SANJEEV AGRAWAL GLOBAL EDUCATIONAL (SAGE) UNIVERSITY, BHOPAL

Scheme & Syllabus

for

Master of Technology (Data Science)

School of Advanced Computing

Program Educational Objectives (PEOs)

PEO-1: To have successful professional career in Data Science (DS) and allied fields with in-depth knowledge and practical/interpersonal skills.

PEO-2: To produce post graduates who can contribute in research and development in the field of Data Science.

PEO-3: To engage in sustainable development and demonstrate data analysis skills for effective interpretation and decision making to solve real life problems.

PEO-4: To develop team work capability so that they can work on multidisciplinary projects and exhibits high level of professional and ethical values.

PEO-5: Promote design, research, and implementation of products and services in the field of Data Science through strong communication and entrepreneurial skills.

Program Outcomes (POs):

PO-1: Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO-2: Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO-3: Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO-4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO-5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO-6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO-7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO-8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO-9: Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO-10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO-11: Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO-12: Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Curriculum Components

Components	Credits
Program Core (08 Courses)	28
Program Electives (Discipline Specific Electives) (04 Courses)	16
Project Based Learning (PBL)/MOOCs (04 courses)	12
Project (02 Courses)	28
Total	**84**

First Semester																
Course Code	**Course Title**	**Contact Hours Per Week**			**Credits**	**ESE Duration (Hours)**	**Theory**						**Practical**			
		L	**T**	**P**			**MSE**	**ASG**	**TA**	**ATTD**	**ESE**	**Tot**	**CE**	**ESE**	**Tot**	**GT**
MA20M101	Advanced Mathematics	3	1	-	4	3	30	05	05	10	50	100	-	-	-	100
AI20M102	Artificial Intelligence and Machine Learning	3	-	2	4	3	30	05	05	10	50	100	20	30	50	150
AI20M103	Essentials of Data Science	3	-	2	4	3	30	05	05	10	50	100	20	30	50	150
Table-1	DSE-I	3	-	2	4	3	30	05	05	10	50	100	20	30	50	150
Table-1	DSE-II	3	1	-	4	3	30	05	05	10	50	100	-	-	-	100
AI20M104	Software Lab-I	-	-	4	2	2	-			-	-		20	30	50	50
PB20M101	Project Based Learning-I	-	-	4	2	2	-			-	-		50^	50	100	100
		Total			**24**											**800**

L-Lecture, T-Tutorial, P-Practical, MSE- Mid Semester Exam, ASG- Assignment, TA- Teacher's Assessment, ATTD-Attendance, CE-Continuous Evaluation ,ESE- End Semester Exam, Tot-Total, GT-Grand Total, ^ - Two assessment by panel of Experts

Second Semester																
Course Code	**Course Title**	**Contact Hours per Week**			**Credits**	**ESE Duration (Hours)**	**Theory**						**Practical**			**GT**
		L	**T**	**P**			**MSE**	**ASG**	**TA**	**ATTD**	**ESE**	**Tot**	**CE**	**ESE**	**Tot**	
DS20M201	Natural Language Processing	3	-	2	4	3	30	05	05	10	50	100	20	30	50	150
AI20M202	Deep Learning	3	-	2	4	3	30	05	05	10	50	100	20	30	50	150
AI20M203	Big data Analytics	3	1	2	4	3	30	05	05	10	50	100	20	30	50	150
Table-1	DSE-III	3	1	0	4	3	30	05	05	10	50	100	-	-	-	100
Table-1	DSE-IV	3	1	-	4	3	30	05	05	10	50	100	-	-	-	100
AI20M204	Software Lab-II	-	-	4	2	2	-				-	-	20	30	50	50
PB20M201	Project Based Learning-II	-	-	4	2	2	-				-	-	50^	50	100	100
		Total			**24**											**800**

L-Lecture, T-Tutorial, P-Practical, MSE- Mid Semester Exam, ASG- Assignment, TA- Teacher's Assessment, ATTD-Attendance, CE-Continuous Evaluation ,ESE- End Semester Exam, Tot-Total, GT-Grand Total, ^ - Two assessment by panel of Experts

Third Semester																	
Course Code	**Course Title**	**Contact Hours per Week**			**Credits**	**ESE Duration (Hours)**	**Theory**						**Practical**			**GT**
		L	**T**	**P**			**MSE**	**ASG**	**TA**	**ATTD**	**ESE**	**Tot**	**CE**	**ESE**	**Tot**	
	MOOC-1	-	-	8	4	-	-	-	-	-	-	-	50	50	100	100
	MOOC-2	-	-	8	4	-	-	-	-	-	-	-	50	50	100	100
DS20M303	Dissertation Phase-I	-	-	24	12	2	-				-	-	150	150	300	300
		Total			**20**											**500**

L-Lecture, T-Tutorial, P-Practical, MSE- Mid Semester Exam, ASG- Assignment, TA- Teacher's Assessment, ATTD-Attendance, CE-Continuous Evaluation ,ESE- End Semester Exam, Tot-Total, GT-Grand Total

Fourth Semester																
Course Code	**Course Title**	**Contact Hours per Week**			**Credits**	**ESE Duration (Hours)**	**Theory**						**Practical**			**GT**
		L	**T**	**P**			**MSE**	**ASG**	**TA**	**ATTD**	**ESE**	**Tot**	**CE**	**ESE**	**Tot**	
DS20M401	Dissertation Phase-II	-	-	32	16						-	-	200	200	400	400
		Total			**16**											**400**

L-Lecture, T-Tutorial, P-Practical, MSE- Mid Semester Exam, ASG- Assignment, TA- Teacher's Assessment, ATTD-Attendance, CE-Continuous Evaluation ,ESE- End Semester Exam, Tot-Total, GT-Grand Total

Distribution of credits across all components

SEM No.	Prog. Core	Discipline Specific Electives (DSE)	Project Based Learning (PBL)/ MOOCs	Project	Total Credit
I.	14	08	02	-	24
II.	14	08	02	-	24
III.	-	-	08	12	20
IV.	-	-	-	16	16
Total	**28**	**16**	**12**	**28**	**84**

Table-1
List of Discipline Specific Electives (DSE)

SN	Course Code	DSE-I
1.	AI20M106	Internet of Things
	DS20M107	Thick Data Analysis
	DS20M108	Web Databases and Information Systems
SN	**Course Code**	**DSE-II**
2.	AI20M109	Data Mining and Warehousing
	DS20M110	Advanced DBMS
	DS20M111	Statistical Foundations for Data Science
SN	**Course Code**	**DSE-III**
1.	DS20M207	GPU Computing
	DS20M208	Bioinformatics
	DS20M209	Optimization Techniques
SN	**Course Code**	**DSE-IV**
2.	DS20M210	Geospatial Information Management
	DS20M211	Information Retrieval
	AI20M212	Semantic Web

SANJEEV AGRAWAL GLOBAL EDUCATIONAL (SAGE) UNIVERSITY, BHOPAL

Syllabus

for

Master of Technology (Data Science)

I Semester

School of Advanced Computing

COURSE CODE	ADVANCED MATHEMATICS	Total Lecture : 60 Theory : 45 Tutorial : 15
MA20M101		(LTP=3 – 1 – 0 = 4)

Course Objectives:

- To introduce students to the theoretical distributions, sampling distributions and their applications
- To introduce the students to the solution of partial differential equation
- Demonstrate an understanding to the theory and applications of linear algebra
- To extend the concept of the computer algorithms related to dimensionality reduction and feature extraction.
- To introduce the concepts of Stochastic process and Markov process transition.

