Hindi Vidya Prachar Samiti's

RAMNIRANJAN JHUNJHUNWALA COLLEGE OF ARTS, SCIENCE & COMMERCE (Autonomous)

University of Mumbai

Refer to page nos: 04 and 05 highlighting component of Research Project/Internship

Affiliated to

University of Mumbai

Syllabus for the M. Sc. Part I

Program: M. Sc. in Data Science & Artificial Intelligence

Program Code: RJSPGDSAI

(Choice Based Credit System with effect from the academic year 2022 – 2023 for Part I and 2023 – 2024 for Part II)

Syllabus for Approval

Sr. No.	Heading	Particulars
1	Title of course	M.Sc. in Data Science and Artificial Intelligence
2	Eligibility for admission	B.Sc. Computer Science, B.Sc. Information Technology and B.Sc. Statistics. B.Sc. Mathematics, BTech.
3	Passing Marks	40%
4	No. of Years, Semesters	2 Years, 4 Semesters
5	Level	Post Graduate
6	Pattern	Semester
7	Status	Introduced
8	To be implemented from Academic Year	Part I: 2020 – 2021, Part II: 2021 – 2022

Program Educational Objectives

1. To enable graduates to excel professionally by adapting to the dynamic needs of the academia, industry and research in the field of Data Science and Artificial Intelligence.
2. To develop in depth understanding of the key technologies in Data Science and Artificial Intelligence.
3. To enable student to excel in the field of Data Analytics, Data Mining, Machine Learning, Visualization Techniques, Predictive Analysis and Statistical modelling.
4. To practice the problems of analysis and decision making using big data.
5. To gain practical, hands-on experience with programming languages, data analysis tools and frameworks through coursework.
6. To enable graduates to use the concepts of machine learning, deep learning and natural language processing in the applications of Artificial Intelligence.

Program Outcomes

Students who have completed the M.Sc. in Data Science and Artificial Intelligence will be able to:

1. To apply statistical modelling and data analysis techniques to the solution of real-world business problems, effectively present results using data visualization techniques.
2. To test and train various machine learning algorithms for real world data and applications.
3. To create data warehouse and mine the data for analysis.
4. To analysis big data using various languages and tools.
5. To apply machine learning and deep learning algorithms to real-world problems.
6. To create the applications for analysis using deep learning and natural language processing concepts.
7. To recognize and analyze ethical issues in data security, integrity and privacy.

Course Structure

Semester I

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGITDSAI101	Statistics for Data Science	CC	4	-	4
RJSPGITDSAI102	Data Warehousing	CC	4	-	4
RJSPGITDSAI103	Artificial Intelligence	CC	4	-	4
RJSPGITDSAI1L1	PG Lab – I (FDS)	PGL	-	2	2
RJSPGITDSAI1L2	PG Lab – II (DW)	PGL	-	2	2
RJSPGITDSAI1L3	PG Lab – III (AI)	MNP	-	2	2
RJSPGITDSAI1S1	Seminar – I	SE	-	1	1
RJSPGITDSAI1P1	Professional Elective – I (Python for DS/ Applied Mathematics for DS)	PE	3	-	3
RJSPGITDSAI1I1	Career Advancement Course (RIC)	CAC	2	-	2
	Total		**16**	**8**	**24**

Semester II

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGITDSAI201	Machine Learning	CC	4	-	4
RJSPGITDSAI202	Big Data Technology	CC	4	-	4
RJSPGITDSAI203	Soft computing	CC	4	-	4
RJSPGITDSAI2L1	PG Lab – IV (ML)	PGL	-	2	2
RJSPGITDSAI2L2	PG Lab – V (BDT)	PGL	-	2	2
RJSPGITDSAI2L3	PG Lab – VI (SC)	MNP	-	2	2
RJSPGITDSAI2R1	Mini Project	SE	-	1	1
RJSPGITDSAI2P2	Professional Elective – II (IP&CV/CC)	PE	3	-	3
RJSPGITDSAI2I1	Career Advancement Course (BI)	CAC	2	-	2
	Total		**16**	**8**	**24**

Semester III

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGITDSAI301	Natural Language Processing	CC	4	-	4
RJSPGITDSAI302	Big Data Analytics	CC	4	-	4
RJSPGITDSAI303	Deep Learning	CC	4	-	4
RJSPGITDSAI3L1	PG Lab – VII (NLP)	PGL	-	2	2
RJSPGITDSAI3L2	PG Lab – VIII (BDA)	PGL	-	2	2
RJSPGITDSAI3L3	PG Lab – IX (DL)	DES	-	2	2
RJSPGITDSAI3PL4	PG Lab – X (E-I / E-II)	PEL		2	2
RJSPGITDSAI3P3	Professional Elective – III (Robotics & RPA/ Block chain)	PE	4	-	4
	Total		**16**	**8**	**24**

Semester IV

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGITDSAI4D2	Dissertation – II (Major Project)	DES	-	30	12
RJSPGITDSAI4D3	Industrial Internship	II	-	30	12
	Total		**-**	**60**	**24**

Note:

Student have to register for the courses as per the following guidelines:

