis, clustering methods (density-based methods, hierarchical-based methods, portioning-based methods, grid-based methods), clustering algorithms (k-means, k-nearest neighbours).	10

References

1. "An Introduction to Statistical Learning with Application in R", By Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer Texts in Statistics.
2. "Machine Learning", Mitchell Tom, McGraw Hill,1997.
3. "Pattern classification", 2nd edition, Richard O. Duda, Peter E. Hart, David G. Stork. Wiley, New York, 2001.
4. "Machine Learning: A Probabilistic Perspective", Kevin P. Murphy, MIT Press, 2012
5. "Practical Data Science", Andreas Francois Vermeulen, APress, 2018
6. "Principles of Data Science", Sinan Ozdemir, Packt, 2016.

Course Code	**Course Name**	**Group**	**Teaching Scheme (Hrs/Week)**		**Credits: 4 Max. Lectures Required: 40**
			Lectures	**Practical**	
RJSPGDSAI202	**Big Data Technology**	**CC**	**4**	**-**	

Course Objectives

1. To provide knowledge of basic and advanced methods of big data technology and tools.
2. To provide the knowledge of MapReduce, Hadoop and its ecosystem.
3. To provide hands-on training that enable effective participation in big data projects.

Learning Outcomes

Upon completion of this course, the student should be able to

1. Apply Hadoop ecosystem components.

2. Build and maintain reliable, scalable and distributed systems with Apache Hadoop.
3. Apply big data concepts to various use cases.
4. Develop application using Zookeeper and Monitoring the cluster.

Unit	Topics	Lectures
Unit I	**INTRODUCTION TO BIG DATA** Introduction: Distributed file system, Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce. **INTRODUCTION HADOOP** Big Data: Apache Hadoop & Hadoop Ecosystem, Moving Data in and out of Hadoop, Understanding inputs and outputs of MapReduce, Data Serialization.	10
Unit II	**HADOOP ARCHITECTURE** Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands, Anatomy of File Write and Read., Name Node, Secondary Name Node, and Data Node, Hadoop MapReduce paradigm, Map and Reduce tasks, Job, Task trackers, Cluster Setup, SSH & Hadoop Configuration, HDFS Administering, Monitoring & Maintenance.	10
Unit III	**HADOOP ECOSYSTEM AND YARN** Hadoop ecosystem components: Schedulers, Fair and Capacity, Hadoop 2.0 New Features Name Node High Availability, HDFS Federation, MRv2, YARN, Running MRv1 in YARN.	10
Unit IV	**Hive and HiveQL, HBase** Hive Architecture and Installation, Comparison with Traditional Database. **HiveQL** Querying Data, Sorting and Aggregating, Map Reduce Scripts, Joins & Subqueries. **HBase concepts**	10

	Advanced Usage, Schema Design, Advance Indexing, PIG, Zookeeper, how it helps in monitoring a cluster, HBase uses Zookeeper and how to Build Applications with Zookeeper.	

References

1. "Professional Hadoop Solutions", Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, Wiley, ISBN: 9788126551071, 2015.
2. "Understanding Big data", Chris Eaton, Dirk deroos et al, McGraw Hill, 2012.
3. "HADOOP: The definitive Guide", Tom White, O Reilly 2012. 6 IT2015 SRM(E&T).
4. "Big Data Analytics with R and Hadoop", Vignesh Prajapati, Packet Publishing 2013.
5. "Oracle Big Data Handbook", Tom Plunkett, Brian Macdonald et al, Oracle Press, 2014.
6. "Big Data and Business analytics", Jy Liebowitz, CRC press, 2013.
7. http://www.bigdatauniversity.com/

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits: 4 Max. Lectures Required: 40
			Lectures	Practical	
RJSPGDSAI203	Soft Computing	CC	4	-	

Course Objectives

1. To provide the knowledge of soft computing concepts like fuzzy logic, neural networks and genetic algorithm, where Artificial Intelligence is mother branch of all.
2. To learn effective techniques and their roles in building intelligent systems.
3. To learn how to use neural networks for classification and regression problems.

Learning Outcomes

Upon completion of this course, the student should be able to

1. Identify and describe soft computing techniques and their roles in building intelligent machines.
2. Select soft computing methodology to solve a particular problem.
3. Apply fuzzy logic and reasoning to solve engineering problems.
4. Apply genetic algorithms to combinatorial optimization problems.

5. Apply neural networks for classification and regression problems.
6. Evaluate and compare solutions by various soft computing approaches for a given problem.

