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

INSTITUTE VISION

To nurture excellence in various fields of engineering by imparting timeless core values to the learners and to mould the institution into a centre of academic excellence and advanced research.

INSTITUTE MISSION

To impart high quality technical education in order to mould the learners into globally competitive technocrats who are professionally deft, intellectually adept and socially responsible. The institution strives to makethe learners inculcate and imbibe pragmatic perception and pro-active nature so as to enable them to acquire a vision for exploration and an insight for advanced enquiry.

DEPARTMENT VISION

To provide excellent information technology and computer science education by building strong teaching and research environment

DEPARTMENT MISSION

To offer high quality graduate and post graduate programs in information technology and computer science education and to prepare students for professional career or higher studies. The department promotes excellence in teaching, research, collaborative activities and positive contributions to society

PROGRAM EDUCATIONAL OBJECTIVES (PEOS)

PEO1: Graduate shall have successful professional career in data science and allied fields with in-depth knowledge and practical/interpersonal skills.

PEO2: Graduates shall have the ability to apply knowledge across the disciplines and in emerging areas of Data Science and Engineering for higher studies, research, employability/entrepreneurship, product development and handle the realistic problems.

PROGRAM OUTCOMES

PO1: Students are able to carry out research/investigation and development work to solve complex data analytical problems.

PO2: An ability to write a substantial technical report/document and to communicate effectively with a wide range of audience

PO3: An ability to demonstrate a degree of mastery over the area of Data Engineering by applying current techniques, skills and modern tools.

PO4: Conceptualize and solve data science problems using Mathematical, Statistical and Machine Learning concepts to arrive at feasible and optimal solutions for societal benefits.

PO5: Use data visualization and analytics tools and technologies to analyze business models and solve real world problems.

PO6:An ability to learn independently and engage in lifelong learning with understanding of professional, social and ethical responsibilities for the need of sustainable development

PROGRAM SPECIFIC OUTCOMES (PSOS)

PSO1: Mastery in emerging areas of Artificial Intelligence and Machine Learning withthe necessary functional and practical knowledge in Statistical and Evolutionary Techniques.

PSO2: Competent in Big Data Technologies, Cloud Computing, Deep Learning,Data Analytics and Business Analytics to work on industry/Societal problems.

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 Type	Course Code	Title of the Course	L	T	P	C
1.	Programme Core - I	19ITDS1001	Mathematical Foundations For Data Science	3	0	0	3
2.	Programme Core - II	19ITDS1002	Machine Learning	3	0	0	3
3.	Programme Core - III	19ITDS1003	Advanced Algorithms	3	0	0	3
4.	Programme Elective - I	19ITDS1014	A. Cloud Computing and Virtualization B. R For Data Science	3	0	0	3
5.	Programme Elective - II	19ITDS1015	A. Social and Information Network Analysis B. Optimization Techniques For Data Analysis C. Industry need based Course	3	0	0	3
6.	Mandatory Learning Course	19ITDS1026	Research Methodology and IPR	2	0	0	0
7.	Laboratory-I	19ITDS1051	Machine Learning Lab	0	0	3	1.5
8.	Laboratory-II	19ITDS1052	Python for Data Science Lab	0	0	3	1.5
TOTAL				17	0	06	18

SEMESTER II **Contact Hours : 25**

S.No.	Course Type	Code No.	Subject	L	T	P	C
1.	Programme Core - IV	19ITDS2001	Data Visualization	3	0	0	3
2.	Programme Core - V	19ITDS2002	Bigdata	3	0	0	3
3.	Programme Core - VI	19ITDS2003	Business Analytics	3	0	0	3
4.	Programme Elective - III	19ITDS2014	A. Computer Vision B. Deep Learning	3	0	0	3
5.	Programme Elective - IV	19ITDS2015	C. Natural Language Processing D. Cyber Security E. Industry Need Based Course	3	0	0	3
6.	Audit Course	19ITDS2036	Technical Report Writing	2	0	0	-
7.	Term Paper	19ITDS2066	Term Paper seminar – Literature Review for the proposed problem	2	0	0	1
8.	Laboratory-I	19ITDS2051	Bigdata And Visualization Lab	0	0	3	1.5
9.	Laboratory-II	19ITDS2052	Business Analytics Lab	0	0	3	1.5
TOTAL				19	0	6	19