Sr. No.	Category	Credits				Total Credits
		Semester I	Semester II	Semester III	Semester IV	
1	Core Courses (CC)	12 (3 Courses)	12 (3 Courses)	12 (3 Courses)	-	36
2	PG Labs (PGL) PEL	6 (3 Courses) -	6 (3 Courses) -	6 (3 Courses) 2 (1 Course)	-	12 2
3	Mini Project (MNP)	-	1	-	-	1
4	Seminar (SE)	1	-	-	-	1
5	Professional Electives (PE)	3	3	4	-	10
6	Career Advancement Course (CAC)	2	2	-	-	4
7	Dissertation – II (Major Project) (DES)	-	-	-	12	12
8	Industrial Internship (II)	-	-	-	12	12
Total Credits		24	24	24	24	**96**

Core Courses

Semester I

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGITDSAI101	Statistics of Data Science	CC	4	-	4

Course Objectives

1. To provide basic knowledge of data science.
2. To provide the foundation on topics in statistical methods and applied probability which forms the basis for data science.
3. To provide the foundation on topics of mathematics which 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, Linear Transformations Data Collection, Modelling and Compilation, Data Analysis, Data Presentation and Visualization Data Science Software Tools, Programming Languages for Data Science, Applications of Data Science	10
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, Oneway 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,	10
	Descriptive Statistics, Compound probability, Conditional probability. **Data Transformations and quality analysis** Merge, Rollup, Transpose and Append, Missing Analysis and Treatment, Outlier analysis and treatment.	
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 twoway ANOVA, Chi-square test, Fisher's exact test, Power and sample size. **Regression and Prediction** 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.	10

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
			Lectures	Practical	
RJSPGITDSAI102	Data Warehousing	CC	4	-	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, 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 Modeling Primer** Different Worlds of Data Capture and Data Analysis, Goals of Data Warehousing and Business Intelligence, Dimensional Modeling Introduction, Kimball's DW/BI Architecture, Alternative DW/BI Architectures, Dimensional Modeling Myths.	10
Unit II	**Kimball Dimensional Modeling 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 Data Integration Approaches, Low Latency Reality Check.	10

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
			Lectures	Practical	
RJSPGITDSAI103	Artificial Intelligence	CC	4	-	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 RealTime 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 **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.	10
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	**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

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
			Lectures	Practical	
RJSPGITDSAI1L1	**PG Lab – I** **Statistics of Data Science**	**PGL**	**-**	**2**	**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
			Lectures	Practical	
RJSPGITDSAI1L2	PG Lab – II Data Warehousing	PGL	-	2	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.

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGITDSAI1L3	PG Lab - III Artificial Intelligence	PGL	-	2	2

Practical List:

1. Write a programme to conduct uninformed search.(BFS,DFS etc)
2. Write a programme to conduct informed search.(A*, AO*)
3. Write a programme to conduct heuristic search.(Hill Climbing)
4. Write a programme to conduct game search.(Tic-tac-toe,N-Queen,Tower of Hanoi)
5. Simple/Multiple Linear Regression.
6. Bayesian Classification.
7. Decision tree Classification.
8. Write a programme to do reinforcement learning in a grid world.
9. Write a programme to run value and policy iteration in a grid world.
10. Implement Artificial Neural Network.

Professional Electives

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGITDSAI1P1	Professional Elective – I: Python for Data Science	PE	3	-	3

Course Objectives

1. To learn the concepts of python programming.
2. To provide in-depth knowledge python.
3. To learn how to build and create program for various applications in data science, Data Analysis, Data analysis.

Learning Outcomes

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

1. Create the programs in python for different application.
2. Perform cleansing on data and various functions and classes.
3. Enable students to programming in object-oriented programming.

Unit	Topics	Lectures
Unit I	**Introduction of high-level language:** Keywords and identifiers, statements & comments, python variables, Datatypes and type conversion, I/O and import, python operators, Namespace. **Python Flow Control:** If-else, for and while loop, continue and break statement, pass, try catch, **Python Functions:** Functions and arguments, recursion function, lambda function, Built-in function global local functions, Global keywords, Modules and Packages.	10
Unit II	**Python Datatypes:** Numbers, List, Tuple, Array, string, Set, Dictionary. **Python Files:** File operation, Directory, Exception handling. **Python Object & Class:** Introduction of OPPs, class, Inheritance, operator overloading.	10
Unit III	**Exploratory Data Analysis**: data analysis with pandas dataframe, data cleansing, report generation and web scrapping. **Visual Data Analysis**: Pyplot, Plotting, Markers, Line, Labels, Grid, Subplot, Scatter, Bars, histogram, Piecharts, heatmap, boxplot.	10

Unit IV	**Advance Python Programming** File handling, read and write files, delete file, with statement. map, itr and zip functions, Regular Expression, abstract classes, constructors and destructors, decorators and Generators	10

References

1. "Programming Python, Book by Mark Lutz. "
2. "Fluent Python, Book by Luciano Ramalho."
3. "https://www.w3schools.com/python/default.asp"

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGITDSAI1P1	**Professional Elective – I: Mathematics for Data Science**	**PE**	**3**	**-**	**3**

Course Objectives

1. To learn the concepts of mathematics that used in field of Data Science.
2. To provide in-depth knowledge of Linear Algebra and Calculus.

Learning Outcomes

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

1. Demonstrate understanding of basic mathematical concepts in data science, relating to linear algebra, and calculus
2. Employ methods related to these concepts in a variety of data science applications.
3. Apply logical thinking to problem-solving in context.
4. Use appropriate technology to aid problem-solving and data analysis.

