W.E.F. 2019-20

MTECH-19

M. Tech.

Data Science

Scheme of Instruction and Syllabus

w.e.f. 2019-20

Department of Information Technology

(M.Tech. Programme)

VELAGAPUDI RAMAKRISHNA
SIDDHARTHA ENGINEERING COLLEGE
(An Autonomous, ISO 9001:2015 Certified Institution)
(Approved by AICTE, Accredited by NAAC with 'A' Grade, Affiliated to JNTUK, Kakinada)
(Sponsored by Siddhartha Academy of General & Technical Education)
Kanuru, Vijayawada
Andhra Pradesh - 520007, INDIA.
www.vrsiddhartha.ac.in

VELAGAPUDI RAMAKRISHNA
SIDDHARTHA ENGINEERING COLLEGE
SCHEME OF INSTRUCTIONS FOR TWO YEAR PG PROGRAMME[M.TECH 19]

M.Tech (DATA SCIENCE)

SEMESTER I **Contact Hours : 23**

S.No	Course Code	Title of the Course	L	T	P	C	CE	SE	Total
1.	19ITDS1001	Mathematical Foundations For Data Science	3	0	0	3	40	60	100
2.	19ITDS1002	Machine Learning	3	0	0	3	40	60	100
3.	19ITDS1003	Advanced Algorithms	3	0	0	3	40	60	100
4.	19ITDS1014	A. Cloud Computing and Virtualization B. R For Data Science	3	0	0	3	40	60	100
5.	19ITDS1015	A. Social and Information Network Analysis B. Optimization Techniques For Data Analysis C. Industry need based Course	3	0	0	3	40	60	100
6.	19MTMC1026	Research Methodology and IPR	2	0	0	-	40	60	100
7.	19ITDS1051	Machine Learning Lab	0	0	3	1.5	40	60	100
8.	19ITDS1052	Python for Data Science Lab	0	0	3	1.5	40	60	100
TOTAL			17	0	6	18	320	480	800

L-Lecture, T-Tutorial, P-Practical, C-Credits, CE-Continuous Evaluation, SE-Semester End, T-Total Marks

SEMESTER II **Contact Hours : 25**

S.No.	Course Code	Title of the Course	L	T	P	C	CE	SE	Total
1.	19ITDS2001	Data Visualization	3	0	0	3	40	60	100
2.	19ITDS2002	Bigdata Management	3	0	0	3	40	60	100
3.	19ITDS2003	Business Analytics	3	0	0	3	40	60	100
4.	19ITDS2014	A. Computer Vision B. Deep Learning	3	0	0	3	40	60	100
5.	19ITDS2015	A. Natural Language Processing B. Cyber Security C. Industry Need Based Course	3	0	0	3	40	60	100
6.	19MTAC2036	Technical Report Writing	2	0	0	-	-	-	-
7.	19ITDS2063	Term Paper	0	0	2	1	40	60	100
8.	19ITDS2051	Bigdata And Visualization Lab	0	0	3	1.5	40	60	100
9.	19ITDS2052	Business Analytics Lab	0	0	3	1.5	40	60	100
TOTAL			17	0	8	19	320	480	800

L-Lecture, T-Tutorial, P-Practical, C-Credits, CE-Continuous Evaluation, SE-Semester End, T-Total Marks

Semester III

Contact Hours : 24

S.No.	Course Code	Title of the Course	L	T	P	C	CE	SE	Total
1.	19ITDS3011	Programme Elective-V#	0	0	0	3	-	-	100
2.	19ITDS3061	Project Part-A	0	0	20	10	40	60	100
2.	19ITDS3052	Intership	0	0	4	2	-	100	100
TOTAL			0	0	24	15	40	160	300

L-Lecture, T-Tutorial, P-Practical, C-Credits, CE-Continuous Evaluation, SE-Semester End, T-Total Marks

 # Evaluation done by MOOCS providers will be considered

Semester IV

Contact Hours : 32

S.No.	Course Code	Title of the Course	L	T	P	C	CE	SE	Total
1	19ITDS4061	Project Part-B	0	0	32	16	40	60	100
TOTAL			0	0	32	16	40	60	100

L-Lecture, T-Tutorial, P-Practical, C-Credits, CE-Continuous Evaluation, SE-Semester End, T-Total Marks

Total Credits : 68

SEMESTER I

19ITDS1001 - MATHEMATICAL FOUNDATIONS FOR DATA SCIENCE

Course Category:	Programme Core	**Credits:**	3
Course Type:	Theory	**Lecture-Tutorial-Practice:**	3-0-0
Prerequisites:	Basic Mathematics	**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Analyze the need and importance of Calculus to a data scientist Understand basic mathematical concepts like calculus and linear algebra
	CO2	Derive the probability mass and density functions of transformation of random variables
	CO3	Apply the mathematical and probabilistic foundations of statistical inference in computing
	CO4	Interpret the results of Regression and Correlation Analysis, for forecasting , perform analysis of variance

Contribution of Course Outcomes towards achievement of Program Outcomes(L-Low, M-Medium, H- High)

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PSO 1	PSO 2
CO1	L			H		M	L	L
CO2				H				
CO3	M			M	L	L	H	M
CO4	L		L	L	M		L	L

Course Content

UNIT I:

Derivatives: Derivatives and rates of change, The derivative as a function, Differentiation formulas, Derivatives of trigonometric functions, The chain rule, Implicit differentiation. Applications of Differentiation, Maximum and minimum values, The mean value theorem, How derivatives affect the shape of a graph, summary of curve sketching, Graphing with calculus and calculators, Integrals ,The definite integral, The fundamental theorem of calculus, The substitution rule.

Linear Algebra: Introduction to Vectors, solving linear equations

Eigen values and Eigen vectors: Diagonalzing a matrix, Systems of Differential Equations.

UNIT II:

Probability and statistics:

Introduction- Probability models, sample space, events, algebra of events, graphical methods of representing events, probability axioms, conditional probability, independence of events.

Discrete Random Variables:The Probability Mass Function, Special Discrete Distributions, The Probability Generating Function, Independent Random Variables

Continuous Random Variables: Introduction, The Exponential Distribution, The Reliability and Failure Rate, Some Important

	Distributions, Functions of a Random Variable, Jointly Distributed Random Variables, Order Statistics.
	UNIT III: **Expectation:** Introduction, Moments, Expectation Based on Multiple Random Variables, Transform Methods, Moments and Transforms of Some Distributions, Computation of Mean Time to Failure. **Stochastic Process:**Classification of Stochastic Processes, The Bernoulli Process, The Poisson Process
	UNIT IV: **Statistical Inference:** Introduction, parameter estimation, hypothesis testing **Regression and Analysis of variance:** Introduction, Least-squares Curve Fitting, The Coefficients of Determination, Correlation Analysis, Simple Nonlinear Regression, Higher-dimensional Least-squares Fit, Analysis of Variance.
Text books and Reference books	**Text Book(s):** **[1]** Calculus, 7th Edition by James Stewart 2015 **[2]** Gilbert Strang, Linear Algebra and its applications, Wellesley-Cambridge Press, Fifth Edition, 2016 **[3]** Kishor S. Trivedi, Probability and Statistics with Reliability, Queuing, and Computer Science Applications, John Wiley & Sons, 2016 **Reference Book(s):** **[1]** M. Mitzenmacher and E. Upfal, .Probability and Computing: Randomized Algorithms andProbabilistic Analysis, Cambridge, 2005 **[2]** John Vince, Foundation Mathematics for Computer Science, Springer, 2015
E-resources and other digital material	[1] Maggie Myers, Robert van de Geijn, (24,06,2019). Linear Algebra - Foundations to Frontiers, UTAustinX, https://www.edx.org/ course/linear-algebra-foundations-to-frontiers-0 [2] Statistics And Probability Tutorial | Statistics And Probability for Data Science | Edureka, https://www.youtube.com/watch?v=XcLO4f1i4Yo [3] Dr Nic's Maths and Stats, Understanding Statistical Inference - statistics help, https://www.youtube.com/watch? v=tFRXsngz4UQ

19ITDS1002 - MACHINE LEARNING

Course Category:	Programme Core	**Credits:**	3
Course Type:	Theory	**Lecture-Tutorial-Practice:**	3-0-0
Prerequisites:	Basic Satistics	**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Recognize the characteristics of machine learning, binary classification
	CO2	Solve classification problems using concept learning and multiclass classification
	CO3	Apply Tree based and Linear learning models to real world problems
	CO4	Analyze Bayesian classifiers, Distance based classification and clustering algorithms

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M-Medium, H- High)	**PO 1**	**PO 2**	**PO 3**	**PO 4**	**PO 5**	**PO 6**	**PSO1**	**PSO2**
CO1	L	L	L	L			M	L
CO2	M	L	M	M			H	L
CO3	M	L	M	M			H	L
CO4	M	L	M	M			H	L

Course Content

UNIT I:

The ingredients of machine learning, **Tasks:** the problems that can be solved with machine learning, Looking for structure, Evaluating performance on a task, Models: the output of machine learning: Geometric models, Probabilistic models, Logical models, Grouping and grading, Features: the workhorses of machine learning, Two uses of features, Feature construction and transformation.

Binary classification and related tasks: Classification, Assessing classification performance, Visualizing classification performance ,Class probability estimation, Assessing Class probability estimates

UNIT II:

Beyond binary classification: Handling more than two classes, Multi class classification Multi class scores and probabilities, Regression, Unsupervised and descriptive learning, Predictive and descriptive clustering.