L-Lecture, T-Tutorial, P-Practical, C-Credits

***Students to be encouraged to go to industrial training for atleast two months during semester break.**

#Students should conduct the Literature Survey for the proposed research topic and they need to develop a prototype or simulation based(must be outcome oriented)- the same to be presented in any conference(nation or international)

SEMESTER III **Contact Hours : 20**

S.No.	Course Type	Code No.	Subject	L	T	P	C
1.	Programme Elective - V	19ITDS3011	Choice for students to complete course in any MOOCS Platform	0	0	0	3
2.	Project (Part-A)	19ITDS3062	Dissertation */Project/Research Organization	0	0	20	10
2.	Internship	19ITDS3051	Intership/Summer Training in Research Organization/Institutions of Higher Learning(After II Sem)	0	0	0	2
			Total	0	0	20	15

L-Lecture, T-Tutorial, P-Practical, C-Credits
***To be continued in the IV Semester**
Programme Elective V and VI may be completed in semester I or II by satisfying the pre-requisites those who are going for industrial projec

SEMESTER IV **Contact Hours : 32**

S.No.	Course Type	Code No.	Subject	L	T	P	C
1	Project (Part-B)	19ITDS4061	Dissertation / Industrial Project	0	0	32	16
			Total	0	0	32	16

L-Lecture, T-Tutorial, P-Practical, C-Credits
Total Credits : 68
Course Type
Programme Core-0
Programme Elective – 1
Mandatory Course – 2
Audit Course – 3
Open Elective – 4
Internship/Laboratory – 5
Term Paper/Project – 6

19ITDS1001 MATHEMATICAL FOUNDATIONS FOR DATA SCIENCE

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	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,Confidence Intervals in Linear Regression, Trend Detection and Slope Estimation, 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

19ITDS1002 MACHINE LEARNING

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	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:** Problems that can be solved with machine learning, Looking for structure, Evaluating performance of 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
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:	-	**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-Moderate, H-High)

	PO1	PO2	PO3	PO4	PO5	PO6	PSO1	PSO2
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, Available: http://nptel.ac.in/courses/106101060/ [2].Prof.Tim Roughgarden, (03, 05, 2018). Kleinberg and Tardos, Algorithm Design,2015,Available:http://openclassroom.stanford.edu/MainFolder/CoursePage.php?course=IntroToAlgorithms [3] Sudarshan Iyengar: IIT Ropar (12, August, 2018). Data Structures and Algorithms[NPTEL]. Available: http://nptel.ac.in/

19ITDS1014A CLOUD COMPUTING AND VIRTUALIZATION

Course Category:	Programme Elective	**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 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-MODERATE, H- High)

	PO 1	PO 2	PO 3	PO 4	PO 5	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

19ITDS1014B R FOR DATA SCIENCE

Course Category:	Programme Elective IV	**Credits:**	3
Course Type:	Theory	**Lecture-Tutorial-Practice:**	3-0-0
Prerequisites:	**Mathematical Foundations For Data Science**	**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 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	PO4	PO5	PO6	PSO1	PSO2
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). Available: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. Available: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 network analysis
	CO2	Derive various relationships that exists in the network D
	CO3	Derive communities in the society and then generate recommendations.
	CO4	Apply link analysis and web search techniques for a given web application

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

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

Course Content

UNIT I:
Overview: Introduction to Social and Information 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