UNIT	CONTENTS	HOURS
I.	Probability, compound probability and discrete random variable. Binomial, Normal and Poisson's distributions, Sampling distribution, elementary concept of estimation and theory of hypothesis, recurred relations.	8
II.	Solution of Partial Differential Equation (PDE) by separation of variable method, numerical solution of PDE (Laplace, Poisson's, Parabola) using finite difference methods, Elementary properties of FT, DFT, WFT, Wavelet transform, Haas transform.	10
III.	Finite differences: forward, backward and central difference operators, polynomial interpolation: equally spaced and unequally spaced data; Numerical Differentiation, Numerical integration- Trapezoidal and Simpson1/3rd and 3/8th rules; Initial value problems - Taylor series method, Euler and modified Euler methods, Runge- Kutta methods.	10
IV.	Solution of Linear systems– Gaussian elimination method, LU factorization method, Cholesky's factorization method. Linear least-squares problems - Normal equations, QR method (or Gram Schmidt Ortho- normalization), Singular value decomposition (SVD) for linear least-squares problems, numerical rank determination via SVD, Principal Component Analysis.	10

V.	Stochastic process, Markov process transition probability transition probability matrix, just and higher order Markov process, Application of Eigen value problems in Markov Process, Markov chain. Queuing system, transient and steady state, traffic intensity, distribution queuing system, concepts of queuing models (M/M/1: Infinity/ Infinity/ FC FS), (M/M/1: N/ Infinity/ FC FS), (M/M/S: Infinity/ Infinity/ FC FS)	07

Course Outcomes as per Bloom's Taxonomy

At the end of the course the students should be able to:

CO 1	**Understand**[2] probability, sampling distribution and discrete random variable.
CO 2	**Understand**[2] the terms and their applications of Solution of Partial Differential Equations
CO 3	**Understand**[2] the numerical methods and their use in obtaining approximate solutions to otherwise intractable linear/non-linear system of equations and differential equations.
CO 4	**Analyse**[4] the fundamental use of matrices in the computer algorithms related to dimensionality reduction and feature extraction.
CO 5	**Implement**[3] Stochastic process, Markov process transition probability transition probability matrix and Markov process.
Text Books	• *Gupta* S C & *Kapoor* V K (2014): **Fundamentals of Mathematical Statistics**, New Delhi :Sultan Chand & Sons,. • Jimmie Gilbert (2010): **Linear Algebra And Matrix Theory**, India: Elsevier. • Grewal B S (2014): **Numerical Methods in Engineering & Science: With Programs in C, C++ & MATLAB,** 10th Edition, Delhi: Khanna Publishers.
Reference Books	• Rohatgi, V.K., and Saleh, A.K.Md. Ehsanes (2009): **An introduction to probability and statistics- Second Edition**, India: Wiley. • Trefethen L. N. and Bau David (1997): **Numerical Linear Algebra**, Philadelphia: SIAM.

COURSE CODE	ARTIFICIAL INTELLIGENCE & MACHINE LEARNING	Total Lecture : 60 Theory : 45 Practical : 15
AI20M102		(LTP= 3 – 0 – 2 = 4)

Course Objectives:

- Learning basic concepts of various machine learning methods is primary objective of this course.
- To be able to formulate machine learning problems corresponding to different applications.
- To understand a range of machine learning algorithms along with their strengths and weaknesses.
- To be able to apply machine learning algorithms to solve problems of moderate complexity.
- Make student able to learn mathematical concepts, and algorithms used in machine learning techniques for solving real world problems

UNIT	CONTENTS	HOURS
I.	Introduction: Artificial Intelligence, AI Problems, AI Techniques, The Level of the Model, Criteria For Success. Defining the Problem as a State Space Search, Problem Characteristics, Production Systems, Search: Issues in The Design of Search Programs, Un-Informed Search, BFS, DFS; Heuristic Search Techniques: Generate-And- Test, Hill Climbing, Best-First Search, A*Algorithm, Problem Reduction, AO*Algorithm, Constraint Satisfaction, Means-Ends Analysis.	6
II.	Introduction to Machine Learning: Applications of ML, Difference between Data Mining and Predictive Analysis, Tools and Techniques of Machine Learning. What is Machine Learning, Basic Terminologies of Machine Learning	6
III.	Types of Machine Learning: Reinforcement Learning, Machine Learning Lifecycle, Supervised Learning, Unsupervised Learning, Introduction to ANN and deep neural networks, Principle component analysis.	6
IV.	Supervised Learning: Classification and Regression: Non-Linear Regression, Classification: K-Nearest Neighbour, Decision Trees, Linear Regression, Logistic Regression, Naïve Bayes, Regression: Model Representation, Support Vector Machines.	6

V.	Unsupervised and Reinforcement Learning: Genetic Algorithm, Clustering: K-Means Clustering, Density-Based Clustering, Hierarchical clustering	6

Course Outcomes as per Bloom's Taxonomy

At the end of the course the students should be able to:

CO 1	**Demonstrate**[2] fundamental understanding of the history of artificial intelligence (AI) and its foundations.
CO 2	**Apply**[3] basic principles of AI in solutions that require problem solving, inference, perception, knowledge representation, and learning.
CO 3	**Demonstrate**[2] awareness and a fundamental understanding of various applications of AI techniques in intelligent agents, expert systems, artificial neural networks and other machine learning models.
CO 4	**Understand**[2] about various learning and their types.
CO 5	**Demonstrate**[2] an ability to share in discussions of AI, its current scope and limitations, and societal implications.
Text Books	• Luger F George: **Artificial Intelligence**, Singapore Pearson Education Publications • Rich Elaine and Knight: ***Artificial Intelligence,*** Mcgraw-Hill Publications.
Reference Books	• Patterson: ***Introduction To Artificial Intelligence & Expert Systems***,PHI • G. Weiss: ***Multi Agent systems- a modern approach to Distributed Artificial intelligence,*** MIT Press. • Russell and Norvig: ***Artificial Intelligence : A modern Approach***, Prentice Hall

COURSE CODE	ESSENTIALS OF DATA SCIENCE	Total Lecture : 60 Theory : 45 Practical : 15
AI20M103		(LTP= 3 – 0 – 2 = 4)

Course Objectives:

- To provide an understanding of the data science.
- To understand exploratory data analysis, and various machine learning methods
- To learn and understand analysis of data and pattern.
- To understand data Visualization techniques.
- To understand Feature Generation & Feature Selection methods.