Concept learning: The hypothesis space, Least general generalization, Internal disjunction ,Paths through the hypothesis space, Most general consistent hypotheses, Closed concepts, Beyond conjunctive concepts

	UNIT III: **Tree models**: Decision trees, Ranking and probability estimation trees, Sensitivity to skewed class distributions **Linear models**: The least-squares method, nultivariate linear regression, regularized regression, using lesat-squres regression for classification, Support vector machines, Soft margin SVM **Case study 1**: Imlplement decision tree learning algorithm using iris data set for predicting the species of a given sample and plot the decision surface using paired features.
	UNIT IV: **Bayesian Learning**: Introduction, Bayes Theorem, Bayes Optimal Classifier, Gibbs Algorithm, Naïve Bayes Classifier, Learning to classify Text **Distance Based Models:** Ways of measuring diatance, Neighbours and exemplars, Nearest Neighbours classification, Distance based clustering, k means algorithm, Clustering around mediods, Silhouettes, Hierarchical Clutsering **Case study 2**: Implement kmeans clustering on iris data set to group the samples automatically, without 'training' the algorithm. https://constantgeeks.com/2017/01/11/playing-with-iris-data-kmeans-clustering-in-python/
Text books and Reference books	**Text Book(s):** **[1]** Machine Learning: The art and Science of algorithms that make sense of data, Peter Flach, Cambridge University Press, 2012 **[2]** Tom M. Mitchell, Machine Learning, India Edition 2013, McGraw Hill Education **[3]** Chris Albon : Machine Learning with Python Cookbook , O'Reilly Media, Inc.2018 **Reference Books:** **[1]** Stephen Marsland, "Machine Learning – An Algorithmic Perspective", Second Edition, Chapman and Hall/CRC Machine Learning and Pattern Recognition Series, 2014 **[2]** EthemAlpaydın, Introduction to machine learning, second edition, MIT press. **[3]** T. Hastie, R. Tibshirani and J. Friedman, "Elements of Statistical Learning", Springer Series , 2nd edition
E-resources and other digital material	**[1]** Kevin Murphy,"MachineLearning:AProbabilisticPerspective" , MIT Press, 2012, https://www.cs.ubc.ca/~murphyk/MLbook/pml-intro-5nov11.pdf **[2]** Professor S. Sarkar IIT Kharagpur "Introduction to machine learning" https://www.youtube.com/playlist?list=PLYihddLF-CgYuWNL55Wg8ALkm6u8U7gps **[3]** Professor Carl Gustaf Jansson, KTH, Video Course on Machine Learning https://nptel.ac.in/noc/individual_course.php?id=noc19-cs35 **[4]** Tom Mitchell, "Machine Learning", http://www.cs.cmu.edu/~tom/10701_sp11/lectures.shtml

19ITDS1003 - ADVANCED ALGORITHMS

Course Category:	Programme Core	**Credits:**	3
Course Type:	Theory	**Lecture-Tutorial-Practice:**	3-0-0
Prerequisites:	Any programming language	**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Analyze the Performance of algorithms using Time and Space complexities.
	CO2	Analyse operations on various types of tree data structures
	CO3	Understand graph datastructure and its operations
	CO4	Identify data structures suitable to solve novel problems.

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M-Medium, H- High)	**PO 1**	**PO 2**	**PO 3**	**PO 4**	**PO 5**	**PO 6**	**PSO 1**	**PSO 2**
CO1	L		M	L			L	
CO2	L							M
CO3	M		L			L	H	
CO4	H		L		M		M	L

Course Content	**UNIT I:** **Performance analysis:** Performance, timecomplexity **Asymptotic notation:** Introduction, Asymptotic notation, Asymptotic mathematics, complexity analysis examples. **Performance measurement:** Introduction, choosing instance size, developing the test data. **Binary and other trees:** Trees, binary tree, properties of binary tree
	UNIT II: **Priority queues:** definition and application, the abstract data type, heaps **Applications:** heapsort **Binary search trees:** definitions, abstract data types, binary search tree operations and implementation **Applications:** histogramming **Balanced search trees:** **AVL Tree:** Definition,Height of an AVL Tree, Representation of an AVL Tree, Searching an AVL Search Tree, Inserting into an AVL Search Tree, Deletion from an AVL Search Tree.
	UNIT III: **Red-black Trees:** Definiton, Representation of a Red-Black Tree, Searching a Red-Black Tree, Inserting into a Red-Black Tree, Deletion from a Red-Black

	Tree, Implementation Considerations and Complexity. **B-trees:** M-Way Search Trees, B-Trees of Order m, Height of a B-Tree, Searching a B-Tree, Inserting into a B-Tree, Deletion from a B-Tree, Node Structure.
	UNIT IV: Graph algorithms: Definitions,Topological sort, shortest-path algorithms,Introduction to NP-Completeness.
Text books and Reference books	**Text Book(s):** **[1]** Sartaj Sahni, ―Datastructures, algorithms and applications in C++, University Press(India)Pvt. Ltd, 2 Edition 2005. **[2]** Mark Allen Weiss –"Data Structures and Algorithm Analysis in C++", Pearson, 4th edition 2014. **Reference Books:** **[1]** S.K.Basu, ―Design Methods and Analysis of Algorithms, PHI Learning Private Limited, New Delhi, 2008 **[2]** T.H.Cormen, et al, ―Introduction to Algorithms, 2 ed, PHI Pvt. Ltd. / Pearson Education, 2001.
E-resources and other digital material	**[1]** Prof. Abhiram Ramade, (03, 05, 2018).Computer Science Department, IIT-Bombay, http://nptel.ac.in/courses/106101060/ **[2]** Prof.Tim Roughgarden, (03, 05, 2018). Kleinberg and Tardos, Algorithm Design, 2015, http://openclassroom.stanford.edu /MainFolder/ CoursePage.php? course=IntroToAlgorithms **[3]** Sudarshan Iyengar: IIT Ropar (12, August, 2018). Data Structures and Algorithms[NPTEL]. http://nptel.ac.in/

19ITDS1014A - CLOUD COMPUTING AND VIRTUALIZATION

Course Category:	Programme Elective I	**Credits:**	3
Course Type:	Theory	**Lecture-Tutorial-Practice:**	3-0-0
Prerequisites:	Computer Networks	**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Understand the basics of cloud computing and its services
	CO2	Analyze the cloud architecture and the technologies driving virtualization
	CO3	Explore the functioning of different cloud platforms and their applications
	CO4	Identify the need of security in cloud and its mechanisms to manage the cloud environment

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M-Medium, H- High)

	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1	L						L	H
CO2	M			L	H		L	H
CO3	M		L				L	H
CO4	M					H	L	H

Course Content

UNIT I:
Fundamental Cloud Computing and Models
Understanding Cloud Computing: Origin and influences, Basic concepts and terminology, Goals and benefits, Risks and challenges.
Fundamental concepts and Models : Roles and boundaries, Cloud characteristics, Cloud Delivery models, Cloud deployment models.

UNIT-II:
Cloud Computing Architecture and Virtualization
Cloud Computing Architecture: Introduction, The cloud reference model: Architecture, Infrastructure-and hardware-as-a-service, Platform as a service, Software as a service.
Virtualization : Introduction, Characteristics of Virtualized Environments, Taxonomy of Virtualization Techniques -Execution Virtualization, Other types of Virtualization, Virtualization and Cloud Computing, Pros and Cons of Virtualization, Technology examples – VMware, Microsoft Hyper-V.

UNIT III:
Cloud Technologies and Eco-Systems
Building Cloud Computing Environments: Application development, Infrastructure and system development, Computing platforms and technologies.
Working with Clouds : **Cloud delivery Models : The cloud provider**

	perspective : Building IaaS Environments, Equipping PaaS Environments, Optimizing SaaS Environments, **Cloud Consumer perspective :** Working with IaaS Environments, Working with PaaS Environments, Working with SaaS Environments.
	UNIT-IV: **Cloud Management and Security Mechanisms** **Cloud management Mechanisms :** Remote Administration System, Resource Management System and SLA Management System. **Fundamental Cloud Security:** Basic Terms and Concepts, Threat Agents, Cloud Security Threats, **Cloud Security Mechanisms -** Encryption, Hashing, Digital Signature, IAM, SSO. **Case studies :** **Amazon web services** - Compute services, Storage services. **Google AppEngine** - Architecture and core concepts.
Text books and Reference books	**Text Book(s):** **[1]** Thomas Erl and RicardoPuttini Cloud Computing-Concepts, Technology and Architecture, Pearson, 2013. **[2]** Rajkumar Buyya, Christian Vecchiola, S Tamarai Selvi "Mastering Cloud Computing Foundations And Applications Programming" , McGraw Hill Education, 2016. **Reference Books :** **[1]** Ivanka Menken and Gerard Blokdijk, Cloud Computing Virtualization Specialist Complete Certification Kit-Study GuideBook, Lightning Source, 2009 **[2]** Barrie Sosinsky, "Cloud Computing Bible", Wiley Publishers, 2012 **[3]** John W. Rittenhouse and James F. Ransome, Cloud Computing Implementation, Management and Security,CRC Press, Taylor& Francis Group, 2010.
E-resources and other digital material	**[1]** Sanjay Pathak, "Cloud Concepts", https://www.youtube.com/watch?v=vv16c3BazSs **[2]** Edureka, "Salesforce Training", https://www.youtube.com/watch?v=9Gsmiff27do **[3]** MaciejArkit,"GoogleAppEngine", https://www.youtube.com/watch?v=UBa4ZsEAvP4

19ITDS2014 B - R FOR DATA SCIENCE

Course Category:	Programme Elective I	**Credits:**	3
Course Type:	Theory	**Lecture-Tutorial-Practice:**	3-0-0
Prerequisites:	Any programming language and Basic Mathematics	**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Understand the semantics, data handling and control statements in R.
	CO2	Analyze the libraries for data manipulation and conduct hypothesis tests for statistical inference.
	CO3	Synthesize data to fit linear and nonlinear models.
	CO4	Implement clustering, optimization and data visualization using R.

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M- Medium, H- High)

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PSO 1	PSO 2
CO1	L		M	H			M	
CO2	M			L		L	M	
CO3	M			L	M	L		M
CO4	H		L	M	H	M	H	H

Course Content

UNIT I:
Introduction and Basics of R: Basic Math, Variables, Data Types, Vectors, Missing Data.
Advanced Data Structures: data.frame, Lists, Matrices, and Arrays.
Reading Data into R: Reading CSV's, Excel Data, Reading from Databases, and Extract Data from Web Sites.
Control Structures & Loops: if and else, switch, if else, compound tests, for loops, while loops, controlling loops.

UNIT II:
Group Manipulation: Apply Family, aggregate, plyr and data.table.
Data Reshaping: cbind, rbind, joins and reshape2.
Manipulating Strings: paste, sprint, extracting text and regular expressions.
Probability Distributions: Normal Distribution, Binomial Distribution, Poisson distribution.
Basic Statistics: Summary Statistics, Correlation and Covariance, T-Tests and ANOVA.