Unit	Topics	Lectures
Unit I	**Linear Algebra:** Scalars, Vectors, Matrices and their properties, Vector Addition and Multiplication, Norm of a vector, Dot product of two vectors, Cross products , Relation between norm and dot product, Orthogonal and Orthonormal Vectors, Linear Independence of vectors, Linear Dependence and span **Scalars, Vectors and Matrices:** Matrices, Visualizing matrices, Determinants, Properties of Matrices, Matrix multiplication, Types of Matrices, Transpose of matrix, Identity and Inverse of a Matrix, Determinant of a Matrix **Eigens:** Eigen values, Eigen Vectors	10

Unit II	**Functions:** Functions, New Functions from Old, Families of Functions, Inverse Functions; Inverse Trigonometric Functions, Exponential and Logarithmic Functions **LIMITS AND CONTINUITY:** Limits , Computing Limits, Limits at Infinity; End Behaviour of a Function, Continuity, Continuity of Trigonometric, Exponential, and Inverse Functions **THE DERIVATIVE:** Tangent Lines and Rates of Change, The Derivative Function, Introduction to Techniques of Differentiation, The Product and Quotient Rules, Derivatives of Trigonometric Functions, The Chain Rule, Euclidian Geometry	10
Unit III	**DIFFERENTIATION:** Implicit Differentiation, Derivatives of Logarithmic Functions, Derivatives of Exponential and Inverse Trigonometric Functions **THE DERIVATIVE IN GRAPHING AND APPLICATIONS:** Analysis of Functions I: Increase, Decrease, and Concavity, Analysis of Functions II: Relative Extrema; Graphing Polynomials, Absolute Maxima and Minima, Applied Maximum and Minimum Problems, Rolle's Theorem; Mean-Value Theorem	10
Unit IV	**INTEGRATION:** An Overview of the Area Problem, The Indefinite Integral , Integration by Substitution , The Definition of Area as a Limit; Sigma Notation, The Definite Integral, The Fundamental Theorem of Calculus, Rectilinear Motion Revisited Using Integration, Average Value of a Function and its Applications, Evaluating Definite Integrals by Substitution, Logarithmic and Other Functions Defined by Integrals **PARTIAL DERIVATIVES:** Functions of Two or More Variables, Limits and Continuity, Partial Derivatives, Differentiability, Differentials, and Local Linearity, The Chain Rule, Directional Derivatives and Gradients, Tangent Planes and Normal Vectors, Maxima and Minima of Functions of Two Variables	10
References 1. Calculus: by Howard Anton		

Career Advancement Course

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGITDSAI1I1	Research in Computing	CAC	2	-	2

Course Objectives

1. To be able to conduct business research with an understanding of all
2. the latest theories.
3. To develop the ability to explore research techniques used for solving
4. 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:** research meaning and characteristic, research objectives, Positivism and postpositivistic approach to research. **Business Research:** Role of Business Research, Information Systems and Knowledge Management, Theory Building, Organization ethics and Issues. **Beginning Stages of Research Process:** Problem definition, Qualitative research, Quantitative Research, primary and Secondary data research.	10
Unit II	**Research Methods and Data Collection:** Survey research, communicating with respondents, Observation methods, Descriptive and experimental Research type, Inductive and deductive approach, Action research, research steps. **Formulation of research problem:** problem selection, literature review, formulation of hypothesis. **Variables:** dependent, independent and Intervening variables.	10
Unit III	**Data collection and sampling:** Probability sampling, Non probability sampling, Survey method, contact method, questioner. **Selection of project domain:** Publication ethics, Tools and evaluation. Selection of tentative project area and process of literature survey – Literature survey components and procedures Basic components of a research paper – procedures and processes, Journal types, Scopus, web of science, Science Citation Index, H-index, Google citations.	10

Unit IV	**Research Paper Writing** Title selection, paragraph writing, report design, conclusion formation, diagrams and equation, citations, plagiarism, paper format, scopes index journals, predatory journals, digital object identifier/ISBN number and publication, research ethics. **Presentation of selected project proposal:** Oral presentation. Preparation of a report on the selected project proposal, Attending special invited lectures, practical orientation in searching and collecting literature through library, online tools, presenting a seminar on selected project.	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
			Lectures	Practical	
RJSPGITDSAI201	**Machine Learning**	**CC**	**4**	**-**	**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:** What is Dataset, different sources of repository/different format, data pipeline and flow, web scraping and ethics, understandability of data, training and testing data, Instance-base and model base learning, ML vs traditional programming, MLDLC, challenges in ML **Feature Engineering:** Feature transformation, feature construction, feature scaling(Standardization & Normalization), feature extraction(PCA,LDA),function and power transformation.	10