UNIT	CONTENTS	HOURS
I.	Introduction: What is Data Science? - Big Data and Data Science hype – and getting past the hype – Datafication - Current landscape of perspectives - Skill sets needed, Statistical Inference - Populations and samples - Statistical modeling, probability distributions, fitting a model - Intro to R.	10
II.	Exploratory Data Analysis and the Data Science Process - Basic tools (plots, graphs and summary statistics) of EDA - Philosophy of EDA - The Data Science Process - Case Study: RealDirect (online real estate firm), Data Cleaning, Web Scraping. Linear Regression, Logistic Regression, k-Nearest Neighbors (k-NN), k-means, Clustering, Usage in Applications - Filtering Spam - Why Linear Regression and k-NN are poor choices for Filtering Spam - Naive Bayes and why it works for Filtering Spam - Data Wrangling: APIs and other tools for scrapping the Web	10
III.	Feature Generation and Feature Selection (Extracting Meaning From Data) - Motivating application: user (customer) retention - Feature Generation (brainstorming, role of domain expertise, and place for imagination) - Feature Selection algorithms – Filters; Wrappers; Decision Trees; Random Forests.	10
IV.	Recommendation Systems: Building a User-Facing Data Product - Algorithmic ingredients of a Recommendation Engine - Dimensionality Reduction - Singular Value Decomposition - Principal Component Analysis - Exercise: build your	8

	own recommendation system.	
V.	Mining Social-Network Graphs - Social networks as graphs - Clustering of graphs - Direct discovery of communities in graphs - Partitioning of graphs - Neighborhood properties in graphs, Data Visualization - Basic principles, ideas and tools for data visualization.	7

List of Experiments:

1. Use R to carry out basic statistical modeling and analysis.
2. Apply basic tools (plots, graphs, summary statistics) to carry out EDA.
3. Use APIs and other tools to scrap the Web and collect data.
4. Apply EDA and the Data Science process in a case study.
5. Apply basic machine learning algorithms (Linear Regression, k-Nearest Neighbors (k-NN), k-means, Naive Bayes) for predictive modeling. Explain why Linear Regression and k-NN are poor choices for Filtering Spam.
6. Explain why Naive Bayes is a better alternative.
7. Identify basic Feature Selection algorithms (Filters, Wrappers, Decision Trees, Random Forests) and use in applications.
8. Identify and explain fundamental mathematical and algorithmic ingredients that constitute a Recommendation Engine (dimensionality reduction, singular value decomposition, principal component analysis). Build own recommendation system using existing components.
9. Create effective visualization of given data (to communicate or persuade).

Course Outcomes as per Bloom's Taxonomy

At the end of the course the students should be able to:

CO 1	**Understand**[2] what Data Science is and the skill sets needed to be a data scientist.
CO 2	**Understand**[2] how to perform evaluation of learning algorithms and model selection.
CO 3	**Apply**[3] and use APIs and other tools to scrap the Web and collect data.
CO 4	**Experiment**[5] with Feature Generation algorithms (Filters, Wrappers, Decision Trees, Random Forests) and use in applications.
CO 5	**Create**[6] effective visualization of given data (to communicate or persuade).

Text Books	• Mitchell M Tom (2017): ***Machine Learning***, Ist Edition, India :Tata McGraw Hill • O'Neil Cathy and Schutt Rachel, (2013): ***Doing Data Science, Straight Talk From The Frontline***, Ist Edition, O′Reilly Publishers.
Reference Books	• VanderPlas Jake (2016***): Python Data Science Handbook,*** O'Reilly Media, Inc. • Leskovek Jure, Rajaraman Anand and Ullman Jeffrey (2014): ***Mining of Massive Datasets,*** v2.1, Cambridge University Press. • Murphy P Kevin (2012): ***Machine Learning: A Probabilistic Perspective***, Illustrated edition, MIT Press. • Provost Foster and Fawcett Tom (2013): ***Data Science for Business:What You Need to Know about Data Mining and Data-analytic Thinking***, 1st Edition, O′Reilly Publishers. • Hastie Trevor, Tibshirani Robert and Friedman Jerome (2009): ***Elements of Statistical Learning,*** IInd Edition, Springer Series. • Blum Avrim, Hopcroft John and Kannan Ravindran (2018): ***Foundations of Data Science***. • Zaki J Mohammed and Miera Wagner (2014): ***Data Mining and Analysis: Fundamental Concepts and Algorithms***, Cambridge University Press. • Han Jiawei, Kamber Micheline and Pei Jian (2011): ***Data Mining: Concepts and Techniques***, Third Edition.

COURSE CODE	DSE-I INTERNET OF THINGS	Total Lecture : 60 Theory : 45 Practical : 15
AI20M106		(LTP= 3 – 0 – 2 = 4)

Course Objectives

- To Understand the Architectural Overview of IoT
- To Understand the IoT Reference Architecture and Real World Design Constraints
- To Understand the various IoT Protocols (Data link, Network, Transport, Session, Service)
- To understand security aspect in IoT..
- To design an IoT application.

UNIT	CONTENTS	HOURS
I.	Evolution of IoT: Review of computer communication concepts (OSI layers, components, packet communication, Networks, TCP-IP, subnetting,IPV4 addressing and challenges). IPV6 addressing. IoT architecture reference layer.	10
II.	Introduction to IoT components: Characteristics IoT sensor nodes, Edge computer, cloud and peripheral cloud, single board computers, open source hardwares, Examples of IoT infrastructure	10
III.	IoT protocols and softwares: MQTT, UDP, MQTT brokers, publish subscribe modes, HTTP, COAP,XMPP and gateway protocols, IoT Communication Pattern, IoT protocol Architecture, Selection of Wireless technologies (6LoWPAN, Zigbee, WIFI, BT, BLE,SIG,NFC, LORA,Lifi,Widi)	10
IV.	IoT security: Need for encryption, standard encryption protocol, light weight cryptography, Quadruple Trust Model for IoT-A – Threat Analysis and model for IoT-A, Cloud security. open source IoT platforms, cloud dashboards.	8
V.	IoT application and its Variants: Case studies: IoT for smart cities, health care, agriculture, smart meters.M2M, Web of things, Cellular IoT, Industrial IoT, Industry 4.0,IoT standards.	7

List of Practical's

1. Familiarization with Arduino/Raspberry Pi and perform necessary software installation.
2. To interface LED/Buzzer with Arduino/Raspberry Pi and write a program to turn ON

LED for 1 sec after every 2 seconds.