UNIT III:
Linear Models: Simple Linear Regression and Multiple Regressions.
Generalized Linear Models: Logistic Regression, Poisson Regression.
Model Diagnostics: Residuals, Comparing Models, Cross-Validation, Bootstrap and Stepwise Variable Selection.

	Case Study: **Popularity Contest:** Develop a test to compare two different Twitter topics to see which one is most popular(or at least which one has a higher posting rate)
	UNIT IV: **Non-Linear Models:** Nonlinear Least Squares, Splines, Generalized Additive Models, Decision Trees and Random Forests. **Clustering:** K-means, PAM and Hierarchical Clustering. **Plots:** Base Graphics and ggplot2. **Case Study:** **String Theory:** To focus on manipulating unstructured data, which in most cases means natural language texts. Tweets are again a useful source of data for this because tweets are mainly a short (140 characters or less) character strings.
Text books and Reference books	**Text Book(s):** **[1]** Jared P. Lander, R for Everyone, Addison Wesley Data & Analytics Series, Pearson, 2014. **[2]** Jeffrey Stanton, An Introduction to Data Science, 2012. **Reference Books:** **[1]** G. Jay Kerns, Introduction to Probability and Statistics using R, First Edition, 2010 **[2]** Peter Dalgaard, Introductory Statistics with R, Springer, Second Edition, 2008
E-resources and other digital material	**[1]** Rafael Irizarry, Michael Love, Statistics with R, Harvard University (18, 04, 2018). https://www.edx.org/course/statistics-r-harvardx-ph525-1x-1 **[2]** Mine Çetinkaya-Rundel, David Banks, Colin Rundel, Merlise A Clyde, Duke University, (18, 04, 2018). Statistics with R Specialization. https://www.coursera.org/specializations/statistics

19ITDS1015A - SOCIAL AND INFORMATION NETWORK ANALYSIS

Course Category:	Programme Elective II	**Credits:**	3
Course Type:	Theory	**Lecture-Tutorial-Practice:**	3-0-0
Prerequisites:	-	**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Understand the basic notation and terminology used in social network
	CO2	Analyze the structure and balance of the social network
	CO3	Derive the similarities of people in the society and find the communities in the society.
	CO4	Apply link analysis and web search techniques for a given web application and generate recommendations.

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M-Medium, H- High)

	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1			M		L		L	
CO2	M		H		L		M	L
CO3	M		M	M	L		L	M
CO4	M		M	M	L		M	M

Course Content

UNIT I:
Overview: Introduction to Social Network Analysis.
Graph Essentials: Graph basic definitions, Paths and connectivity, Distance and Breadth-first search, Network Datasets.
Strong and Weak Ties : Triadic closure, The strength of weak Ties, Tie strength and network structures in large scale data, tie strength, social media and passive engagement, closure, structural holes and social capital.

UNIT II:
Networks in Their Surrounding Contexts: Homophily, Mechanisms Underlying Homophily: Selection and Social Influence, Affiliation.
Positive and Negative Relationships: Structural Balance, Characterizing the Structure of Balanced Networks, Applications of Structural Balance, A weaker form of Structural balance.

UNIT III:
Community analysis: Community Detection, Node degree, Node Reachability, Social Communities, Community Detection Algorithms, Member Based Community Detection
Group Based Community Detection Algorithms: Balanced Communities, Robust Communities, Modular Communities, Dense Communities and Hierarchical communities.
Case Study: Find the communities in the society using social network data.

	UNIT IV: **Information Networks and World Wide Web:** The structure of the web: The World Wide Web, Information Networks, Hypertext, and Associative Memory **Link Analysis and Web Search:** Searching the Web: The Problem of Ranking, Link Analysis Using Hubs and Authorities, PageRank. **Recommendations:** Recommendation System challenges, classical recommendation algorithms, Recommendation using social context. **Case study:** Develop a system to generate recommendations for the given data using social network data.
Text books and Reference books	**Text Book(s):** **[1]** "Networks, Crowds, and Markets Reasoning about a Highly Connected World", David Easley, Cornell University, New York, Jon Kleinberg, Cornell University, New York, 2010. **[2]** Reza Zafarani, Mohammad Ali Abbasi , Huan Liu Social Media Mining: An Introduction **Reference Books:** **[1]** Charu c. aggarwal "Social network data analytics" Springer **[2]** M. E. J. Newman Hardback "Networks: An Introduction by M. E. J. Newman, a college-level textbook about the science of networks.", Oxford University Press, 2010.
E-resources and other digital material	**[1]** Dr Bernie Hogan Social network analysis - Introduction to structural thinking:, University of Oxford, 2018. https://www.youtube.com/watch?v=2zhuj8ubinm **[2]** S.R.S.Lyengar "Introduction to Social Networks, 2017. https://www.youtube.com/watch?v=b7Ug1h6EGNk

19ITDS1015B - OPTIMIZATION TECHNIQUES FOR DATA ANALYSIS

Course Category:	Programme Elective II	**Credits:**	3
Course Type:	Theory	**Lecture-Tutorial-Practice:**	3-0-0
Prerequisites:	-	**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Understand the concept of optimality criteria for various types of optimization problem
	CO2	Analyze optimization algorithms for Linear Programming
	CO3	Solve various constrained and unconstrained nonlinear programmingproblems
	CO4	Apply the modern optimization methods to provide optimal solution for a given problem.

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M-Medium, H- High)	**PO1**	**PO 2**	**PO 3**	**PO4**	**PO5**	**PO6**	**PSO 1**	**PSO2**
CO1	M			L		L	L	
CO2	M		L					L
CO3	M					L	L	
CO4	H		L	L			L	L

Course Content	**UNIT I:** **Introduction to Optimization:** Introduction, Historical Development, Engineering Applications of Optimization, Statement of an Optimization Problem, Classification of Optimization Problems. **Classical Optimization Techniques:** Single-Variable Optimization, Multivariable Optimization with No Constraints, Multivariable Optimization with Equality Constraints, Multivariable Optimization with Inequality Constraints.
	UNIT II: **Linear Programming** Introduction, Applications of Linear Programming, Standard Form of a Linear Programming Problem, Geometry of Linear Programming Problems, Solution of a System of Linear Simultaneous Equations, Pivotal Reduction of a General System of Equations **Simplex Method:**Motivation of the Simplex Method, Simplex Algorithm, Improving a Nonoptimal Basic Feasible Solution, Two Phases of the Simplex Method

	UNIT III: **Nonlinear Programming Algorithms**: Unconstrained Algorithms – Direct Search Method, Gradient method, Constrained Algorithms - Separable Programming, Quadratic Programming, Chance- Constrained Programming, Linear Combinations method, SUMT Algorithm. **Case Study 1**: Chance Constrained Problem
	UNIT IV: **Modern Methods of Optimization** Introduction, Genetic Algorithms, Simulated Annealing, Particle Swarm Optimization, Ant Colony Optimization, Optimization of Fuzzy Systems, Neural-Network-Based Optimization **Case Study 2:**Travelling Salesperson Problem
Text books and Reference books	**Text books:** **[1]** Singiresu S Rao, "Engineering Optimization Theory and Practice", John Wiley and sons, 4th Edition, 2009. **[2]** HamdyA.Taha, "OprationResearch : An Introduction", 8th Edition, Pearson Prentice Hall, 2007. **[3]** Paulo Cortez, "Modern Optimization with R", Springer series, 2014. **Reference Books:** **[1]** S.Rao, "Engineering optimization: Theory and practice", 4th Edition, New Age International, 2009. **[2]** Edwin K. P. Chong and Stanislaw. Zak "An Introduction to Optimization", John Wiley and sons, 2nd Edition 2001. **[3]** Andreas Antoniou, "Practical Optimization Algorithms and Engineering Applications", **[4]** An Introduction to Optimization Edwin K., P. Chong & Stanislaw h. Zak. **[5]** Andreas Antoniou. "Practical Optimization Algorithms and Engineering Applications", Springer Series, 2007.
E-resources and other digital material	**[1]** Prof.A.Goswami, Department of Mathematics, IIT Kharagpur, "Optimization", 2014 https://nptel.ac.in/courses/111105039/ **[2]** Dr.AdityaJagannadham, IIT Kanpur, "Applied Optimization for Wireless, Machine Learning and Big Data, 2018 https://nptel.ac.in/courses/108104112/

19MTMC1026 – RESEARCH METHODOLOGY AND IPR

Course Category:	Mandatory learning Course	**Credits:**	0
Course Type:	Theory	**Lecture-Tutorial-Practice:**	2-0-0
Prerequisites:	-	**Internal Assessment:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Acquire an overview of the research methodology andtechniquesto define research problem
	CO2	Review the literature and identify the problem.
	CO3	Analyze the optimum sampling techniques for collected data.
	CO4	Apply various forms of the intellectual properties for research work.

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M-Medium, H- High)	**PO1**	**PO 2**	**PO 3**	**PO4**	**PO5**	**PO6**	**PSO 1**	**PSO2**
CO1	H	M	H	H	H	M	L	M
CO2	H	L	M	H	M	M	M	L
CO3	M	H	L	M	L	H	H	M
CO4	L	M	H	L	H	M	M	M

Course Content	**UNIT I:** **Research Methodology**: Introduction, Meaning of Research, Objectives of Research, Motivation in Research, Research Approaches, Significance of Research, Research and Scientific Methods, Research Process, Criteria of Good Research, and Problems Encountered by Researchers in India. **Research Problem:**Defining theResearch Problem, Selecting the Problem, Necessity of Defining the Problem, Technique Involved in Defining a Problem, an Illustration.
	UNIT II: **Reviewing the literature**: Place of the literature review in research, improving research methodology, broadening knowledge base in research area, enabling contextual findings. **Research Design**: Meaning of Research Design, Need for Research Design, Features of a Good Design, Important Concepts Relating to Research Design, Basic Principles of experimental Designs, Important Experimental Designs.