Unit II	**Supervised Learning:** **Regression:** what is regression, linear model and linear seperability, types of LR, bestfit-line, relationship, cost function, accuracy metrics, where LR fails, optimize using gradient decent, Polynomial regression. **Regularization**: Overfitting vs underfitting, Ridge, lasso and ElasticNet regression. **Classification:** what is classification, logistic regression, SVM, KNN, Decision tree, Random forest, Naive bayes, data imbalance, classification metrics and precision and recall. Repressor vs classifier. **Ensemble Learning:** Introduction to ensemble learning, Voting ensembles, Bagging techniques, Hyperparamter tuning, AdaBost Algorithm, stacking and blending, bagging vs boosting.	10
Unit III	**Unsupervised Learning:** **Clustering:** whyclustering, partition, Hierarchical, Density and grid based clustering application of unsupervised learning. **Association Rule Learning:** Market Basket analysis, support, confidence and lift, Apriori algorithm, F-P growth. **Dimensionality Reduction:** curse of dimensionality, Principal Component Analysis, t-sne, auto-encoders.	10
Unit IV	**Reinforcement Learning**: Value functions, Bellman's equation, value iteration, policy iteration, Markov Decision processes(MDPs) Q-Learning. **Time series analysis(TSA):**what is TSA and it significance, component of time series, limitation of TSA, station and non-station data and conversion, auto regressive model(AR),ARIMA,SARIMA model, time series model selection.	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
			Lectures	Practical	
RJSPGITDSAI202	**Big Data Technology**	**CC**	**4**	**-**	**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** 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.	10

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
			Lectures	Practical	
RJSPGITDSAI203	**Soft Computing**	**CC**	**4**	**-**	**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, McCullohPitts 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**	10

	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 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
RJSPGITDSAI2L1	**PG Lab – III** **Machine Learning**	**PGL**	**-**	**2**	**2**

Practical List:

1. All Feature Engineering Operations.
2. Simple and Multiple linear regression.
3. Logistics regression, SVM and KNN
4. Regularization with Penalty
5. Ensemble learning
6. Clustering using K-means and DBScan.
7. Dimensionality reduction techniques
8. Market Basket Analysis using Apriori Algorithm
9. Time Series Model
10. Q Learning

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
RJSPGITDSAI2L2	**PG Lab – IV** **Big Data Technology**	**PGL**	**-**	**2**	**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.

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
RJSPGITDSAI2L3	**PG Lab – III** **Soft Computing**	**PGL**	**-**	**2**	**2**

Practical List:

Implement the following:

1. Design a simple linear neural network model.
2. Calculate the output of neural net using both binary and bipolar sigmoidal function.

Implement the following:

1. Generate AND/NOT function using McCulloch-Pitts neural net.
2. Generate XOR function using McCulloch-Pitts neural net.

Implement the Following

1. Write a program to implement Hebb's rule.
2. Write a program to implement of delta rule.

Implement the Following

1. Write a program for Back Propagation Algorithm
2. Write a program for error Backpropagation algorithm.

Implement the Following

1. Write a program for Hopfield Network.
2. Write a program for Radial Basis function.

Implement the Following

1. Kohonen Self organizing map.
2. Adaptive resonance theory.

Implement the Following

1. Write a program for Linear separation.

Implement the Following

2. Membership and Identity Operators | in, not in, is, is not.

Implement the Following

3. Find ratios using fuzzy logic
4. Solve Tipping problem using fuzzy logic.

Implement the Following

1. Implementation of Simple genetic algorithm
2. Create two classes: City and Fitness using Genetic algorithm

Professional Electives

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGITDSAI2P2	Data Science in Cloud Computing	PE	3	-	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 Products available for Virtualization.	10
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 AppsClient-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**	10
	Privacy, Data Life Cycle, Key Privacy Concems in the Cloud. Protecting Privacy, Privacy Risk Management and Compliance in Relatior to Cloud Computing.	

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
			Lectures	Practical	
RJSPGITDSAI2P2	**Image Processing & Computer Vision**	**PE**	**3**	**-**	**3**

Course Objectives

1. To study the image fundamentals and mathematical transforms necessary for image processing.
2. To study the image enhancement, restoration and compression techniques.
3. To study computer vision technique.

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 computer vision.
3. Categorize various OpenCv techniques.
4. Interpret computer vision standards.
5. Interpret image segmentation and representation techniques.

Unit	Topics	Lectures
Unit I	**Introduction to Image-processing System :** Introduction, Image Sampling, Quantization, Resolution, Human Visual Systems, Elements of an Image-processing System, Applications of Digital Image Processing **2D Correlation Image Transforms:** Need for transform, image transforms, Fourier transform, 2D Discrete Fourier Transform, Properties of 2D DFT, Importance of Phase, Walsh transform, Hadamard transform, Haar transform, Slant transform, Discrete Cosine transform, KL transform	10
Unit II	**Image Enhancement** : Image Enhancement in spatial domain, Enhancement trough Point operations, Histogram manipulation, Linear and nonlinear Gray Level Transformation, local or neighborhood operation, Median Filter, Spatial domain High pass filtering, Bit-plane slicing, Image Enhancement in frequency domain, Homomorphic filter, Zooming operation, **Colour Image processing :** Colour images, Colour Model, Colour image quantization, Smoothing and sharpening,Image segmentation bases on colour, Noise in colour images	10
Unit III	**Morphological Image Processing**: Introduction, Erosion and Dilation, Opening and closing, History, Miss transformation, Basic morphological algorithms, Gray scale morphology. **Image Segmentation:** Image segmentation techniques, Region approach, Clustering techniques, Thresholding, Edge-based segmentation, Edge detection, Edge Linking, Hough Transform **Image Compression:** Need for image compression, Redundancy in images, Image-compression scheme, Fundamentals of Information Theory, Run-length coding, Shannon-Fano coding, Huffman Coding, Arithmetic Coding, Transform-based compression, Image-compression standard.	10
Unit IV	**Introduction to computer vision:** Image processing vs computer vision, Overview of problems of machine vision and pattern classification, Image formation and processing, Feature extraction from images, Biological object recognition, Bayesian modeling and inference, Object detection and recognition, Morphable models, object tracking and detection, Genture recognition. Image formation:radiometry, shape from shading, image formation, background modeling and motion estimation, sementic segmentation **Deep learning for computer vision:** Convolutional neural network, Recurrent neural network and image caption generator, deep generative model, recent trends.	10