3. To interface Push button/Digital sensor (IR/LDR) with Arduino/Raspberry Pi and write a program to turn ON LED when push button is pressed or at sensor detection
4. To interface DHT11 sensor with Arduino/Raspberry Pi and write a program to print temperature and humidity readings.
5. To interface motor using relay with Arduino/Raspberry Pi and write a program to turn ON motor when push button is pressed.
6. To interface OLED with Arduino/Raspberry Pi and write a program to print temperature and humidity readings on it.
7. To interface Bluetooth with Arduino/Raspberry Pi and write a program to send sensor data to smartphone using Bluetooth.
8. To interface Bluetooth with Arduino/Raspberry Pi and write a program to turn LED ON/OFF when '1'/'0' is received from smartphone using Bluetooth.
9. Write a program on Arduino/Raspberry Pi to upload temperature and humidity data to thingspeak cloud.
10. Write a program on Arduino/Raspberry Pi to retrieve temperature and humidity data from thingspeak cloud.
11. To install MySQL database on Raspberry Pi and perform basic SQL queries.
12. Write a program on Arduino/Raspberry Pi to publish temperature data to MQTT broker.
13. Write a program on Arduino/Raspberry Pi to subscribe to MQTT broker for temperature data and print it.
14. Write a program to create TCP server on Arduino/Raspberry Pi and respond with humidity data to TCP client when requested.
15. Write a program to create UDP server on Arduino/Raspberry Pi and respond with humidity data to UDP client when requested.

Course Outcomes as per Bloom's Taxonomy

At the end of the course the students should be able to:

CO 1	**Understand**[2] the definition and significance of the Internet of Things
CO 2	**Understand**[2] the architecture, operation, and business benefits of an IoT solution
CO 3	**Understand**[2] various layers and Examine the potential business opportunities that IoT can uncover
CO 4	**Interpret**[2] the relationship between IoT, cloud computing, and big data
CO 5	To **Identify**[1] how IoT differs from traditional data collection systems
Text Books	• Holler Jan, Tsiatsis Vlasios, Mulligan Catherine, Avesand Stefan, Karnouskos Stamatis, Boyle David (2014): ***From Machine-to-Machine to the Internet of Things:Introduction to a New Age of Intelligence***, 1st Edition, Academic Press. • Scholz-Reiter Bernd, Michahelles Florian: ***Architecting the Internet of***

	Things, Springer • Madisetti Vijay and Bahga Arshdeep (2014): ***Internet of Things (A Hands-on Approach)***, 1 st Edition, VPT. • http://www.cse.wustl.edu/~jain/cse570-15/ftp/iot_prot/index.html
Reference Books	• Minoli Daniel: ***Building the Internet of Things with IPv6 and MIPv6: The Evolving World of M2M Communications***, Willy Publications • Waher Peter: ***Learning Internet of Things***, MUMBAI: PACKT publishing, BIRMINGHAM.

COURSE CODE	DSE-I	Total Lecture:45 Practical:15
DS20M107	THICK DATA ANALYSIS	3 – 0–2 = 4

Course Objectives

- Learn the basic concepts of thick data.
- Identify problems that matter to users or stakeholders.
- Propose solutions to the problems.
- Test and refine solutions on a small scale.
- To implement solutions on a large scale.

Unit	Contents	Hours
1	Introduction to big data : Introduction to Big Data Platform – Challenges of Conventional Systems - Intelligent data analysis – Nature of Data - Analytic Processes and Tools - Analysis vs Reporting. Thick data – overview, Challenges of Conventional Systems , compare with NOSQL data , Big data vs. Thick Data, Role in problem solving, Strengths and limitations.	10
2	Case Studies : A WORLD BANK SOCIAL OBSERVATORY , Background and Motivations, Embedded research, Improving the "adaptive capacity", Participatory Tracking, Standard vs. designed-by-participants survey Findings and Insights.	7
3	CASE Studies : APHIS OUTREACH PROGRAM , Background and Motivations, Identifying the Problem, Analyzing Data , Converting Results into Program Design , Audience Segmentation, Bridging Big Data and Human Expertise , Bridging Big Data and Human Expertise , Data scientists must know the agency's context.	10
4	Learned and Recommendations, Summary of key characteristics of case studies(World Bank's Social Observatory, APHIS Public Outreach, LA Express Park), Key Benefits of Mixed Analytics, Relevance Ownership and empowerment , Enhanced communication , Mid-course correction.	8
5	Introduction to HR Analytics: Overview of HR Process, HR as an expense, the analytics and prediction Strategic Human capital measures, business analysis and rational action. Benefits of Analytics in Improving HR Process, Intersection of people and profits. Technology Used, SWOT Analysis of HR analytics.	10

	Employee Engagement Measurement Process: Attracting, motivating and retaining people Organization Gap and Alignment Analytics. Process to assess and prioritize organization gaps and identify alignment opportunities. HR Alignment Inventory. Developing performance metrics/Predicting future 'performance' Developing metrics to capture the fallouts of HR Policies.	

Course Outcomes

At the end of the course the students should be able to:

CO 1	**understand**[2] big data and thick data.
CO 2	To **understand**[2] the different case study of world bank.
CO 3	To **understand**[2] Bridging Big Data and Human Expertise in big data analysis.
CO 4	To **understand**[2] the key characteristics of case studies in Thick data.
CO 5	To **understand**[2] the SWOT Analysis of HR analytics and business analysis.
Text Books	• Yuen Yuen Ang, (2019). **Integrating Big Data and Thick Data to Transform Public Services Delivery**. • Gilbert N. Nyaga , Gary J. Young, **Ten Actions to Improve Inventory Management in Government: Lessons From VA Hospitals**.
Reference Books	• Alfred T. Ho, Bo McCall, **Ten Actions to Implement Big Data Initiatives: A Study of 65 Cities**. • Kevin C. Desouza, **Delivering Artificial Intelligence in Government: Challenges and Opportunities**.

Code	DSE-I	Total Lecture:45 Tutorial: 15
DS20M108	WEB DATABASES AND INFORMATION SYSTEMS 3–1-0 = 4	

Course Objectives

- This course covers fundamental concepts of database and information systems.
- Enable students with principles and knowledge of web databases and semi-structured data in the Internet environment.
- Teach students with sound techniques in designing and querying web database.
- Provide detailed examples of how advance techniques are being applied in web database applications now and the near future.
- Understand the concepts of database modeling and design.