	UNIT III: **Design of Sampling**: Introduction, Sample Design, Sampling and Non-sampling Errors, Sample Survey versus Census Survey, Measurement and Scaling: Qualitative and Quantitative Data, Classifications of Measurement Scales, Goodness of Measurement Scales, sources of error in measurement tools. **Data Collection**: Experimental and Surveys, Collection of Primary Data, Collection of Secondary Data, Selection of Appropriate Method for Data Collection, Case Study Method **Testing of Hypotheses:** Hypothesis, Basic Concepts, Testing of Hypothesis, Test Statistics and Critical Region, Critical Value and Decision Rule, Procedure for Hypothesis Testing.
	UNIT IV: **Interpretation and Report Writing**: Meaning of Interpretation, Technique of Interpretation, Precaution in Interpretation, and Significance of Report Writing **Intellectual Property:** The Concept, Intellectual Property System in India, Development of TRIPS Complied Regime in India, Patents Act, 1970, Trade Mark Act, 1999, The Designs Act, 2000, The Geographical Indications of Goods (Registration and Protection) Act1999, Copyright Act, 1957, Trade Secrets, Utility Models WTO, Paris Convention for the Protection of Industrial Property, National Treatment, Right of Priority, Common Rules, Patents, Marks, Industrial Designs, Trade Names, Indications of Source, Unfair Competition, Patent Cooperation Treaty (PCT), Trade Related Aspects of Intellectual Property Rights(TRIPS) Agreement.
Text books and Reference books	**Text books:** **[1]** Research methodology: Methods and Techniques, C.R. Kothari, GauravGarg, New Age International, 4th Edition, 2018. **[2]** Research Methodology a step-by-step guide for beginners.Ranjit Kumar, SAGE Publications Ltd.,3rd Edition, 2011 **[3]** Study Material, Professional Programme Intellectual Property Rights, Law and Practice, The Institute of Company Secretaries of India, Statutory Body under an Act of Parliament, September 2013. **Reference Books:** **[1]** An introduction to Research Methodology, Garg B.L et al ,RBSA Publishers 2002 **[2]** An Introduction to Multivariate Statistical Analysis Anderson T.W,Wiley 3rd Edition **[3]** Research Methodology, Sinha, S.C, Dhiman,EssEss Publications2002 **[4]** Research Methods: the concise knowledge base ,Trochim ,Atomic Dog Publishing, 2005 **[5]** How to Write and Publish a Scientific Paper, Day R.A,Cambridge University Press, 1992 **[6]** Conducting Research Literature Reviews: From the Internet to Paper,Fink A,Sage Publications,2009 **[7]** Proposal Writing,Coley S.M. Scheinberg, C.A,Sage Publications,1990 **[8]** Intellectual Property Rights in the Global Economy, Keith Eugene Maskus,Institute for International Economics.
E-resources and other digital material	**[1]** Prabuddhaganguli, Intellectual property right (1stedition) [English]. http://www.slideshare.net/harshhanu/intellectual-property-rights-13551183 **[2]** U.S Government Printing office, 1986 http://www.e-booksdirectory.com/details.php?ebook=10758

19ITDS1051 - MACHINE LEARNING LAB

Course Category:	Programme Core	**Credits:**	1.5
Course Type:	Practical	**Lecture-Tutorial-Practice:**	0-0-3
Prerequisites:		**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Implement classification problems with decision trees, support vectors
	CO2	Demonstrate Neural network, genetic algorithms
	CO3	Apply different Bayesian learning techniques
	CO4	Solve distance based supervised and unsupervised learning problems

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M-Medium, H- High)	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PSO1**	**PSO2**
CO1	M		M	M			H	L
CO2	M		M	M			H	L
CO3	M		M	M			H	L
CO4	M		M	M			H	L

Course Content

Week 1:
For a given set of training examples stored in .CSV format,
1. Write some hypotheses possible for concept learning in hypothesis space H
2. Implement and demonstrate Find-S algorithm to display the most specific hypothesis

Week 2:
For the training examples considered in task of week1 stored in .CSV format,
implement and demonstrate candidate elimination to display the version space

Week 3:
Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.

Week 4:
Write a program to implement multiple linear regression. Use an appropriate data set and show the relationship between dependent and independent variables.

Week 5:
Build an Artificial Neural Network using sigmoid function by implementing the Back propagation algorithm and test the same using appropriate data sets.

	Week 6: Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.
	Week 7: Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set.
	Week 8: Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering. You can add Java/Python ML library classes/API in the program.
	Week 9: Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions. Java/Python ML library classes can be used for this problem
	Week 10: Implement Kernel SVM for non linear classification to design XOR logic gate
	Week 11,12: **Case Study 1:** Consider telecommunications dataset provided by IBM, do the customer segmentation and predict the group of customers who are going to churn using k-means clustering. **Case Study 2**: Process automation in finance can be of chatbots, gamification of employee training and paper automation. Apply ID3 classification technique to predict the nature of any given financial ecosystem. **Case Study 3:** Recommend a personalized insurance plan to a particular customer or user by applying association rule mining **Case Study 4:** Use KDD cup dataset, split the data into training and test datasets, build the classification model using C4.5 to predict the respondents. **Case Study 5:** Consider movie review dataset and classify positive and negative reviews using Naïve Bayes algorithm. **Case study 6:** Consider women crime data of different region of India and predict the future crime rate of a given state using SVM.
Text Books and Reference Books	**Text Book(s):** **[1]** Stephen Marsland, "Machine Learning – An Algorithmic Perspective" Second Edition, Chapman and Hall/CRC Machine Learning and Pattern Recognition Series, 2014 **Reference Books:** **[1]** Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, " An Introduction to Statistical Learning with Applications in R", Springer texts in series 2014

	[2] Tom M. Mitchell, Machine Learning, India Edition 2013, McGraw Hill Education
E-resources and other digital material	**[1]** Professor S. Sarkar IIT Kharagpur "Introduction to machine learning" , https://www.youtube.com/playlist?list=PLYihddLF-CgYuWNL55Wg8ALkm6u8U7gps **[2]** Professor Carl Gustaf Jansson, KTH, Video Course on Machine Learning https://nptel.ac.in/noc/individual_course.php?id=noc19-cs35 **[3]** Tom Mitchell, "Machine Learning", http://www.cs.cmu.edu/~tom/10701_sp11/lectures.shtml **[4]** Barber, David. "Machine learning a probabilistic approach." (2006). https://pdfs.semanticscholar.org/7bc7/54bc548f32b9ac53df67e3171e8e4df66d15.pdf

19ITDS1052 - PYTHON FOR DATASCIENCE LAB

Course Category:	Programme Core	**Credits:**	1.5
Course Type:	Practical	**Lecture-Tutorial-Practice:**	0-0-3
Prerequisites:		**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Implement python programming constructs to build small to large scale applications
	CO2	Manipulate one-dimensional and multi -dimensional numpy arrays, and pandas series and data frames
	CO3	Perform data loading, cleaning ,transformation and merging
	CO4	Create different plots for basic exploratory data analysis

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M-Medium, H- High)

	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1	M		M				L	L
CO2	M		M				L	M
CO3	M		M			M	M	M
CO4	M		M		H	M	H	H

Course Content

Week 1:

1. Creation of Python scripts that uses Operators, Controlflow statements
2. Create Python Script that uses functions with various types of arguments such as default arguments, keyword argument and variable lengtharguments

Week 2:

1. Create python script to implement fruitful functions and void functions
2. Simulate the calculator application to perform all the operations to be implemented onit.
3. Write a function that takes an ordered list of numbers (a list where the elements are in order from smallest to largest) and another number. The function decides whether or not the given number is inside the list and returns (then prints) an appropriateboolean

Week 3:

1. Creation of python programs on the modules Numpy
2. Analyse the given series of data using pandas
3. Python programs that uses the dictionaries, tuples and other data structures

Week 4:

1. Python Program on Text Files reading, handling and manipulation on these files
2. Creation of python programs to handle missing data and hierarchical indexing, data aggregation on multi indexes

	Week 5: 1. Python scripts that access the data from a given database.
	Week 6: 1. Creation of Python forms for the department library/Lab/attendance etc., by entering student details of each student . Validate the form using Python validators and display error message
	Week 7: 1. Python programs on data transformation and string manipulation. 2. Python Programs to simulate Queue Operations 3. Implement the data structure of binary search trees, using classes, with operations for inserting and finding an element
	Week 8: 1. Python programs on Scatter plots with histograms and a Scatter plot matrix for a given data 2. Find the root words of the given list of words using Porter and Snowball Stemming 3. Perform tokenization and parts of speech tagging for the given sentence
	Week 9 & 10: 1. Malicious URL is one of the dangerous threats to the web users in today‟s world and cyber security. These URL‟s are mainly used by the attackers and hackers to steal our valuable information like monetory loss, stealing of private information, and installation of of malware. As a python data analyst, develop a suitable algorithm to detect malicious URL from a given set of URL‟s. 2. Implement depth first search traversal for a graph which contains 6 vertices. Keep the elements in the stack, the lower order number first. Also check the traversal if you keep the higher order number first. 3. Apply classification/clustering on a given remotely sensed data. Use python language to extract the relevant features and implement a classification/clustering algorithm. 4. Create an application that simulates bank operations by using the python framework. 5. Analyze the sentiments on a given topic from the data available in social media. Identify the appropriate features and from there analyze the sentiments through a classification algorithm 6. Develop a python application that will detect objects in a given video. 7. Develop a python algorithm that detects a fraud in banking Transactions. This project aims at classify a given transaction is fraud or not from a universally accepted dataset by extracting relevant features and with a classifier. 8. Develop a python application that extracts the information on Land Cover and Land Use from a satellite image. 9. Design a python framework to extract the meaningful information from the images available from Bhuvan.
	Week 11 & 12: 1. Implement Binary search tree operations using python a. Create Binary search Tree b. Searching

	c. Inserting into an Binary search Tree d. Deletion from an Binary search Tree 2. Implement AVL tree operations using python a. Create AVL Tree b. Searching c. Inserting into an AVL Tree d. Deletion from an AVL Tree 3. Implement Red-Black tree operations using python a. Create Red-Black Tree b. Searching c. Inserting into an Red-Black Tree d. Deletion from an Red-Black Tree 4. Implement B tree operations using python a. Create B Tree b. Searching c. Inserting into an B Tree d. Deletion from an B Tree
Text Book(s) and Reference Books	**Text Book(s):** **[1]** Jeffrey Stanton, Syracuse University , An Introduction to Data Science **[2]** VamsiKurama, "Python Programming: A Modern Approach", Pearson India, 2017 **[3]** Wes McKinney , "Python for Data Analysis" ,OReilly Media Inc. 2013 Samir Madhavan, "Mastering Python for Data Science", PACKT publishing, 2015 **Reference Books:** **[1]** Jake Vanderplas, "Python Datascience Handbook" , OReilly Media Inc 2017 **[2]** Joel Grus, "Data science from Scratch, First Prniciples with Python ", OReilly Media Inc 2015
E Resources	**[1]** Charles Severance: University of Michigan,Python for Everybody [COURSERA]. https://www.coursera.org/ **[2]** MadhavanMukund, (12, may, 2018). Programming, Data Structures & Algorithms using Python [NPTEL]. http://nptel.ac.in/ **[3]** Keith Galli Complete Python NumPy Tutorial (Creating Arrays, Indexing, Math, Statistics, Reshaping) https://www.youtube.com/watch?v=GB9ByFAIAH4 **[4]** Keith Galli Complete Python Pandas Data Science Tutorial! (Reading CSV/Excel files, Sorting, Filtering, Groupby) https://www.youtube.com/watch?v=vmEHCJofslg **[5]** CS Dojo, Intro to Data Analysis / Visualization with Python, Matplotlib and Pandas | Matplotlib Tutorial https://www.youtube.com/watch?v=a9UrKTVEeZA