Unit	Topics	Lectures
Unit I	**Introduction** Introduction of soft computing, soft computing vs. hard computing, various types of soft computing techniques, Fuzzy Computing, Neural Computing, Genetic Algorithms, Associative Memory, Adaptive Resonance Theory, Classification, Clustering, Bayesian Networks, Probabilistic reasoning, applications of soft computing.	10
Unit II	**Artificial Neural Network** Fundamental concept, Evolution of Neural Networks, Basic Models, McCulloh-Pitts Neuron, Linear Separability, Hebb Network. **Supervised Learning Network** Perceptron Networks, Adaptive Linear Neuron, Multiple Adaptive Linear Neurons, Backpropagation Network, Radial Basis Function, Time Delay Network, Functional Link Networks, Tree Neural Network. **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 memory networks.	10
Unit III	**Unsupervised Learning Networks** Fixed weight competitive nets, Kohonen self-organizing feature maps, learning vectors quantization, counter propagation networks, adaptive resonance theory networks. **Special Networks** Simulated annealing, Boltzmann machine, Gaussian Machine, Cauchy Machine, Probabilistic neural net, cascade correlation	10

	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.	
Unit IV	**Introduction to Fuzzy Logic, Classical Sets and Fuzzy sets** 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. **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.	10

References

1. "Artificial Intelligence and Soft Computing", Anandita Battacharya Das, SPD 3rd, 2018.
2. "Principles of Soft computing", S.N.Sivanandan, 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, 2004.
4. "Neural Networks, Fuzzy Logic and Genetic Algorithms: Synthesis & Applications", S.Rajasekaran, G. A. Vijayalakshami , Prentice Hall of India. 2004.
5. "Fuzzy Logic with Engineering Applications", Timothy J.Ross, McGraw-Hill, 1997.

PG Labs

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits: 2 Max. Lectures Required: 20
RJSPGDSAI2L3	PG Lab – III Machine Learning - I	PGL	-	2	

Practical List:

1. Simple and linear regression.
2. Linear regression
3. Multiple linear regression
4. Logistics regression
5. Linear discriminant analysis
6. Quadratic discriminant analysis
7. Clustering using K-means
8. K-nearest neighbors
9. Program component analysis
10. Hierarchical methods

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits: 2 Max. Lectures Required: 20
RJSPGDSAI2L4	PG Lab – IV Big Data Technology	PGL	-	2	

Practical List:

1. Setting Single node Hadoop cluster using Ubuntu and HDFS.
2. Configuration of Multiple node Hadoop cluster.
3. Fie management in HDFS.
4. Creating application using MapReduce.
5. Word Count application using Hadoop Eclipse.
6. Handling unstructured data using NoSQL.

7. Querying, Sorting and Aggregating data using HiveQL.
8. Map Reduce Scripts, Joins & Subqueries using HiveQL.
9. Schema design using HBase.
10. Using Mahout Library for big data analysis.
11. Building application with Zookeeper.

Professional Electives

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits: 4 Max. Lectures Required: 40
			Lectures	Practical	
RJSPGDSAI2P2A	Professional Elective – II Data Science in Cloud Computing	PE	3	-	

Course Objectives

1. To study the fundamental aspects of cloud environment, deployment models and different services offered by cloud.
2. To study various techniques of virtualization.
3. To Study security issues in cloud computing.

Learning Outcomes

Upon completion of this course, the student should be able to

1. Design of computer clusters for scalable parallel computing.
2. Understand virtualization of clusters and Data centers along with various cloud computing and Service models-PaaS, SaaS, IaaS.
3. Apply various aspects of security to cloud clusters.

Unit	Topics	Lectures
Unit I	**Cloud Computing Basics** Cloud Enabling Technologies, Characteristics of Cloud Computing, Benefits of Cloud Computing, Cloud Service Models, Cloud Deployment models, Cloud computing Infrastructure, Cloud Challenges.	10
Unit II	**Virtualization Fundamentals** Virtualization-Enabling technology for cloud computing, Types of Virtualization, Server Virtualization, Desktop Virtualization, Memory Virtualization, Application and Storage Virtualization, Tools and	10

	Products available for Virtualization.	
Unit III	**SaaS, PaaS, Iaas And Cloud Storage** Getting started with SaaS, Understanding the multitenant nature of SaaS solutions, Understanding Open SaaS Solutions, Understanding Service Oriented Architecture, PaaS, Benefits and Limitations of PaaS. Understanding IaaS, improving performance through Load balancing, Server Types within IaaS solutions, utilizing cloud based NAS devices. Understanding Cloud based data storage, Cloud based backup devices, Cloud based database solutions, Cloud based block storage.	10
Unit IV	**Cloud Application Development and Security Management & Privacy in Cloud** Client Server Distributed Architecture for cloud – Traditional apps vs. Cloud Apps-Client-side programming model: Web clients. Mobile clients, Server-Side **Programming Technologies** AJAX, JSON, Web Services (RPC, REST), MVC Design Patterns for Cloud Application Development. **Security Management in the Cloud** Security Management Standards, Security Management in the Cloud, Availability Management, SaaS Availability Management, PaaS Availability Management, IaaS Availability Management, Access Control, Security Vulnerability, Patch, and Configuration Management. **Privacy in Cloud** Privacy, Data Life Cycle, Key Privacy Concems in the Cloud. Protecting Privacy, Privacy Risk Management and Compliance in Relatior to Cloud Computing.	10