Unit	Contents	Hours
1	Introduction to Web Database Systems: Review of relational, object-oriented, and XML databases. Semistructured Data: XML basics, the simple API for XML, parsing XML, W3C document object model, SAX parsing, , XML graph model. XML Data Modeling: DTD and XML Schema. Native XML databases, transforming XML data to relations, and storing XML data in relational databases.	10
2	Querying of Web Databases: XPath, XQuery, XQL, XMLQL, unQL, grouping with nested queries, binding elements and contents, querying attributes, joining elements by value, tag variables, mediators for data integration, distributed evaluations, query processing and evaluations.	10
3	Typing and Indexing: Schema formalisms, Datalog, extracting schemas from data, data guides, inferring schemas from queries, attribute multiplicity, path constraints in semistructured data, XML schema, XML views, modelling data types, indexing and extending databases in XML. Web Transactions Management: Serializing relational query results in XML, prefetching and caching, XML transaction servers.	10
4	Web Database Systems: Using XML with relational databases, XML support in MS/SQL and Oracle, compressing XML objects, XMill, Web intermediary, and XML wrappers.	8
5	Web Services and Applications: Dynamic media contents composition, B2B and B2C e-commerce applications, web services, UDDI, EDI applications, ebXML, VBL, PML and education applications.	7

Course Outcomes

At the end of the course the students should be able to:	
CO 1	**Understand**[2] and use XML for web page design.
CO 2	**Use**[3] queries for web database access.
CO 3	**Use**[3] various queries for database access.
CO 4	To **use**[3] various typing and indexing techniques.
CO	**Understand**[2] EDI applications.
Text Books	• Bhowmick, Madria, Ng (2013): **Web Data Management: A Warehouse Approach.** IInd Ed, Springer. • Abiteboul, S., Manolescu, I., Rigaux, P., Rousset, M.C., Senellart, P (2011):, **Web Data Management**, Ist Edition, Cambridge University Press.
Reference Books	• Silberschatz, Korth, and Sudarshan (2005): **Database System Concepts**, Vth Edition, McGraw-Hill Education. • Rajshekhar Sunderraman (2003): **Oracle 9i Programming: A Primer**, Ist Edition, Addison Wesley.

COURSE CODE	DSE-II DATA MINING AND WAREHOUSING	Total Lecture : 60 Theory : 45 Tutorial : 15
AI20M109		(LTP= 3 – 1– 0 = 4)

Course Objectives:

- Understand the components, architecture and other important tools of data warehousing.
- To understand data pre-processing and data visualization techniques
- To study algorithms for finding hidden and interesting patterns in data
- To understand and apply various classification and clustering techniques using tools.
- To understand types of association rules & algorithms.

UNIT	CONTENTS	HOURS
I.	Introduction to DWH : Data warehouse (DWH): Need, Definition, Advantages of DWH, OLTP Vs DWH, 3-tier Architecture, DWH Design Process, ETL Process, DWH Back-end Tools and Utilities, Metadata Repository, Models of DWH: Enterprise Warehouse, Data Mart, Virtual Warehouse, Comparison.	10
II.	Dimensional Modeling: Dimensional Model Vs ER Model, DWH Schemas: Star, Snowflake, Fact Constellation, their Comparison, Techniques to Handle Changing Dimensions, Aggregation, Families of Fact Tables, Fact Less Fact Tables; Data Warehouse Indexing: Factors used to select an Indexing Technique, Properties of a Good Indexing Technique for DWH, Indexing Techniques: Projection Index, Bitmap Index (Pure and Encoded), Join Index and their Comparison.	10
III.	Data Mining and Functionalities: Need of Data Mining, Knowledge Discovery in Database (KDD), Architecture of Data Mining System, Data Mining on Different kind of Data, Data Mining Functionalities; Data Preprocessing: Need, Cleaning, Integration, Transformation, Reduction, Discretization, Concept Hierarchy Generation.	10
IV.	Cluster Analysis: Categories of Clustering methods, Partitioning methods: k-Means, kMedoids; Prediction: Numerical Prediction, Linear, Non-Linear	8

	Regression; Outlier Analysis: Applications, Techniques.	
V.	Classification: Decision Tree Classifier, Rule Based Classification, Bayesian Classification, Neural Network Classification: Back Propagation Algorithm, Lazy Learner: kNN Classifier, Case-Based Reasoning, Other: Fuzzy Set Approach, Classifier Accuracy Measures, Techniques for Evaluating Classifier Accuracy; Frequent Itemset Mining: Interesting Item Set Mining: Market Basket Analysis, APriori Algorithm, Generating Association Rules, Types of Association Rules, Correlation Analysis. Data Mining on different Databases: Multimedia Data Mining, Web Mining, Text Mining, Spatial Data Mining, Mining on Social Networks, Multi-relational Data Mining.	7

Course Outcomes as per Bloom's Taxonomy

At the end of the course the students should be able to:

CO 1	**Construct**[6] an end-to-end data warehousing solution.
CO 2	**Evaluate**[5] various data processing algorithms in their applicability to different problems.
CO 3	**Display**[4] the process of converting data into a user defined format required for particular analysis.
CO 4	**Utilize**[2] statistical tools in deriving insights from data.
CO 5	**Describe**[1] various techniques for clustering and classification. Apply various techniques to solve real-world data analysis problems
Text Books	• Kimball, Reeves, Ross, waite Thornth, Wiley John. (2002): ***The Data Warehouse Lifecycle Toolkit***, ISBN 9971-51-415-X. • Han Jiawei and Kamber Micheline, Kaufman Morgan: ***Data Mining: Concepts and Techniques***, 2nd Edition, ISBN 978-81-312-0535-8.

Reference Books	• Mallach G Efrem (2009): ***Decision Support and Data Warehouse Systems***, Tata McGraw Hill. • Berry M and . Linoff G, Wiley John (2008): ***Mastering Data Mining: The art and science of customer relationship management***, Ist Edition, Wiley India Pvt. Ltd.

Code	DSE-II	Total Lecture:45 Tutorial: 15
DS20M110	ADVANCED DBMS	3-1-0 = 4

Course Objectives

- Learn the data structures used in the implementation of physical layer of a DBMS.
- To understand the basic concepts and terminology related to DBMS and Relational Database Design.
- Tell how the various relational operators are evaluated in a DBMS.
- Compare the Relational DBMS with Object Databases and Distributed Databases.
- Outline the security and authorization used in a DBMS.
- To understand advanced DBMS techniques to construct tables and write effective queries, forms, and reports.