References

1) Digital Image Processing, S Jayaraman, S Esakkirajan, T Veerakumar,Tata McGraw-Hill Education Pvt. Ltd., 2009.
2) Computer Vision: Algorithms and Applications 2nd Edition Richard Szeliski.

Career Advancement Course

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGITDSAI2I1	**Business Intelligence and Analytics**	**CAC**	**2**	**-**	2

Course Objectives

1. To understand how accurately represent voluminous complex data set in business intelligence analytics.
2. To understand the methodologies used to visualize large data sets.
3. To understand the process involved in business processes and visualization 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 business inteligence
2. Design the process involved Power BI and Tableu present.

Unit	Topics	Lectures
Unit I	**Introduction to Business Intelligence:** BI concept, I architecture, BI in today's perspective, BI Process Applications of BI like Financial analysis, statistical analysis, sales analysis CRM,ERP , result pattern, and ranking analysis, Balanced Scorecard, BI in Decision Modelling: Optimization, Decision making under uncertainty. Ethics and business intelligence. **Logistic and production models:** Supply chain optimization, Optimization modelsfor logistics planning, Revenue management systems. **Data envelopment analysis**: Efficiency measures, Efficient frontier, The CCR model, Identification of good operating practices	10
Unit II	**Data Visualization and Dashboard Design** : Responsibilities of BI analysts by focusing on creating data visualizations and dashboards. Importance of data visualization, types of basic and composite charts **Performance Dashboard:** Measuring, Monitoring and management of Business, KPIs and dashboard, the types of dashboards, the common characteristics of Enterprise dashboard, design of enterprise dashboards, and the common pitfalls of dashboard design.	10
Unit III	**Power BI:** introduction to power BI, natural-language queries, power BI Visualization charts, BI reports and dashboard. **Tableau:** uploading excel file, text and different files, Bins, Joining tables and data blending, Report generation, set and combined sets, data labels, sorting data and perform aggregation function, data visualization, custom sql, tableau advance reports and calculations, dashboard design dataserver.	10

Unit IV	**Modelling and Analysis:** Exploring Excel Modeling capabilities to solve business problems, summarize and present selected data, introduction to business metrics and KPIs, creating cubes using Microsoft Excel **Future of Business Intelligence:** Emerging Technologies, Machine Learning, Predicting the Future with the help of Data Analysis, BI Search & Text Analytics – Advanced Visualization – Rich Report.	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.

Semester III

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGITDSAI301	**Machine Learning - II**	**CC**	**4**	**-**	**4**

Course Objectives

1. To give a comprehensive coverage of analysis methods like cross validation and bootstrap.
2. To study linear and non-linear models, model selection and regularization.
3. To study highly effective analysis methods like decision trees, Forests and Support Vector machines.

Learning Outcomes

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

1. Apply various data analysis concepts to the data sets.
2. Apply quantitative modelling and data analysis techniques to solve real-world business problems.
3. Analyse data using efficient linear and non-linear models.
4. Analyse data using decision trees, Forests and Support Vector machines.

Unit	Topics	Lectures
Unit I	**Resampling Methods** Cross validation, the bootstrap. **Linear model Selection and Regularization** Subset selection, shrinkage methods, dimension reduction methods, consideration in high dimensions.	10
Unit II	**Moving Beyond Linearity** Polynomial regression, step functions, basis function, regression splines, smoothing splines, local regression, generalized addictive models.	10

Unit III	**Tree and forest** The basics of decision tree classifier, decision tree regressor, bagging, random forest classifier, random forest regressor, boosting.	10
Unit IV	**Support Vector Machines** Maximal margin classifier, support vector classifiers, support vector machines, SVMs with more than two classes, relationship to Logistic regression.	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. "Practical Data Science", Andreas Francois Vermeulen, APress, 2018
5. "Principles of Data Science", Sinan Ozdemir, Packt, 2016.
6. "Machine Learning: A Probabilistic Perspective", Kevin P. Murphy, MIT Press, 2012

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGITDSAI302	**Big Data Analytics**	**CC**	**4**	**-**	**4**

Course Objectives

1. To learn the computational approaches to Modeling and Feature Extraction.
2. To Learn the need and application of Map Reduce.
3. To learn the various search algorithms applicable to Big Data.
4. To analyse and interpret streaming data
5. To learn how to handle large data sets in main memory.
6. To learn the various clustering techniques applicable to Big Data.

Learning Outcomes

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

1. Design algorithms by employing Map Reduce technique for solving Big Data problems.
2. Design algorithms for Big Data by deciding on Features set.
3. Design algorithms for handling big size datasets and propose solutions for Big Data by optimizing main memory consumption.
4. Design solutions for problems in Big Data by suggesting appropriate clustering techniques.