References

1. "Cloud Computing: A Practical Approach", Anthony T. Velte, Toby J. Velte, Robert Elsenpeter, Tata McGraw Hill Edition, Fourth Reprint, 2010.
2. "Cloud Computing: SaaS, PaaS, IaaS, Virtualization, Business Models, Mobile, Security and more", Kris Jamsa, Jones & Bartlett Learning Company LLC, 2013.
3. "Cloud Security: A Comprehensive Guide to Secure Cloud Computing", Ronald L.

Krutz, Russell vines, Wiley Publishing Inc., 2010.

4. "Cloud Security and Privacy an Enterprise perspective on Risk and Compliance", Tim Mather, Subra Kumaraswamy, and Shahed Latif, O'Reilly.
5. "Security and privacy in Internet of Things Models Algorithms and Implementations", Fe Hu, CRC Press.
6. "Cloud Security", Ronald Krutz and Russell Dean Vines, Wiley, India.

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits: 4 Max. Lectures Required: 40
			Lectures	Practical	
RJSPGDSAI2P2B	Professional Elective – II Virtual and Augmented Reality	PE	3	-	

Course Objectives

1. To introduce the concepts of virtual reality, visual perception, visual rendering and virtual reality systems.
2. To learn the principles of Virtual and Augmented Reality, VR examples and applications.
3. Understand trends and technology future VR experiences.

Learning Outcomes

Upon completion of this course, the student should be able to

1. Apply the concepts of Virtual and Augmented Reality to various problems.
2. Understand trends and technology future VR experiences.
3. Evaluate Virtual Reality Systems.

Unit	Topics	Lectures
Unit I	**Introduction to Virtual Reality** What is Virtual Reality? Modern VR Experiences, History. **Introduction to Augmented Reality**	10

	Definition and scope, Brief history, Examples, Related Fields. **Birds Eye View** Hardware, Software, Human Physiology and Perception. **Geometry of Virtual Worlds** Geometric models, changing position and orientation, viewing and chaining transformations.	
Unit II	**Light and optics** Basic behaviour of light, lenses, optical aberrations, human eye, camera, displays. **Physiology of Human Vision** From Cornea to photoreceptors, From photoreceptors to visual Cortex, Eye Movements, Implications for VR. **Visual Perception** Perception of depth, Perception of motion, perception of color, combining sources of Information, Visual Rendering.	10
Unit III	**Visual Rendering** Ray Tracing and Shading Models, Rasterization, Correcting Optical Distortions, Improving Latency and Frame Rates, Immersive photos and videos. **Motion in Real and Virtual World** Velocities and accelerations, vestibular systems, mismatched motion and eviction. **Tracking and Interaction** Tracking 2D orientation, tracking 3D orientation, Tracking attached bodies, 3D scanning environments, Motor programs and remapping, Locomotion, Manipulation, Social Interaction, Additional Interaction mechanisms.	10
Unit IV	**Audio** Physics of sound, Physiology of human hearing, Auditory perception, Auditory rendering. **Evaluating VR Systems and Experiences** Perceptual Training, Recommendations for developers, Comfort and	10

	VR sickness, experiences on Human subjects.	

References

1. "Virtual Reality", Steven M LaValle, Cambridge university Press.
2. "Understanding Virtual Reality: Interface, Application and Design", William R. Sherman, Alan B. Craig.
3. "Augmented Reality Principles and Practice", Dieter Schmalstieg Tobias Hollered.

Career Advancement Course

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits: 1 Max. Lectures Required : 40
			Lectures	Practical	
RJSPGDSAI2C2	Career Advancement Course - II Data Visualization	CAC	2	-	

Course Objectives

1. To understand how accurately represent voluminous complex data set in web and from other data sources.
2. To understand the methodologies used to visualize large data sets.
3. To understand the process involved in data visualization and security aspects involved in data visualization.