Unit	Contents	Hours
1	Introduction: Overview of storage and indexing, disks and files:Data on external storage; File organizations and indexing; Index data structures; Comparison offile organizations; Indexes and performance tuning. Memory hierarchy; RAID; Disk spacemanagement; Buffer manager; Files of records; Page formats and record formats.Tree structured indexing: Indexed Sequential Access Method(ISAM).	10
2	Hash Based Indexing: Static hashing; Extendible hashing, Linear hashing, comparisons.External Sorting: When does a DBMS sort data, A simple two-way merge sort; Externalmerge sort, Using B+ trees for sorting.Evaluating Relational Operators.	10
3	Query Optimization: Using Heuristics in Query Optimization, Using selectivity and costestimates in Query Optimization, Overview of Query optimization in Oracle, Semantic Query Optimization.Physical Database Design and Tuning: Clustering and indexing; Indexes that enable index-only plans; Overview of database tuning;Choices in tuning the conceptual schema; Choices in tuning queries and views.	10
4	Object Databases: Concepts for Object Databases: Overview of Object-Oriented Concepts, Object Identity,Object Structure, and Type Constructors, Encapsulation of Operations, Methods, andPersistence, Type and Class Hierarchies and Inheritance, Complex Objects;Object Database Standards, Languages, and Design: Overview of the Object Model ofODMG, The Object Definition Language ODL, The Object Query Language OQL, Object-Relational and Extended-Relational Systems: Overview of SQL and its Object Relational features, Object-Relational Features of Oracle.	8

5	Distributed Databases: Distributed Database concepts; Data Fragmentation, Replication, and Allocation Techniques forDistributed Database Design; Types of Distributed Database Systems; Query Processing inDistributed Databases; Overview of Concurrency Control and Recovery in Distributeddatabases; Distributed databases in Oracle. Security and Authorization:Discretionary Access Control, Mandatory Access Control,Role of the Database Administrator.	7

Course Outcomes

At the end of the course the students should be able to:

CO 1	**Analyze**[4] the data structures used in the implementation of physical layer of a DBMS.
CO 2	**Describe**[1] an optimized DBMS query so that the result will be faster.
CO 3	**Describe**[1] how the various relational operators are evaluated in a DBMS.
CO 4	**Compare**[4] the Relational DBMS with Object Databases and Distributed Databases.
CO 5	**Analyze**[4] the security and authorization used in a DBMS.
Text Books	• Khicha Arihant, Kapoor Neeti **(**2018), **Advance Database Management System**, Genius Publications Pvt. Ltd. • Kim Won (2008), **Introduction to Object-oriented Databases**, The M.I.T. Press.
Reference Books	• Jarke Matthias, "**Query Optimization in Database Systems**", Springer. • Khicha Arihant, Kapoor Neeti **(**2018), **Advance Database Management System**, Genius Publications Pvt. Ltd.

Code	DSE-II	Total Lecture:45 Tutorial: 15
DS20M111	STATISTICAL FOUNDATIONS FOR DATA SCIENCE	3-1-0 = 4

Course Objectives

- The course will introduce students to the fundamental mathematical concepts required for a program in data science.
- Understanding of descriptive statistics by practical application of quantitative reasoning and data visualization.
- Learn the basics of inferential statistics by making valid generalizations from sample data.
- Understand the use of probability in the field of data science.
- Greater appreciation for the importance of statistical literacy in today's data rich world.

Unit	Contents	Hours
1	Basics of Data Science: Introduction; Typology of problems; Importance of linear algebra, statistics and optimization from a data science perspective; Structured thinking for solving data science problems.	10
2	Linear Algebra: Matrices and their properties (determinants, traces, rank, nullity, etc.); Eigenvalues and eigenvectors; Matrix factorizations; Inner products; Distance measures; Projections; Notion of hyperplanes; half-planes.	10
3	Probability, Statistics and Random Processes: Probability theory and axioms; Random variables; Probability distributions and density functions (univariate and multivariate); Expectations and moments; Covariance and correlation; Statistics and sampling distributions; Hypothesis testing of means, proportions, variances and correlations; Confidence (statistical) intervals; Correlation functions; White-noise process.	10
4	Optimization: Unconstrained optimization; Necessary and sufficiency conditions for optima; Gradient descent methods; Constrained optimization, KKT conditions; Introduction to non-gradient techniques; Introduction to least squares optimization; Optimization view of machine learning.	8
5	Introduction to Data Science Methods: Linear regression as an exemplar function approximation problem; Linear classification problems.	7

Course Outcomes	
At the end of the course the students should be able to:	
CO 1	**Describe**[1] basic statistical terms and concepts.
CO 2	**Identify**[4] statistical methods that are suitable for exploring, describing and analyzing data.
CO 3	**Understand**[2] and apply various optimization techniques.
CO 4	To **apply**[3] linear algebra in an efficient manner.
CO 5	To **use**[3] different methods of data science.
Text Books	• Strang G. (2016): **Introduction to Linear Algebra**, Fifth edition, Wellesley-Cambridge Press, USA. • Bendat, J. S. and A. G. Piersol (2010): **Random Data: Analysis and Measurement Procedures**". IV Edition, John Wiley & Sons, Inc. • Montgomery, D. C. and G. C. Runger (2011): **Applied Statistics and Probability for Engineers**, V Edition, John Wiley & Sons, Inc. • David G. Luenberger (1969): **Optimization by Vector Space Methods**, John Wiley & Sons.
Reference Books	• Cathy O'Neil and Rachel Schutt,(2013): **Doing Data Science**", O'Reilly Media.

Code	SOFTWARE LAB-I PYTHON PROGRAMMING	Practical: 30
AI20M104		0– 0–4 = 2

Course Objectives:

- This course introduces core programming basics—including data types, control structures, algorithm development, and program design with functions—via the Python programming language.
- The course discusses the fundamental principles of Object-Oriented Programming
- Learn about data and information processing techniques.
- Students will solve problems and explore real-world software development challenges
- Learn to create practical and contemporary applications.

Unit	Contents	Hours
I.	Introduction to Python Programming Language. : Introduction to Python Language, Strengths and Weaknesses, IDLE, Dynamic Types, Naming Conventions, String Values, String Operations, String Slices, String Operators, Numeric Data Types, Conversions, Built In Functions	5
II.	Data Collections and Language Component: Introduction, Control Flow and Syntax, Indenting, The if Statement, Relational Operators, Logical, Operators, True or False, Bit Wise Operators, The while Loop, break and continue, The for Loop, Lists, Tuples, Sets, Dictionaries, Sorting Dictionaries, Copying Collections.	10
III.	Object and Classes : Classes in Python, Principles of Object Orientation, Creating Classes, Instance Methods, File Organization, Special Methods, Class Variables, Inheritance, Polymorphism, Type Identification, Custom Exception Classes	10
IV.	Functions and Modules : Introduction, Defining Your Own Functions, Parameters , Function Documentation, Keyword and Optional Parameters, Passing Collections to a Function, Variable Number of Arguments, Scope, Functions - "First Class Citizens", Passing Functions to a Function, Mapping Functions in a Dictionary, Lambda, Modules, Standard Modules – sys, Standard Modules – math, Standard Modules – time, The dir Function	10
V.	I/O and Error Handling In Python : Introduction, Data Streams, Creating Your Own Data Streams, Access Modes, Writing Data to a File, Reading Data From a File, Additional File Methods, Using Pipes as Data Streams, Handling IO	10

	Exceptions, Working with Directories, Metadata, Errors, Run Time Errors, The Exception Model, Exception Hierarchy, Handling Multiple Exceptions	