Unit	Topics	Lectures

Unit I	**Data Mining and Large-Scale Files** Introduction to Statistical modelling, Machine Learning, Computational approaches to modelling, Summarization, Feature Extraction, Statistical Limits on Data Mining, Distributed File Systems, Map-reduce, Algorithms using Map Reduce, Efficiency of Cluster Computing Techniques.	10
Unit II	**Mining Data Streams** Stream Data Model, Sampling Data in the Stream, Filtering Streams, Counting Distance Elements in a Stream, Estimating Moments, Counting Ones in Window, Decaying Windows **Clustering** Introduction to Clustering Techniques, Hierarchical Clustering, Algorithms: K-Means, CURE, Clustering in Non-Euclidean Spaces, Streams and Parallelism, Case Study: Advertising on the Web-Recommendation Systems	10
Unit III	**Introduction to NOSQL** Definition of NOSQL, History of NOSQL and Different NOSQL products, Exploring MondoDB Java/Ruby/Python, Interfacing and Interacting with NOSQL 2. NOSQL Basics NOSQL Storage Architecture, CRUD operations with MongoDB, Querying, Modifying and Managing NOSQL Data stores, Indexing and ordering datasets (MongoDB/CouchDB/Cassandra)	10
Unit IV	**Introduction to Spark** Introduction to Spark, Components of the Spark unified stack, Resilient Distributed Dataset (RDD). **Resilient Distributed Dataset and Data Frames**	10
	Creation of parallelized collections and external datasets, Resilient Distributed Dataset (RDD) operations, shared variables and keyvalue pairs. **Spark application programming** Purpose and usage of the Spark Context, Initialize Spark with the various programming languages, Describe and run some Spark examples, Pass functions to Spark, Create and run a Spark standalone application, Submit applications to the cluster, Introduction to Spark libraries.	

References

1. "Mining of Massive Datasets", Jure Leskovec, AnandRajaraman, Jeffrey David Ullman, Cambridge University Press, Second Edition, 2014.
2. "Data Mining Concepts and Techniques", Jiawei Han, MichelineKamber, Jian Pei, Morgan Kaufman Publications, Third Edition, 2011.
3. "Data Mining – Practical Machine Learning Tools and Techniques", Ian H.Witten, Eibe Frank, Morgan Kaufman Publications, Third Edition, 2011.
4. "Principles of Data Mining", David Hand, HeikkiMannila and Padhraic Smyth, MIT PRESS.
5. Dan Sullivan,"NoSQL for Mere Mortals",1 stEdition, Pearson Education, 2015. (ISBN-13: 978-9332557338)
6. https://cognitiveclass.ai/courses/what-is-spark

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGITDSAI303	Deep Learning	CC	4	-	4

Course Objectives

1. To study fundamental concepts of Deep Learning and its applications.
2. To provide knowledge on fundamentals of deep networks, activation functions, loss functions and hyperparameters.
3. To study feed-forward and backpropagation approaches of deep networks.
4. To study major architectures of deep networks like Convolutional Neural Network, Recurrent Neural Network, Recursive Neural Network and Autoencoders.

Learning Outcomes

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

1. Train and test the various deep networks like Feed Forward Network, Convolutional Neural Network, Recurrent Neural Network, Autoencoder and Recursive Neural Network.
2. Create the various applications like Face Detection, Handwriting Recognition, Sentiment Analysis, etc. using Deep Neural Networks.
3. Optimize and Fine Tune the Deep Neural Network.

Unit	Topics	Lectures
Unit I	**Fundamentals Concepts of Machine Learning**	10
	Linear Algebra for machine Learning, Testing, Cross-Validation, Dimensionality reduction, Over/Under-fitting, Hyper parameters and validation sets, Estimators, Bias, Variance, Regularization-Introduction to a simple DNN, Platform for deep learning, Deep learning software libraries. **Deep Feed Forward Networks** Learning XOR, Gradient-Based Learning, Hidden units, Various Activation Functions, error functions, Architecture Design and other differentiation algorithms. **Regularization for Deep Learning** Parameter norm penalties, Early Stopping, Drop out.	
Unit II	**Convolutional Neural Networks and Sequence Modeling** Convolutional Networks: Convolutional operation, Motivation, Pooling, Normalization. **Applications in Computer Vision** ImageNet. **Sequence Modeling** Recurrent Neural Networks, Difficulty in Training RNN, Encoder, Decoder.	10
Unit III	**Auto encoders** Under complete, regularized, stochastic, denoising, contractive, applications, dimensionality reduction, classification, recommendation, Optimization for Deep Learning: optimizers. RMS Prop for RNNs, SGD for CNNs.	10

Unit IV	**Deep Architectures in Vision** Alexnet to ResNet, Transfer learning, Siamese Networks, Metric Learning, Ranking/Triplet loss, RCNNs, CNN, RNN, Applications in captioning and video tasks, 3D CNNs.	10

References

1. "Deep Learning", Ian Goodfellow, Yoshua Bengio, Aaron Courville, MIT Press, 2016 (available at *http://www.deeplearningbook.org)*
2. "Deep Learning: A Practitioner's Approach", Josh Patterson and Adam Gibson, O'Reilly, 2017.
3. "Machine Learning: A Probabilistic Perspective", Kevin P. Murphy, MIT Press, 2012
4. "Neural Networks and Deep Learning", Michael Nielsen, Online book, 2016 (http://neuralnetworksanddeeplearning.com/)
5. "Pattern Recognition and Machine Learning", Christopher and M. Bishop, Springer Science Business Media, 2006.
6. "Deep Learning with Python", Jason Brownlee, eBook, 2016
7. "Deep Learning Step by Step with Python: A Very Gentle Introduction to Deep Neural Networks for Practical Data Science", N. D. Lewis.