UNIT III:
Community analysis: Community detection, node degree, node reachability, Social communities, Community detection algorithms.
Recommendations: Recommendation System challenges, classical recommendation algorithms, recommendation using social context.
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 ,The Web as a Directed Graph. **Link Analysis and Web Search:** Searching the Web: The Problem of Ranking, Link Analysis Using Hubs and Authorities, PageRank. **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] "Networks: An Introduction by M. E. J. Newman, a college-level textbook about the science of networks." M. E. J. Newman Hardback, Oxford University Press, 2010.
E-resources and other digital material	[1]. Dr Bernie Hogan https://www.youtube.com/watch?v=2zhuj8ubinm Social network analysis - Introduction to structural thinking:, University of Oxford, 2018. [2]. S.R.S. Lyengar https://www.youtube.com/watch?v=b7Ug1h6EGNk "Introduction to Social Networks, 2017.

19ITDS1015B OPTIMIZATION TECHNIQUES FOR DATA ANALYSIS

Course Category:	Programme Elective II	**Credits:**	
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 programming problems
	CO4	Apply the modern methods of optimization to provide optimal solution.

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

	PO1	PO 2	PO3	PO4	PO5	PO6	PSO1	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**: 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 :**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. **References:** [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. 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/

19ITDS1026 Research Methodology and IPR

Course Category:	Mandatory Learning Course	**Credits:**	
Course Type:	Theory	**Lecture-Tutorial-Practice:**	2-0-0
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	Study the overview of the research methodology and the technique of defining a research problem
CO2	Analyzing various research designs and details of sampling design.
CO3	Apply the art of interpretation and the art of writing research reports
CO4	Recognize the role of intellectual property rights.

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

	PO 1	PO 2	PO 3	PO4	PO5	PO6	PSO1	PSO 2
CO1	M	H				M	L	L
CO2	M	L			L	L	L	L
CO3	H	M		M	L	L	L	L
CO4	M	H			M	H	L	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 the Research 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 Book(s):** [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 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:	Laboratory-I	**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,
Write some hypotheses possible for concept learning in hypothesis space H
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
(a) 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 Book(s):	**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	[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:	Laboratory-II	**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-Moderate, 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

Week1:

- Creation of Python scripts that uses Operators, Control flow statements
- Create Python Script that uses functions with various types of arguments such as default arguments, keyword argument and variable length arguments

Week2:

- Create python script to implement fruitful functions and void functions
- Simulate the calculator application to perform all the operation to be implemented on it.
- 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 appropriate boolean.

	Week3: • Creation of python programs on the modules Numpy • Analyse the given series of data using pandas • Python programs that uses the dictionaries, tuples and other data structures
	Week4: • Python Program on Text Files reading, handling and manipulation on these files • Creation of python programs to handle missing data and hierarchical indexing, data aggregation on multi indexes
	Week5: • Python scripts that access the data from a given database
	Week6: • 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
	Week7: • Python programs on data transformation and string manipulation • Python Programs to simulate Queue Operations • Implement the data structure of binary search trees, using classes, with operations for inserting and finding an element
	Week8: • Python programs on Scatter plots with histograms and a Scatter plot matrix for a given data • Find the root words of the given list of words using Porter and Snowball Stemming • 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. Design a python framework to extract the meaningful information from the images available from Bhuvan.
	Week11 & 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 Books 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 [4] 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]. Available: https://www.coursera.org/ [2]. MadhavanMukund, (12, may, 2018). Programming, Data Structures & Algorithms using Python [NPTEL]. Available: http://nptel.ac.in/ [3]. Keith Galli Complete Python NumPy Tutorial (Creating Arrays, Indexing, Math, Statistics, Reshaping) Available: https://www.youtube.com/watch?v=GB9ByFAIAH4 [4]. Keith Galli Complete Python Pandas Data Science Tutorial! (Reading CSV/Excel files, Sorting, Filtering, Groupby) Available: https://www.youtube.com/watch?v=vmEHCJofslg [5]. CS Dojo, Intro to Data Analysis / Visualization with Python, Matplotlib and Pandas | Matplotlib Tutorial Available: https://www.youtube.com/watch?v=a9UrKTVEeZA