SEMESTER II

19ITDS2001 - DATA VISUALIZATION

Course Category:	Programme Core	**Credits:**	3
Course Type:	Theory	**Lecture-Tutorial-Practice:**	3-0-0
Prerequisites:		**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Comprehend the importance of the exploratory data analysis paradigm
	CO2	Understand basic concepts of data visualization
	CO3	Select appropriate data visualization technique for given data
	CO4	Design visualizations for presenting stories from data

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M-Medium, H- High)		**PO1**	**PO 2**	**PO 3**	**PO4**	**PO5**	**PO6**	**PSO 1**	**PSO2**
	CO1			L		L			L
	CO2	L		L		L		L	H
	CO3	L	H			H			
	CO4	H	H	H		H	L	H	

Course Content

UNIT I:

Exploratory data analysis (EDA): Typical data format and the types of EDA,Univariate non-graphical EDA,Categorical data ,Characteristics of quantitative ,Central tendency,Spread,Skewness and kurtosis

Univariate graphical EDA: Histograms, stem-and-leaf plots, boxplots, Qunatile-normal plots.

Multivariate non-graphical EDA: Cross-tabulation, Correlation for categorical data, Univariate statistics by category, Correlation and covariance, Covariance and correlation matrices.

Multivariate graphical EDA:Univariate graphs by category,Scatterplots

UNIT II:

The Context of Data Visualization : Visualization as a discovery tool, The bedrock of visualization knowledge, Defining data visualization, Visualization skills for the masses

The data visualization methodology**.**

Setting the Purpose and Identifying Key Factors: Establishing intent – the visualization's function, Establishing intent – the visualization's tone, Key factors surrounding a visualization project, The " eight hats" of data visualization design

	UNIT III: **Conceiving and Reasoning Visualization Design Options:** Data visualization design is all about choices, The visualization anatomy – data representation, The visualization anatomy – data presentation **Taxonomy of Data Visualization Methods:** Data visualization methods, Choosing the appropriate chart type, Assessing hierarchies and part-to-whole relationships.
	UNIT IV: **Constructing and Evaluating Your Design Solution:** For constructing visualizations, technology matters, The construction process, Approaching the finishing line, Post-launch evaluation, Developing your capabilities, mapping geo-spatial data
Text books and Reference books	**Text Book(s):** **[1]** Howard J. Seltman," Experimental Design andAnalysis",2018 **[2]** Andy Kirk, "Data Visualization: a successful design process", Packt Publishing,2012 **Reference Books:** **[1]** Claus O. Wilke," Fundamentals of Data Visualization A Primer on Making Informative and Compelling Figures", O'Reilly,2019, ISBN:978-1-492-03108-6 **[2]** Stephen Few, "Now you see it: Simple Visualization techniques for quantitative analysis", Analytics Press, 2009.
E-resources and other digital material	**[1]** Prof.ShankarNarasimhan,RagunathanRengasamy,IIT ,Data Visualization in R Basic graphics, Madras, https://nptel.ac.in/courses/106106179/11,2011 **[2]** Dr. Ed Vul,Dr. Mike Frank,Massachusetts Institute of Technology,Statistics and Visualization for Data Analysis and Inference, https://ocw.mit.edu/resources/res-9-0002-statistics-and-visualization-for-data-analysis-and-inference-january-iap-2009/.2009 **[3]** Python for Data Analysis - Python for Data Visualisation, https://www.youtube.com/watch?v=nXr2Xt52MfA,2017 **[4]** Python Data Visualization, https://www.coursera.org/learn/python-visualization **[5]** Data Visualization with Python and Matplotlib, https://www.udemy.com/data-visualization-with-python-and-matplotlib/,2018

19ITDS2002 - BIG DATA MANAGEMENT

Course Category:	Programme Core	**Credits:**	3
Course Type:	Theory	**Lecture-Tutorial-Practice:**	3-0-0
Prerequisites:	Concepts of Database Management	**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Understand The Fundamental Concepts Of Big Data and HDFS.
	CO2	Solve Big Data Problems Using Mapreduce, Pig And Hive.
	CO3	Use Nosql Databases To Process Different Varieties of Data.
	CO4	Perform In Memory Data Analytics With Spark and Spark Streaming.

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M- Medium, H- High)	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PSO1**	**PSO2**
CO1	M			M	L			H
CO2	L			M		L		L
CO3	L		L	M			L	M
CO4				L		H		H

Course Content

UNIT I:

Introduction to big data: Data, Types of digital data, Evolution and Definition of big data, Challenges of big data, Characteristics and Need of big data.

Introduction to Hadoop: Introducing Hadoop, need of Hadoop, limitations of RDBMS, RDBMS versus Hadoop, Distributed Computing Challenges, History of Hadoop, Hadoop Overview, Hadoop Distributors.

HDFS (Hadoop Distributed File System): HDFS Daemons, Anatomy of file read, Anatomy of file write, working with HDFS commands.

UNIT II:

Introduction to MAPREDUCE Programming: Processing Data with Hadoop, Managing Resources and Applications with Hadoop YARN (Yet another Resource Negotiator), Mapper, Reducer, Combiner, Partitioner, Searching, Sorting , Compression, Hadoop EcoSystem.

Introduction to Pig: Key Features of pig, The Anatomy of Pig, Pig on Hadoop , Pig Philosophy, Pig Latin Overview, Data Types in Pig, Running Pig, Execution Modes of Pig, Relational Operators.

UNIT III:

Introduction to HIVE: HIVE features, HIVE architecture, HIVE datatypes, HIVE File Formats, HIVE Query Language

NoSQL: Introduction to NOSQL, Types of NoSQL Databases, and

	Advantages of NoSQL databases, CAP Theorem, BASE, SQL versus NoSql.
	UNIT IV: **Spark**: Introduction to data analytics with Spark, Spark Stack, Programming with RDDS, Working with key/value pairs, Spark SQL, Schema RDD. **Sparking Streaming**: High level architecture of Spark Streaming, DStreams, Transformations on DStreams, Different Types of Transformations on DStreams.
Text books and Reference books	**Text Book(s):** **[1]** SeemaAcharya, SubhashiniChellappan, "Big Data and Analytics", Wiley Publishers, 2015, First Edition. **[2]** Holden Karau, Andy Konwinski, Patrick Wendell, MateiZaharia, "Learning Spark: Lightning-Fast Big Data Analysis", O'Reilly Media, Inc, 2015 First Edition. **Reference Books:** **[1]** Tomwhite, Hadoop, "Thedefinitiveguide", 3rdedition, O'reillypublications, 2012. **[2]** David Loshin, "BigDataAnalytics: From Strategic Planning to Enterprise IntegrationwithTools,Techniques,NoSQL,andGraph", MorganKaufmannPublishers, 2013. **[3]** Hadoop in PracticebyAlexHolmes, MANNING. **[4]** Dirk deRoos, Chris Eaton, George Lapis, Paul Zikopoulos, Tom Deutsch "Understanding Big Data Analytics for Enterprise ClassHadoopandStreamingData", 1st Edition, TMH,2012.
E-resources and other digital material	**[1]** Big Data Use cases for Beginners | Real Life Case Studies | Success Stories https://www.youtube.com/watch?v=HHR0-iJp2sM. **[2]** Alexey Grishchenko, Hadoopvs Mpp, https://0x0fff.Com/Hadoop-Vs-Mpp/ **[3]** Random Notes On Bigdata- Slideshare: Available Www.Slideshare.Net/Yiranpang/Random-Notes-On-Big-Data-26439474

19ITDS2003 - BUSINESS ANALYTICS

Course Category:	Programme Core	**Credits:**	3
Course Type:	Theory	**Lecture-Tutorial-Practice:**	3-0-0
Prerequisites:		**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Understand Business Analytics and manipulate data
	CO2	Analyze the fundamental tools and methods of data analysis and statistics.
	CO3	Develop approaches for applying forecasting techniques and data mining techniques.
	CO4	Identify, model and solve decision problems in different settings

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M-Medium, H- High)		**PO1**	**PO 2**	**PO 3**	**PO4**	**PO5**	**PO6**	**PSO 1**	**PSO2**
	CO1	L		M		M		H	
	CO2	L			M			L	
	CO3	M			M	M		M	M
	CO4	H		M	H	H	M	H	H

Course Content

UNIT I:
Foundations of Business Analytics
Introduction, Evolution of Business Analytics, Scope of Business Analytics, Data for Business Analytics,Models in Business Analytics, Problem Solving with Analytics.

UNIT II:
Descriptive Analytics:
Probability Distributions and Data Modeling: Basic concepts of probability,Random variables and probability distribution, Discrete Probability Distributions , Continuous Probability Distributions, Random Sampling from Probability Distributions, Data Modeling and Distribution Fitting.
Sampling and Estimation: Statistical Sampling, Sampling Error, Sampling Distributions.