PG Labs

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
RJSPGITDSAI1L3	**PG Lab – V** **Machine Learning - II**	**PC**	**-**	**2**	**2**

Practical List:

1. Cross-validation, model evaluation and selection
2. Bootstrap
3. Dimensionality reduction using feature extraction.
4. Dimensionality reduction using feature selection.
5. Polynomial Regression
6. Tree classifier and regressor
7. Random forest classifier and regressor
8. Training linear classifier
9. Creating predicated probabilities
10. Identifying support vectors

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
RJSPGITDSAI1L3	**PG Lab – VI** **Big Data Analytics**	**PC**	**-**	**2**	**2**

Practical List:

1. Feature extraction, summarization and data modelling for big data sets.
2. Mining data streams and estimating moments.
3. Implementation of Classification Algorithms Using Big Data.
4. Implementation of Clustering Algorithms Using Big Data.
5. Installation of Spark and Scala.
6. Creating spark application.
7. Using spark libraries.
8. Web and Social Analytics.
9. Finance and Risk Analytics.
10. Supply Chain and Logistic Analysis.

Professional Electives

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGITDSAI3P3A	**Data Science for Cyber Security**	**PE**	**3**	**-**	**3**

Course Objectives

1. To learn different Cyber Threats, Various techniques of collecting Cyber Threat Intelligence Requirements and Information.
2. To learn analysis and disseminating Cyber Threat Intelligence.
3. To identify and document Risks, Threats and vulnerabilities for organization's productions for the infrastructure and assets.

Learning Outcomes

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

1. Identify and understand various cyber threats, accurately access threats, risks, and vulnerabilities, to minimize the potential for incidents.
2. Identify Vulnerabilities and document it for organization's Productions for the infrastructure and assets.

Unit	Topics	Lectures
Unit I	**Defining Cyber Threat Intelligence** The Need for Cyber Threat Intelligence: The menace of targeted attacks, the monitor and respond strategy, Why the strategy is failing, Cyber Threat Intelligence Defined, Key Characteristics: Adversary based, Risk focused, Process oriented, Tailored for diverse consumers, The Benefits of Cyber Threat Intelligence.	10

Unit II	**Developing Cyber Threat Intelligence Requirements** Assets That Must Be Prioritized: Personal information, Intellectual property, Confidential business information, Credentials and IT systems information, Operational systems. Adversaries: Cybercriminals, Competitors and cyber espionage agents, Hacktivists. Intelligence Consumers: Tactical users, Operational users, Strategic users.	10
Unit III	**Collecting Cyber Threat Information** Level 1: Threat Indicators, File hashes and reputation data, Technical sources: honey pots and scanners, Industry sources: malware and reputation feeds. Level 2: Threat Data Feeds, Cyber threat statistics, reports, and surveys, Malware analysis. Level 3: Strategic Cyber Threat Intelligence, Monitoring the underground, Motivation and intentions, Tactics, techniques, and procedures. **Analyzing and Disseminating Cyber Threat Intelligence** Information versus Intelligence, Validation and Prioritization: Risk scores, Tags for context, Human assessment. Interpretation and Analysis: Reports, Analyst skills, Intelligence platform, Customization. Dissemination: Automated feeds and APIs, Searchable knowledge base, Tailored reports.	10
Unit IV	**Data Gathering Sampling**	10
	The RIIOT Method of Data Gathering, Administrative Data Gathering, Technical Data Gathering., Physical Data gathering. **Risk Analysis** Determining Risk, Creating Risk Statements, Team Review of Security Risk Statements, Security Risk,Mitigation, Security Risk Assessment Reporting Security Risk Assessment Approaches. Quantitative vs. Qualitative Analysis, Qualitative Analysis, Tools, Security Risk Assessment Methods, Relevant Standards and Regulations.	

References

1. "Cyber Threat Intelligence", Definitive Guide TM, By Jon Friedman. Mark Bouchard, CISSP. Foreword by John P. Watters, 2015.
2. "Intelligence, Driven Incident Response: Outwitting the Adversary", Scott J. Roberts, Rebekah Brown, , O'Reilly Media, 2017.
3. "How to Define and Build an Effective Cyber Threat Intelligence Capability Elsevier Science & Technology", Henry Dalziel, 2014.
4. "Dark Web Cyber Threat Intelligence Mining", Cambridge University Press, John Robertson, Ahmad Diab, Ericsson Marin, Eric Nunes, Vivin Paliath, Jana Shakarian, Paulo Shakarian, 2017.
5. "The Cyber Threat", Bob Gourley, CreateSpace Independent Pub, 2014.

Course Code	Course Name	Group	Teaching Scheme (Hrs/Week)		Credits
			Lectures	Practical	
RJSPGITDSAI3P3B	**Natural Language Processing**	**PE**	**3**	**-**	**3**

Course Objectives

1. To learn structure of sentence.
2. To learn Morphological analysis, Lexical analysis, Syntactic and Semantic analysis.
3. To learn feature engineering concepts and rule-based systems for NLP.
4. Using Machine learning and deep learning for NLP.