Learning Outcomes

Upon completion of this course, the student should be able to

1. Design and use various methodologies present in data visualization
2. Discuss the process involved and security issues present in data visualization

Unit	Topics	Lectures
Unit I	**Introduction** Context of data visualization: Definition, Methodology, Visualization design objectives. Key Factors: Purpose, visualization function and tone, visualization design options: Data representation, Data Presentation, Seven stages of data visualization, widgets, data visualization tools	10
Unit II	**Visualizing Data Methods** Mapping, Time series, Connections and correlations, Scatterplot maps, Trees, Hierarchies and Recursion, Networks and Graphs, Info graphics	10

Unit III	**Visualizing Data Process** Acquiring data, Where to Find Data, Tools for Acquiring Data from the Internet, Locating Files for Use with Processing, Loading Text Data, Dealing with Files and Folders, Listing Files in a Folder, Asynchronous Image Downloads, Advanced Web Techniques, using a Database, Dealing with a Large Number of Files.	10
Unit IV	**Parsing data** Levels of Effort, Tools for Gathering Clues, Text Is Best, Text Mark-up Languages, Regular Expression, Grammars and BNF Notation, Compressed Data, Vectors and Geometry, Binary Data Formats, Advanced Detective Work.	10

References

1. "Interactive data visualization for the web", Scott Murray, O'Reilly Media, Inc., 2013.
2. "Visualizing Data", Ben Fry, O'Reilly Media, Inc., 2007.
3. "Security Data Visualization: Graphical Techniques for Network Analysis", Greg Conti, No Starch Press Inc, 2007.

Evaluation Scheme

Semester I

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits	Max. Marks	Min. Marks	End Sem Exam Marks	Internal Marks
			Lectures	Practical					
RJSPGDSAI101	Fundamentals of Data Science	CC	4	-	4	100	40	60	40
RJSPGDSAI102	Data Warehousing	CC	4	-	4	100	40	60	40
RJSPGDSAI103	Artificial Intelligence	CC	4	-	4	100	40	60	40
RJSPGDSAI1L1	PG Lab – I	PGL	-	2	2	50	20	50	-
RJSPGDSAI1L2	PG Lab – II	PGL	-	2	2	50	20	50	-
RJSPGDSAI1R1	Mini Project – I	MNP	-	2	2	50	20	50	-
RJSPGDSAI1S1	Seminar – I	SE	-	2	2	50	20	50	-
RJSPGDSAI1P1	Professional Elective – I	PE	3	-	3	100	40	60	40
RJSPGDSAI1I1	Career Advancement Course	CAC	2	-	1	50	20	50	-
	Total		**16**	**8**	**24**	**650**	**-**	**-**	**-**

Semester II

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits	Max Marks	Min. Marks	End Sem Exam Marks	Internal Marks
			Lectures	Practical					
RJSPGDSAI201	Machine Learning – I	CC	4	-	4	100	40	60	40
RJSPGDSAI202	Big Data Technology	CC	4	-	4	100	40	60	40
RJSPGDSAI203	Soft Computing	CC	4	-	4	100	40	60	40
RJSPGDSAI2L1	PG Lab – III	PGL	-	2	2	50	20	50	-
RJSPGDSAI2L2	PG Lab – IV	PGL	-	2	2	50	20	50	-
RJSPGDSAI2R2	Mini Project – II	MNP	-	2	2	50	20	50	-
RJSPGDSAI2S2	Seminar – II	SE	-	2	2	50	20	50	-
RJSPGDSAI2P2	Professional Elective – II	PE	3	-	3	100	40	60	40
RJSPGDSAI2I1	Career Advancement Course	CAC	2	-	1	50	20	50	-
	Total		**16**	**8**	**24**	**650**	**-**	**-**	**-**

Scheme of Examination

Internal Evaluation (40 Marks)

1. Internal Examination 40 marks various modes with different weightage (Presentation, seminar, MCQs, Assignments, Quiz etc.)
    A. 30 marks (Any one of the following):
        a. Written Test

        or

        b. SWAYAM NPTEL (Advanced Course) of minimum 20 hours and certification examination completed

        or

        c. Valid International Certifications (Prometric, Pearson, Certiport, Coursera, Udemy, edx and the like).

        or

        d. One certification mark shall be awarded one course only. For four courses, the students will have to complete four certifications.

    B. 10 Marks

        The marks given out of 40 for publishing the research paper should be divided into four courses and should awarded out of 10 in each of the four courses.
2. One External (Semester End Examination) of 60 marks. Duration 2 ½ hours.
3. One Practical at the end of Semester consisting of 50 marks.
4. Minimum marks for passing Semester End Theory and Practical Exam is 40 %. Separate passing for Internal and Semester End examination.
5. For any KT examinations, there shall be ODD-ODD/EVEN-EVEN pattern followed.
6. A candidate will be allowed to appear for the practical examinations if he/she submits a certified journal or a certificate from the Head of the department / Institute to the effect that the candidate has completed the practical course as per the minimum requirements.
7. In case of loss of journal, a candidate must produce a certificate from the Head of the department /Institute that the practical for the academic year were completed by the

student. However, such a candidate will be allowed to appear for the practical examination but the marks allotted for the journal will not be granted.

8. HOD's decision, in consultation with the Principal, shall remain final and abiding to all.