UNIT III:
Predictive Analytics:
Forecasting Techniques - Qualitative and Judgmental Forecasting, Statistical Forecasting Models, Forecasting Models for Stationary Time Series, Forecasting Models for Time Series with a Linear Trend, Forecasting Time Series with Seasonality, Selecting Appropriate Time-Series-Based Forecasting Models , Regression Forecasting with Causal Variables.
Monte Carlo Simulation and Risk Analysis- Spreadsheet Models with Random Variables, Monte Carlo Simulation Using Analytic Solver Platform, New-Product Development Model, Newsvendor Model.

	UNIT IV: **Prescriptive Analytics:** **Applications of Linear Optimization-**Process Selection Models, Solver Output and Data Visualization, Blending Models, Portfolio Investment Models, Transportation Models. **Decision Analysis**: Formulating Decision Problems, Decision Strategies without Outcome Probabilities, and Decision Strategies with Outcome Probabilities, Decision trees, The value of information, Utility and decision making.
Text books and Reference books	**Text books:** **[1]** James Evans, "Business Analytics, Second Edition, Pearson Publications, 2017. **Reference Books:** **[1]** U. Dinesh Kumar, "Business Analytics - The Science of Data Driven Decision Making", First Edition, Wiley Publications, 2017. **[2]** SeemaAcharya R N Prasad, "Fundamentals of Business Analytics", 2nd Edition, Wiley Publications, 2016
E-resources and other digital material	**[1]** Cody Baldwin, "Introduction to Business Analytics", Feb 2016. https://www.youtube.com/channel/UCHPHVCq_Giziio_y8QEcHyA **[2]** ACADGILD, "Business Analytics for Beginners", 2016 https://www.youtube.com/watch?v=an9PXNtTSSc

19ITDS2014A - COMPUTER VISION

Course Category:	Programme Elective III	**Credits:**	3
Course Type:	Theory	**Lecture-Tutorial-Practice:**	3-0-0
Prerequisites:		**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Understand and master basic knowledge, theories and methods in computer vision
	CO2	Understand various feature extraction methods and its significance.
	CO3	Analyze various clustering and classification techniques.
	CO4	Understand and analyze Video Processing methods.

Contribution of Course Outcomes towards achievement of Program Outcomes(L-Low, M-Medium, H- High)

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PSO1	PSO2
CO1	H		M		L		L	
CO2	M		L					L
CO3						H	M	L
CO4	H		H		M		L	M

Course Content

UNIT I:

Introduction: Computer vision, A brief history.

Image formation: Geometric primitives and transformations**,** Photometric image formation

Image processing: Point operators, Linear filtering, More neighborhood operators, Fourier transforms, Pyramids and wavelets.

UNIT II:

Feature detection and matching:

Points and patches, Application: Performance-driven animation Edges, Application: Edge editing and enhancement, Lines, Application: Rectangle detection.

UNIT III:

Image Segmentation

Split and merge, Mean shift and mode finding, Normalized cuts, Graph cuts and energy-based methods, Application: Medical image segmentation.

Feature-based alignment:

Pose estimation, Application: Augmented reality

UNIT IV:

Dense motion estimation:

Parametric motion, Application: Video stabilization, Optical flow, Application: Video de-noising, Layered motion, Application: Frame interpolation.

Text books and Reference books	**Text Book(s):** **[4]** Richard Szeliski, Computer Vision: Algorithms and Applications, Springer-Verlag London Limited 2011. **Reference Books:** **[1]** Computer Vision: A Modern Approach, D. A. Forsyth, J. Ponce, Pearson Education, 2003. **[2]** K. Fukunaga; Introduction to Statistical Pattern Recognition, Second Edition, Academic Press, Morgan Kaufmann, 1990. **[3]** R.C. Gonzalez and R.E. Woods, Digital Image Processing, Addison-Wesley, 1992.
E-resources and other digital material	**[1]** Dr. Mubarak Shah, (13, 08, 2019). UCF Computer Vision Video Lectures,https://www.youtube.com/watch?v=715uLCHt4jE&list=PLd3hlSJsX_ImKP68wfKZJVIPTd8Ie5u-9 **[2]** Lecture 1 | Introduction to Convolutional Neural Networks for Visual Recognition https://www.youtube.com/watch?v=vT1JzLTH4G4&list=PLf7L7Kg8_FNxHATtLwDceyh72QQL9pvpQ

19ITDS2014B - DEEP LEARNING

Course Category:	Programme Elective III	**Credits:**	3
Course Type:	Theory	**Lecture-Tutorial-Practice:**	3-0-0
Prerequisites:	Machine Learning	**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Understand basic concepts of neural networks and back propagation algorithm
	CO2	Analyze the layers in the architecture of convolution neural networks
	CO3	Acquire knowledge on auto encoders, word2vec architecture
	CO4	Explore deep learning models for sequence analysis

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M-Medium, H- High)

	PO1	**PO2**	**PO3**	**PO 4**	**PO 5**	**PO 6**	**PSO1**	**PSO2**
CO1			L	M			L	L
CO2	M		M	M			M	M
CO3	L		L	L			L	L
CO4	H		M	M			M	M

Course Content

UNIT I:

The Neural Network: Building Intelligent Machines , The Limits of Traditional Computer Programs , The Mechanics of Machine Learning , The Neuron , Expressing Linear Perceptrons as Neurons , Feed-Forward Neural Networks, Linear Neurons and Their Limitations , Sigmoid, Tanh, and ReLU, Softmax output layers

Training Feed-Forward Neural Networks: Gradient Descent , The Delta Rule and Learning Rates , Gradient Descent with Sigmoidal Neurons, The Backpropagation Algorithm , Stochastic and Minibatch Gradient Descent , Test Sets, Validation Sets, and Overfitting, Preventing Overfitting in Deep Neural Networks

UNIT II:

Convolutional Neural Networks: Neurons in Human Vision ,The Shortcomings of Feature Selection, Vanilla Deep Neural Networks, Filters and Feature Maps, Full Description of the Convolutional Layer , Max Pooling Full Architectural Description of Convolution Networks, Closing the Loop on MNIST with Convolutional Networks, Image Preprocessing Pipelines Enable More Robust Models , Accelerating Training with Batch Normalization

UNIT III:

Embedding and Representation Learning: Learning Lower-Dimensional Representations, Principal Component Analysis , Motivating the Autoencoder Architecture , Implementing an Autoencoder in TensorFlow, Denoising to Force Robust Representations, Sparsity in Autoencoders, The Word2Vec Framework , Implementing the Skip-Gram Architecture

	UNIT IV: **Sequence Modeling**: **Recurrent and Recursive nets:** Unfolding Computational Graphs, Recurrentneural networks, Bidirectional RNNS,Encoder-Decoder sequence-to –sequence architectures, Deep Recurrent networks, Recursive neural networks The Challenge of Long-Term Dependencies, Echo State Networks, Leaky Units & Otherstrategies for multiple timescales, The Long Short-Term memory and other Gated RNNs, Optimization for Long-Term Dependencies, Explicit Memory
Text books and Reference books	**Text Book(s):** **[1]** Nikhil Buduma, Nicholas Locascio, "Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms", O'Reilly Media, 2017 **[2]** Ian Goodfellow, YoshuaBengio, Aaron Courville, "Deep Learning(Adaptive Computation and Machine Learning series",MIT Press, 2017 **Reference Books:** **[1]** Li Deng and Dong Yu, "Deep learning Methods and Applications", Now publishers,2013 **[2]** Michael Nielsen,"Neural Networks and Deep Learning", Determination Press 2015
E-resources and other digital material	**[1]** Mitesh Khapra, "Deep Learning", Sep 20, 2018 https://www.youtube.com/watch?v=4TC5s_xNKSs&list=PLH-xYrxjfO2VsvyQXfBvsQsufAzvlqdg9 **[2]** AfshineAmidi and ShervineAmidi ,"Deep Learning cheatsheets for Stanford's CS 230", 2018, https://github.com/afshinea/stanford-cs-230-deep-learning **[3]** YoshuaBengio, Deep learning: "Theoretical Motivations,Canadian Institute for Advanced Research", 2015http://videolectures.net/deeplearning2015_bengio_theoretical_motivations/ **[4]** Geoffrey Hinton's GoogleTech Talk,"Recent developments on Deep Learning"March 2010, https://www.youtube.com/watch?v=VdIURAu1-aU

19ITDS2015A – NATURAL LANGUAGE PROCESSING

Course Category:	Programme Elective IV	**Credits:**	3
Course Type:	Theory	**Lecture-Tutorial-Practice:**	3-0-0
Prerequisites:	Machine Learning	**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Comprehend the concepts of natural language processing, its applications and language modeling techniques
	CO2	Evaluate probabilistic language models and Solve NLP sub problems using tokenizing and tagging
	CO3	Analyze linguistic structure in text, using parsing and CFG
	CO4	Interpret Methods to recognize syntactic and semantics structures of a sentence

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M- Medium, H- High)

	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1	L			L			M	L
CO2	H	M		H			M	L
CO3	H	M		M			M	L
CO4	H						M	L

Course Content

UNIT I:

Introduction – Models and Algorithms

Regular Expressions and Automata Regular Expressions, Finite State Automata, Regular Languages and FSAs

Words and Transducers: Survey of English Morphology, Finite-State Morphological Parsing, Construction of a finite State Lexicon ,Morphological parsing with FST, Transducers and orthographic rules, Combining an FST Lexicon and Rules.

UNIT II:

N-grams- Counting Words in Corpora, Unsmoothed N-grams, Training and Test sets, Smoothing,Backoff, Interpolation, Entropy-Cross entropy for comparing models.

Classes and Part-of-Speech Tagging- English Word Classes, Tag sets for English, Part of Speech Tagging, Rule-Based Part of Speech Tagging, HMM Part of Speech Tagging, Transformation-Based Tagging.

Hidden Markov and Maximum Entropy Models: Markov Chains, The hidden Markov Model.