Learning Outcomes

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

1. Analyse corpus and corpora of NL.
2. Learn the language modeling, formal grammars, statistical parsing, machine translation, and dialog processing.
3. Understanding statistical sequence labeling, information extraction, question answering and summarization, advanced topics in speech recognition, speech synthesis.

Unit	Topics	Lectures
Unit I	**Introduction** Understanding natural language processing, Understanding basic applications, Advantages of togetherness, NLP and Python, Environment setup for NLTK. **Practical Understanding of a Corpus and Dataset** What is a corpus? Why do we need a corpus? Understanding corpus analysis, understanding types of data attributes, exploring different file	10
	formats for corpora, Resources for accessing free corpora, Preparing a dataset for NLP applications, Web scraping.	
Unit II	**Understanding the Structure of a Sentences** Understanding components of NLP, Natural language understanding, Defining context-free grammar, Morphological analysis, Syntactic analysis, Semantic analysis, Handling ambiguity, Discourse integration, Pragmatic analysis. **Pre-processing** Handling corpus-raw text, Handling corpus-raw sentences, Basic preprocessing, Practical and customized pre-processing.	10
Unit III	**Feature Engineering and NLP Algorithms** Understanding feature engineering, Basic feature of NLP, Basic statistical features for NLP, Advantages of features engineering, Challenges of features engineering. **Advanced Feature Engineering and NLP Algorithms** Recall word embedding, Understanding the basics of word2vec, Converting the word2vec model from black box to white box, Understanding the components of the word2vec model, Understanding the logic of the word2vec model, Understanding algorithmic techniques and the mathematics behind the word2vec model.	10

Unit IV	**Rule-Based System for NLP** Understanding of the rule-based system, Purpose of having the rule-based system, Architecture of the RB system, Understanding the RB system development life cycle, Applications, Developing NLP applications using the RB system, Comparing the rule, based approach with other approaches, Advantages of the rule, based system, Disadvantages of the rule-based system **Machine Learning and Deep Learning for NLP Problems** Understanding the basics of machine learning, Development steps for NLP applications, Comparing NLU and NLG	10

References

1. "Python Natural Language Processing", Jalaj Thanaki, Packt.
2. "Natural Language Processing with Python: Analyzing Text with the Natural Launguage Toolkit", By Steven Bird, Ewan Klein, and Edward Loper, NLTK.
3. "Speech and Language Processing", Daniel Jurafskey and James H. Martin, Prentice Hall, 2009.
4. "Foundation of Statistical Natural Language Processing", Christopher D. Manning and Hinrich Schutze, MIT Press, 1999.
5. "Foundations of Computational Linguistics", Ronald Hausser, Springer,Verleg, 1999.
6. "Natural Language Understanding", James Allen, Benjamin/Cummings Publishing Co. 1995.
7. "Corpus – Based Methods in Language and Speech Processing", Steve Young and Gerrit Bloothooft, Kluwer Academic Publishers, 1997.

Evaluation Scheme Internal Evaluation (40 Marks)

The internal assessment marks shall be awarded as follows:

1. **30 marks (Any one of the following):**
   a. **Written Test or**
   b. **SWAYAM (Advanced Course) of minimum 20 hours and certification exam completed or**
   c. **NPTEL (Advanced Course) of minimum 20 hours and certification exam completed or**
   d. **Valid International Certifications (Prometric, Pearson, Certiport, Coursera, Udemy and the like)**
   e. **One certification mark shall be awarded one course only. For four courses, the students will have to complete four certifications.**
2. **10 Marks**

   The marks given out of 40 for publishing the research paper should be divided into four course and should awarded out of 10 in each of the four courses.

I. **Suggested format of Question paper of 30 marks for the written test.**

Q. 1.	**Attempt *any two* of the following:**	**[16 M]**
a)		
b)		
c)		
d)		
Q. 2.	**Attempt *any two* of the following:**	**[14 M]**
a)		
b)		
c)		
d)		

II. 10 marks from every course coming to a total of 40 marks, shall be awarded on publishing of research paper in UGC approved Journal with plagiarism less than 10%. The marks can be awarded as per the impact factor of the journal, quality of the paper, importance of the contents published, social value.

External Examination: (60 marks)

	All Questions are compulsory.	
Q. 1.	**Attempt *any three* of the following:**	**[15 M]**
a)		
b)		
c)		
d)		
e)		
f)		
Q. 2.	**Attempt *any three* of the following:**	**[15 M]**
a)		
b)		
c)		
d)		
e)		
f)		

Q. 3.	**Attempt *any three* of the following:**	**[15 M]**
a)		
b)		
c)		
d)		
e)		
f)		
Q. 4.	**Attempt *any three* of the following:**	**[15 M]**
a)		
b)		
c)		
d)		
e)		
f)		
	~~~~~	

Practical Evaluation (50 marks)

A copy of e-journal is essential to appear for the practical examination.

a)	**Practical Question 1.**	**20**
b)	**Practical Question 2.**	**15**
c)	**Journal**	**10**
d)	**Viva**	**05**

OR

a)	**Practical Question 1.**	**35**
b)	**Journal**	**10**
c)	**Viva**	**05**