Course Outcomes as per Bloom's Taxonomy	
At the end of the course the students should be able to:	
CO 1	**Apply** the principles python programming
CO 2	**Write** clear and effective python code
CO 3	**Create** applications using python programming
CO 4	**Access** database using python programming
CO 5	**Develop** web applications using python programming
Text Books	• Dive into Python, Mike • Learning Python, 4th Edition by Mark Lutz • Programming Python, 4th Edition by Mark Lutz • Kenneth A. Lambert, The Fundamentals of Python: First Programs, 2011, Cengage Learning, ISBN: 978-1111822705.
Reference Books	• Starting Out with Python (2009) Pearson , Tonny Gaddis • Beginning Pyhton Wrox Publication Peter Norton, Alex Samuel • Python Algorithms Apress, Magnus Liet Hetland • Python Object Oriented Programming PACKT Press, Dusty Phillips • Python for Unix and Linux System Administration O'Relly, Noad Gift

COURSE CODE	PROJECT BASED LEARNING-I	Total Lecture:30 Practical:30
PB20M101		**(LTP=0-0-4=2)**
Learning Objectives:	• Integrating the knowledge and skills of various courses on the basis of multidisciplinary projects • Develop the skill of critical thinking and evaluation. • To develop 21st century success skills such as critical thinking, problem solving, communication, collaboration and creativity/innovation among the students. • To enhance deep understanding of academic, personal and social development in students. • Employ the specialized vocabularies and methodologies.	
Course Outcome		
At the end of the course the students will be able to:		
Course Outcomes:	• **Apply**[3]a sound knowledge/skills to select and develop their topicand project respectively. • **Develop**[6] plans and allocate roles with clear lines of responsibilityand accountability. • **Design**[6] solutions to complex problems following a systematic approach like problem identification, formulation and solution. • **Collaborate**[6] with professionals and the community at large in written and in oral forms. • **Correlate**[4]the knowledge, skills and attitudes of a professional.	
General Guidelines:	• PBL will be an integral part of UG/PG Programs at different levels. • Each semester offering PBL will provide a separate Course Code, twocredits will be allotted to it. • Faculty will be assigned as mentor to a group of 30 students minimum byHoS. • Faculty mentor will have 4 hours/week to conduct PBL for assigned students. • Student will select a topic of their choice from syllabus of any course offered in respective semester (in-lines with sustainable development goals). • Student may work as a team maximum 3 or minimum 2 members for single topic. • For MSE, student's performance will be assessed by panel of three experts either from other department/school, or from same department/school based on chosen topic. This will be comprised of apresentation by student followed by viva-voce. It will be evaluated for 30 marks. • 20 marks would be allotted for continuous performance assessment by concerned guide/mentor. For ESE, student will need to submit a project report in prescribed format,duly signed by concerned guide/mentor and head of the school. The report should be	

	comprised of following components: 1. Introduction 2. Review of literature 3. Methodology 4. Result and Discussion 5. Conclusion and Project Outcomes 6. References • Student will need to submit three copies for 1. Concerned School 2.Central Library 3. Self • The integrity of the report should be maintained by student. Any malpractice will not be entertained. • Writing Ethics to be followed by student, a limit of 10 % plagiarism is permissible. Plagiarism report is to be attached along with the report. • Project could be a case study/ analytical work /field work/ experimentalwork/ programming or as per the suitability of the program.

SANJEEV AGRAWAL GLOBAL EDUCATIONAL (SAGE) UNIVERSITY, BHOPAL

Syllabus

for

Master of Technology (Data Science)

II Semester

School of Advanced Computing

COURSE CODE	DSE-III NATURAL LANGUAGE PROCESSING	Total Lecture : 60 Theory : 45 Practical : 15
DS20M201		(LTP= 3 – 0 – 2 = 4)

Course Objectives:

- Students will learn how to process written text from basic of fundamental knowledge starts with Finite automata, Regular expression and probabilistic model with n-grams.
- This NLP course will boost student knowledge to research level where they can conduct new level of research. It really helpful for undergraduate students.
- To get introduced to language processing technologies for processing the text data.
- To acquire knowledge on text data analytics using language model.
- To understand the role of Information Retrieval and Information Extraction in Text Analytics.

UNIT	CONTENTS	HOURS
I.	Introduction: Natural Language Processing tasks in syntax, semantics, and pragmatics, Issues, Applications. The role of machine learning, Probability Basics ,Information theory ,Collocations Ngram Language Models, Estimating parameters and smoothing - Evaluating language models.	10
II.	Morphology and part of speech tagging: Linguistic essentials, Lexical syntax, Morphology and Finite State Transducers, Part of speech Tagging, Rule-Based Part of Speech Tagging, Markov Models, Hidden Markov Models, transformation based Models, Maximum Entropy Models, Conditional Random Fields.	10
III.	Syntax Parsing: Syntax Parsing, Grammar formalisms and tree banks, Parsing with Context Free Grammars, Features and Unification, Statistical parsing and probabilistic CFGs (PCFGs), Lexicalized PCFGs.	10
IV.	Semantic Analysis: Representing Meaning, Semantic Analysis, Lexical semantics, Word sense disambiguation, Supervised, Dictionary based and Unsupervised Approaches Compositional semantics, Semantic Role Labeling and Semantic Parsing, Discourse Analysis.	8

V.	NLP and Information retrieval and ranked information retrieval, semantics, word senses, word similarity, word relations, word net and other thesauri, automatic text summarization, summarizing single document, summarizing multiple documents, question answering systems.	7

Course Outcomes as per Bloom's Taxonomy	
At the end of the course the students should be able to:	
CO 1	**Understand**[2] Natural Language Processing.
CO 2	Probabilistic model of **defining**[1] language and techniques.
CO 3	**Applying**[3] Hidden Markov model and Speech Recognition.
CO 4	**Understand**[2] application of context free grammar and language parsing.
CO 5	Make use of NLP concepts to **solve**[3] Information retrieval problems.
Text Books	• Jurafsky Daniel and Martin H. James (2008): ***Speech and Language Processing***, IInd Edition, Prentice Hall. • Aggarwal C. Charu (2018): ***Machine Learning for Text***, Ist Edition Springer. • D.Manning Christopher and Schuetze Hinrich (1999): ***Foundations of Statistical Natural Language Processing***, MIT press. • Bird Steven, Klein Ewan and Loper Edward (2009): ***Natural Language Processing with Python***, Ist Edition, O'Reilly Media.
Reference Books	• R.Hausser Roland (2011): ***Foundations of Computational Linguistics:Human-Computer Communication in Natural Language***, Paperback, MIT press.

COURSE CODE	DEEP LEARNING	Total Lecture:60 Theory: 45 Practical:15
AI20M202		(LTP= 3 – 0 – 2 = 4)

Course Objectives:

- Understand and learn to implement them in Keras and TensorFlow.
- Build convolutional networks for image recognition, recurrent networks for sequence generation, generative adversarial networks for image generation, and more.
- Understand the implementation of RNN.
- Understand the methods & implementation of CNN.
- Become an expert in neural networks.