	UNIT III: **Automatic Speech Recognition:** Speech Recognition Architecture, HMM applied to Speech. **Formal Grammars of English -** Constituency, Context-Free Grammars,Some Grammar Rules for English, Grammar equivalence and Normal form **Syntactic Parsing**– Parsing as Search, ambiguity, Search in the face of Ambiguity,The Earley Algorithm.
	UNIT IV: **Representing Meaning -** Computational Desiderata for Representaions, First Order Logic, Event and State Representations. **Computational Semantics-** Syntax Driven Semantic Analysis - Semantic augmentations to Syntactic rules. **Question Answering and Summarization:** Factoid Question Answering, Summarization.
Text books and Reference books	**Text Book(s):** **[1]** D.Jurafsky and J. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition", Second Edition, Pearson Education, 2009. **Reference Books:** **[1]** Manning and H. Schutze, "Foundations of Statistical Natural Language Processing", MIT Press,1999. **[2]** Nitin Indurkhya, Fred J. Damerau, "Handbook of Natural Language Processing", 2nd Edition, Chapman and Hall/CRC Press, 2010. **[3]** Tanveer Siddiqui, U.S. Tiwary, "Natural Language Processing and Information Retrieval", Oxford University Press, 2008.
E-resources and other digital material	**[1]** Dan Jurafsky and Christopher Manning, Natural Language Processig Course,Stanford,26th Jun2019, https://web.stanford.edu/~jurafsky/NLPCourseraSlides.html **[2]** Dan Jurafsky and Christopher Manning, Natural Language Processig Course,10th Jun2018, https://www.youtube.com/watch?v=3Dt_yh1mf_U&list=PLQiyVNMpDLKnZYBTUOlSI9mi9wAErFtFm **[3]** Prof.Sudeshna Sarkar and Prof.Anupam Basu, Lecture Series on Artificial Intelligence Department of CSE,I.I.T,Kharagpur, NPTEL,2008,Oct http://nptel.iitm.ac.in

19ITDS2015B - CYBER SECURITY

Course Category:	Programme Elective IV	**Credits:**	3
Course Type:	Theory	**Lecture-Tutorial-Practice:**	3-0-0
Prerequisites:	Computer Networks	**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Identify the assets of information and significance of security.
	CO2	Apply data leakage, protection and security policies on digital systems.
	CO3	Analyse log files and backup strategies for securing the data in real time environment.
	CO4	Implement the issues in handling web vulnerabilities.

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M-Medium, H- High)	**PO 1**	**PO 2**	**PO 3**	**PO 4**	**PO 5**	**PO 6**	**PSO 1**	**PSO 2**
CO1	L	M				M	L	L
CO2	L	M	L	L		M	L	L
CO3	L	M	L	L		M	L	L
CO4	L	L				M	L	L

Course Content

UNIT I: Information Security and Threats

Introduction – Information Security, Information Assets & Threats - Threats to Information Assets, Types of Attacks, Types of Virus, Types of Worms, Types of Trojans, Network Attacks, Common Vulnerabilities and Exposures (CVE).

Fundamentals of Information Security: Elements of information security – Network Security, Application Security, Communications Security. Principles and concepts – data security – Critical Information Characteristics, Information States, Prevention Vs Detection, Types of controls – Access Control Models.

UNIT II: Data Leakage and Prevention

Introduction to Data Leakage, Organizational Data Classification, Location and Pathways, Content Awareness, Content Analysis Techniques, Data Protection

Network Sniffers and Injectors – Sniffers Overview, Tcpdump, Windump, Wireshark, Ettercap, Hping.

UNIT III: Log Correlation and Management

Event Logs - Concepts, Log Management and its need, Log Management Process, IIS Log Files, Log Analysis and Response.

Data Backup: Data Backup -Overview, Types of Backup, Backup Procedures, Types of Storage.

	UNIT IV: **Web Application Hacking :** Scanning for web vulnerabilities : Nikto, HTTP utilities - Curl, OpenSSL, Stunnel, Application Inspection – Zed Attack Proxy, Sqlmap, Password Cracking and Brute-Force Tools.
Text books and Reference books	**Text Book(s):** **[1]** Student Handbook – Security Analyst, NASSCOM **[2]** Anti-Hacker Tool Kit (Indian Edition) by Mike Shema, Publication McGraw Hill **Reference Books:** **[1]** Cyber Security Understanding Cyber Crimes, Computer Forensics and Legal Perspectives by Nina Godbole and SunitBelpure, Publication Wiley **[2]** Nelson Phillips and EnfingerSteuart, "Computer Forensics and Investigations", Cengage Learning, New Delhi, 2009. **[3]** Robert M Slade," Software Forensics", Tata McGraw - Hill, New Delhi, 2005 **[4]** Kevin Mandia, Chris Prosise, Matt Pepe, "Incident Response and Computer Forensics ", Tata McGraw -Hill, New Delhi, 2006. McClure, Stuart, Saumil Shah, and Shreeraj Shah. Web Hacking:attacks and defense. Addison Wesley. 2003.
E-resources and other digital material	**[1]** Hacker HighSchool Available at : http://www.hackerhighschool.org/lessons.html **[2]** Jeremy Koster, "Cyber Security Management",Nov, 19,2015, Available at: https://www.youtube.com/watch?v=A0yaymLmp10 **[3]** E.Rahul Naidu ,"Importance of Cyber Security" Available at : https://www.youtube.com/watch?v=MvK3lIDR3ms

19MTAC2036 - TECHNICAL REPORT WRITING

Course Category:	Audit Course	**Credits:**	0
Course Type:	Theory	**Lecture-Tutorial-Practice:**	2 - 0 - 0
Prerequisites:	Nil	**Total Marks:**	Nil

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Understand the significance of Technical Report Writing.
	CO2	Develop proficiency in writing technical reports.
	CO3	Apply the basic principles to prepare documentation using LATEX.
	CO4	Understanding the need of Bibliography and Reference Books for quality report writing

Contribution of Course Outcomes towards achievement of program outcomes (L-Low, M-Medium, H- High)	**PO 1**	**PO 2**	**PO 3**	**PO4**	**PO5**	**PO6**	**PSO 1**	**PSO 2**
CO1		H				L		
CO2		H				L	L	L
CO3		H						
CO4		H				M		

Course Content	**UNIT I:** **Writing scientific and engineering papers-**Title, Abstract, Introduction, Materials And Methods, Result, Discussion, Conclusion, Reference Books, Acknowledgements, Appendices, Hedging and Criticizing, Paraphrasing and Plagiarism.
	UNIT II: **Effective use of charts, graphs and tables-**Bar Chart, Line Chart, Pie Chart, Area Chart, Cylindrical Chart, Column Bars, Bubble Chart, Flow Diagram, Screen Capture, Tables **Writing Technical Reports-**Objectives Of Technical Report, Types Of Reports, Steps In Writing A Technical Report, Guidelines For Writing A Technical Report.
	UNIT III: LATEX- Introduction, Document Structure- Creating a Title, Sections, Labeling, Table of Contents **Typesetting Text-** Font Effects, Colored Text, Font Sizes, Lists, Comments & Spacing, Special Characters
	UNIT IV: **Tables, Figures, Equations-** Inserting Equations, Mathematical Symbols, Practical. **Inserting Reference Books-** Introduction, The BibTeX file, Inserting the bibliography, Citing Reference Books, Styles, Practical.

Text books and Reference books	**Text Book(s):** **[1]** Barun K Mitra, Effective Technical Communication-A Guide for Scientists and Engineers, Oxford University Press, 2006, ISBN: 978019568291. **[2]** LATEX for Beginners, Workbook Edition 5, March 2014 Document Reference: 3722-2014. **Reference Books:** **[1]** Goldbort R (2006) Writing for Science, Yale University Press (available on Google Books) **[2]** Day R (2006) How to Write and Publish a Scientific Paper, Cambridge University Press
E-resources and other digital material	**[1]** "LaTeX Basics" - https://www.overleaf.com/learn/latex/sections_and_chapters **[2]** "Citation & Style Guide" - https://libguides.cu-portland.edu/citationstyles

19ITDS2063 - TERM PAPER

Course Category:	**Term paper**	**Credits:**	**1**
Course Type:	Term paper	**Lecture-Tutorial-Practice:**	0-0-2
Prerequisites:	NIL	**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Identify real world problems related to Data Science area
	CO2	Analyse the problems from its state of the art for arriving at feasible solutions
	CO3	Prepare an organized report employing elements of technical writing & critical thinking
	CO4	Summarize and communicate the content to audience in an effective manner

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M-Medium, H- High)

	PO1	**PO2**	**PO3**	**PO 4**	**PO 5**	**PO 6**	**PSO1**	**PSO2**
CO1							L	L
CO2	H			L		M	M	M
CO3		M					L	L
CO4		M				H		

Course Content	Student shall collect the literature on the advanced topic in relevant fields and critically review the literature and submit it to the department in a form of report and shall make an oral presentation before the Academic Committee

19ITDS2051 – BIG DATA AND VISUALIZATION LAB

Course Category:	Programme Core	**Credits:**	1.5
Course Type:	Practical	**Lecture-Tutorial-Practice:**	0-0-3
Prerequisites:	Machine Learning	**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Implement big data analytics using Hadoop MapReduce, PIG and Spark.
	CO2	Process Semistructured and unstrctured data using NoSQL databases
	CO3	Construct visualizations for effective data analysis
	CO4	Build interactive dashboards for better decision making

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M-Medium, H- High)

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PSO1**	**PSO2**
CO1	H		H	H			M	
CO2			H	H				L
CO3			M				L	
CO4					M			M

Course Content

Week 1

Hadoop Installation on

[1] a)Single Node b) Multi Node Cluster
[2] Hadoop Distributed File System Commands for File/folder operations.
[3] Graphical User Interface of HDFS to monitor the health of the HDFS

Week 2

[1] Design a distributed application using MapReduce which processes a log file of a system.
[2] List out the users who have logged for maximum period on the system.
[3] Use simple log file from the Internet and process it using MapReduce.

Week 3:

[1] Design and develop a distributed application to find the coolest/hottest year from the available weather data. Use weather data from the Internet and process it using MapReduce.

[2] Monitor the MapReduce Job Interface.

Week 4:

[1] Install and Run Pig then write Pig Latin scripts to sort, group, join, project and filter the data.
[2] Implement data preprocessing operations using Pig Latin
[3] Implement log file processing using Pig Latin

	Week 5: [1] Implement document processing using MongoDB document ordiented database creation and its operations. [2] Write an application using Graph Database Neo4J to store facebook data and find the friend of friend, distance between two people.
	Week 6: [1] Implement data processing using Spark RDDS. [2] Process streamming data using Spark Streamming. [3] Connect Tableau with Spark cluster [4] Analyze and visualize with Tableau
	Week 7: [1] Reading different types of data sets (.txt, .csv) from web and disk and writing in file in specific disk location. And Reading Excel,XML data sheets in R. [2] Using with and without R objects on console, mathematical functions on console create R objects for calculator application and save in a specified location in disk. [3] Write an R script to find basic descriptive statistics using summary,str, quartile unction on mtcars& cars datasets and to find subset of dataset by using subset (),aggregate () functions on dataset.
	Week 8: Implementing data visualization using R [1] Find the data distributions using box and scatter plot. [2] Find the outliers using plot. [3] Plot the histogram, bar chart and pie chart on sample data.
	Week 9: Implementing basic operations in Tableau to get accustomed to its interface and Emphasizing the Results and Map View [1] Tableau Workspace,Connecting to a Data Source,Creating a view and Refining the view [2] Adding Filters to the view,Adding Colors to the view and Key Findings [3] Building a Map View,Getting into details and Identifying the Key points
	Week 10: Creating a dashboard and building story to showcase stories in presentation mode [1] Creating a dashboard and Adding Interactiveness [2] Building a Story and Making a Conclusion
	Week 11& 12: **Case study 1:** Implementation of all types of Join operations using MapReduce for the given multiple datasets. **Case Study 2:** Implementaion of PageRank algorithm using Apache Spark, The algorithm maintains two datasets: one of (pageID, link List) elements containing the list of neighbors of each page, and it has to find (pageID,rank) elements containing the current rank for each page.

	Case Study 3: From the given mobile data analyzing the user behavior, analyzing the network quality of service. **Case study 4:** Tracking Twitter data to see how fast information spreads online: Create a data visualisation to understand the spread of information and miss information insights of individual tweets online. **Case study 5:** Loan risk analysis : Create visualisation to analyse bank loan data to assess the risk of loan default. **Case study 6:** Motivate sales teams by modelling commission rates: Create a visualisation to explore the relationships between compensation type, commission for sales people to motivate them.
Text books and Reference books	**Text Book(s):** **[1]** Cole Nussbaumer Knaflic, Storytelling With Data: A Data Visualization Guide for Business Professionals,Wiley Publications,2015 **[2]** Tom Soukup and Ian Davidson, "Visual Data Mining: Techniques and Tools for Data Visualization and Mining", 1st Edition, John Wiley & Sons, 2002 **[3]** Gauravvaish, "Getting Started with NoSQL"(Kindle Edition),1st edition,2007. **[4]** [TomWhite, Hadoop,"The Definitive Guide",4th Edition,O'Reilly Publications, 2012. **[5]** Holden Karau, Andy Konwinski, Patrick Wendell, MateiZaharia, "Learning Spark: Lightning-Fast Big Data Analysis", O'Reilly Media, Inc. **Reference Books:** **[1]** Andy Kirk, Data Visualisation,C & M Digitals Pvt Ltd.,2016 **[2]** Chun-houh ChenWolfgang Härdle Antony Unwin, ,"Handbook of Data Visualization",Springer-Verlag Berlin Heidelberg,2008.
E-resources and other digital material	**[1]** Data Visualization in R Basic graphics,Prof.Shankar Narasimhan,Ragunathan Rengasamy,IIT Madras, https://nptel.ac.in/courses/106106179/11,2011 **[2]** Statistics and Visualization for Data Analysis and Inference, Dr. Ed Vul, Dr. Mike Frank, Massachusetts Institute of Technology, https://ocw.mit.edu/resources/res-9-0002statistics-and-visualization-for-data-analysis-and-inference-january-iap-2009/. 2009 **[3]** Python for Data Analysis - Python for Data Visualisation, https://www.youtube.com/watch?v=nXr2Xt52MfA,2017 **[4]** Python Data Visualization, https://www.coursera.org/learn/python-visualization **[5]** Data Visualization with Python and Matplotlib, https://www.udemy.com/data-visualization-with-python-and-matplotlib/,2018 **[6]** Integrating Hadoop and BI tools: Analyzing and Visualizing Big Data in Tableau with Spark https://datascience-enthusiast.com/Hadoop/SparkSQL_Tableau.html

19ITDS2052 - BUSINESS ANALYTICS LAB

Course Category:	Programme Core	**Credits:**	1.5
Course Type:	Practical	**Lecture-Tutorial-Practice:**	0-0-3
Prerequisites:	Machine Learning	**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Understand the principles of business analytics
	CO2	Predict the insights using tools and methods of data analysis and statistics.
	CO3	Develop approaches to applying forecasting and data mining techniques.
	CO4	Implement the models to solve decision problems for different applications

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M-Medium, H- High)

	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
CO1	L		M		M		H	
CO2	L			H	M		L	L
CO3	M			M	M		M	M
CO4	H		M	H	H	M	H	H

Course Content

Week 1
Application-Retail
Actionable insights for improving sales of a consumer durables retailer using POS data analytics

Week 2
Application-E-commerce
Customer engagement and brand perception of Indian ecommerce- A social media approach

Week 3:
Application-Web & Social Media
Tapping social media exchanges on Twitter- A case study of the 2015 Chennai floods

Week 4:
Application-Banking
Developing best prediction model of credit default for a retail bank

Week 5:
Application-Supply Chain
Developing a demand forecasting model for optimizing supply chain.

Week 6:
Application-Healthcare
Prediction of user's mood using smartphone data

	Week 7: **Application-**Insurance Personal insurance digital assistant, quartile unction on mtcars& cars datasets and to find subset of dataset by using subset (),aggregate () functions on dataset.
	Week 8: **Application-**Entrepreneurship Getting insights through data analysis
	Week 9: **Application-** Start Ups Start-up insights through data analytics for gaining stakeholders
	Week 10: **Application-**Finance & Accounts Vendor invoicing grief project
	Techniques used: Conditional Inference Tree, Logistic Regression, CART, Random Forest, Univariate and Bivariate Analysis, Multinomial Logistic Regression, NLP (Natural Language Processing), Vector Space Model, Latent Semantic Analysis, ADA Boost, KSVM, Text Mining, K-Means Clustering,Neural Network, Linear Discriminant Analysis, Hierarchical Clustering, Market Basket Analysis, RFM (Recency-Frequency- Monetary) Analysis, Time Series Forecasting.
Text books and Reference books	**Text Book(s):** **[1]** James Evans, "Business Analytics, Second Edition, Pearson Publications, 2017. **Reference Books:** **[1]** U. Dinesh Kumar, "Business Analytics - The Science of Data Driven Decision Making", First Edition, Wiley Publications, 2017. **[2]** SeemaAcharya R N Prasad, "Fundamentals of Business Analytics", 2nd Edition, Wiley Publications, 2016.
E-resources and other digital material	**[1]** https://www.youtube.com/watch?v=dzoe_InL-rE - Analytics Case Study from Telecom Industry, 2015 **[2]** https://www.youtube.com/watch?v=xApFTcsFPcQ - Uber case study, 2017 **[3]** https://www.youtube.com/watch?v=4vkCJcbfXV4 - Marketing Analytics: Predictive Analytics in Marketing, 2017 for-data-analysis-and-inference-january-iap-2009/. 2009

SEMESTER III

19ITDS3011 – SELF LEARNING (MOOCS COURSE)

Course Category:	Program Elective-V	**Credits:**	3
Course Type:	Theory	**Lecture-Tutorial-Practice:**	0-0-0
Prerequisites:		**Continuous Evaluation:**	-
		Semester end Evaluation:	-
		Total Marks:	100

Course Content	The department will recommend the self-learning courses from the available open courseware. The self- learning courses shall be taken from the list of approved MOOCs providers (SWAYAM / NPTEL/ EDX / Others).They must be approved/ratified in the respective Board of Studies and must have a minimum duration of 8 weeks/12 weeks.

19ITDS3061 - PROJECT PART-A

Course Category:	Project Part-A	**Credits:**	10
Course Type:	Project	**Lecture-Tutorial-Practice:**	0-0-20
Prerequisites:	Term paper	**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Identify a topic in relevant areas of Data Science
	CO2	Review literature to identify gaps and define objectives & scope of the project
	CO3	Apply appropriate research methodology to provide a solution to the chosen problem
	CO4	Prepare a technical report effectively using modern tools

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M- Medium, H- High)

	PO1	**PO2**	**PO3**	**PO 4**	**PO 5**	**PO 6**	**PSO1**	**PSO2**
CO1	M					L	L	L
CO2	M		L	L		L	L	L
CO3	H		H	M	L		L	L
CO4		H				H		

Course Content	The project shall be carried out in the major areas pertaining to the program approved by Project Review Committee and may address the societal problems/issues related to the program.

19ITDS3052 – INTERNSHIP

Course Category:	Internship	**Credits:**	2
Course Type:	Practical	**Lecture-Tutorial-Practice:**	0-0-4
Prerequisite:		**Continuous Evaluation:**	-
		Semester end Evaluation:	100
		Total Marks:	100

Course Content	The students shall undergo Internship for a period of six weeks in Industry/ Research organizations/ institute of higher learning approved by the Head of the Department during any time after the second semester

SEMESTER IV

19ITDS4061 - PROJECT PART-B

Course Category:	Project Part-B	**Credits:**	16
Course Type:	Project	**Lecture-Tutorial-Practice:**	0-0-32
Prerequisites:	Project Part-A	**Continuous Evaluation:**	40
		Semester end Evaluation:	60
		Total Marks:	100

Course Outcomes	**Upon successful completion of the course, the student will be able to:**	
	CO1	Identify methods and resources to carry out analysis and experiments
	CO2	Reorganize the procedures with a concern for society, environment and ethics
	CO3	Generate possible alternative solutions to chosen problem, compare, analyze them and derive performance metrics of the result
	CO4	Prepare a comprehensive report of the project work and also explore the possibility of publishing the work .

Contribution of Course Outcomes towards achievement of Program Outcomes (L-Low, M-Medium, H- High)		**PO1**	**PO2**	**PO3**	**PO 4**	**PO 5**	**PO 6**	**PSO1**	**PSO2**
	CO1	M			L		L	L	L
	CO2	L		L	L		L	L	L
	CO3	H		M	H	M	L	L	L
	CO4		H	M		M	M		

Course Content	Project Part B shall be the extension of project Part A.