UNIT	CONTENTS	HOURS
I.	Introduction To Neural Networks, Implementing Gradient Descents, Training Neural Networks, Sentiment Analysis, Keras, Tensorflow.	10
II.	Convolutional Neural Network, CNNs in Tensorflow, Weight Initialization, Autoencoders, Transfer Learning In Tensorflow, Deep Learning For Cancer Detection.	10
III.	Recurrent Neural Networks, Long Short-Term Memory Network, Implementation Of RNN And LSTM, Hyperparameters, Embeddings And Word2vec, Sentiment Prediction RNN.	10
IV.	Generative Adversarial Network, Deep Convolutional GANs, Generate Faces, Semi-Supervised Learning, The RL Framework: The Problem, The RL Framework: The Solution, Dynamic Programming.	8
V.	Monte Carlo Methods, Temporal-Difference Methods, RL In Continuous Spaces, Deep Q-Learning, Policy Gradients, Actor-Critic Methods.	7

Course Outcomes as per Bloom's Taxonomy

At the end of the course the students should be able to:

CO 1	To develop a basic **understanding**[2] of the building blocks of AI as presented in terms of intelligent agents.
CO 2	To **choose**[4] an appropriate problem-solving method and knowledge-representation scheme.

CO 3	Develop an ability to **analyze**[4] and formalize the problem (as a state space, graph, etc.) and select the appropriate search method.
CO 4	To **develop/demonstrate**[2]**/** build simple intelligent systems
CO 5	To **develop/demonstrate**[2]**/** build various classical to y problems using different AI techniques.
Text Books	• Stuart Russell and Peter Norvig: ***Artificial Intelligence: A Modern Approach***, IInd Edition, Pearson Education. • Elaine Rich, Kevin Knight, Nair B Shivshankar: ***Artificial Intelligence***, McGraw Hill, IIIrd Edition. • Elaine Rich, Knight Kevin: ***Artificial Intelligence***, IInd Edition, Tata McGraw Hill.
Reference Books	• Lugar George, (2002): ***AI-Structures and Strategies for Complex Problem Solving***, IVth Edition, Pearson Education. • Nilsson J Nils: ***Principles of Artificial Intelligence***, Narosa Publication. • Patrick H Winston: ***Artificial Intelligence***, IIIrd edition, Pearson Education. • Khemani Deepak (2013): ***A First Course in Artificial Intelligence***, Ist Edition, McGraw Hill Publication.

COURSE CODE	BIG DATA ANALYTICS	Total Lecture : 60 Theory : 45 Practical : 15
AI20M203		(LTP= 3 – 0 – 2 = 4)

Course Objectives:

- To understand Big Data Analytics for different systems like Hadoop.
- To learn the design of Hadoop File System.
- To learn how to analyze Big Data using different tools.
- To understand the importance of Big Data in comparison with traditional databases.
- Understand the concept of Hive Shell.

UNIT	CONTENTS	HOURS
I.	Introduction To Big Data And Hadoop: About database analytics,Database, Design, Model, Functions, Tools. Types of Digital Data, Introduction to Big Data, Big Data Analytics, History of Hadoop, Apache Hadoop, Analysing Data with Unix tools, Analysing Data with Hadoop, Hadoop Streaming, Hadoop Echo System, IBM Big Data Strategy, Introduction to Infosphere BigInsights and Big Sheets.	10
II.	HDFS(Hadoop Distributed File System): The Design of HDFS, HDFS Concepts, Command Line Interface, Hadoop file system interfaces, Data flow, Data Ingest with Flume and Scoop and Hadoop archives, Hadoop I/O: Compression, Serialization, Avro and File-Based Data structures.	10
III.	Map Reduce Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features.	10
IV.	Hadoop Eco System Pig: Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators. Hive: Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, Querying Data and User Defined Functions. Hbase HBasics, Concepts, Clients, Example, Hbase Versus RDBMS. Big SQL : Introduction	8
V.	Data Analytics with R Machine Learning: Introduction, Supervised Learning, Unsupervised Learning, Collaborative Filtering. Big Data Analytics with BigR.	7

Course Outcomes as per Bloom's Taxonomy

At the end of the course the students should be able to:	
CO 1	To gain **knowledge**[1] about working of Hadoop File System.
CO 2	Ability to **analyze**[4] Big Data using different tools.
CO 3	**Build**[4] a complete business data analytics solution
CO 4	**Design**[6] efficient algorithms for mining the data from large volumes
CO 5	Ability to **analyze**[4] Hive Shell.
Text Books	• White Tom (2012): ***Hadoop: The Definitive Guide***, IIIrd Edition, O'Reilly Publications. • De-Roos Dirk, Eaton Chris, Lapis George, Zikopoulos Paul, Deutsch Tom (2012): ***Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data***, Ist Edition, TMH.
Reference Books	• Marconi Katherine Hardcover, Lehmann Harold: ***Big Data and Health Analytics*** • Baesens Bart: ***Analytics in a Big Data World: The Essential Guide to Data Science and its Applications***, Wiley Publications.

COURSE CODE	DSE-III	Total Lecture:45 Tutorial: 15
DS20M207	GPU COMPUTING	3–1–0 = 4

Course Objectives

- To learn parallel programming with Graphics Processing Units (GPUs).
- To understand the basics of GPU architectures.
- To write programs for massively parallel processors.
- To understand the issues in mapping algorithms for GPUs.
- To introduce different GPU programming models

Unit	Contents	Hours
I.	Introduction: History, Graphics Processors, Graphics Processing Units, GPGPUs. Clock speeds, CPU / GPU comparisons, Heterogeneity, Accelerators, Parallel programming, CUDA OpenCL / OpenACC, Hello World Computation Kernels, Launch parameters, Thread hierarchy, Warps / Wavefronts, Thread blocks / Workgroups, Streaming multiprocessors, 1D / 2D / 3D thread mapping, Device properties, Simple Programs.	10
II.	Memory: Memory hierarchy, DRAM / global, local / shared, private / local, textures, Constant Memory, Pointers, Parameter Passing, Arrays and dynamic Memory, Multi-dimensional Arrays, Memory Allocation, Memory copying across devices, Programs with matrices, Performance evaluation with different memories.	10
III.	Synchronization: Memory Consistency, Barriers (local versus global), Atomics, Memory fence. Prefix sum, Reduction. Programs for concurrent Data Structures such as Work lists, Linked-lists. Synchronization across CPU and GPU Functions: Device functions, Host functions, Kernels functions, Using libraries (such as Thrust), and developing libraries.	10
IV.	Support: Debugging GPU Programs. Profiling, Profile tools, Performance aspects Streams: Asynchronous processing, tasks, Task-dependence, Overlapped data transfers, Default Stream, Synchronization with streams. Events, Event-based Synchronization - Overlapping data transfer and kernel execution, pitfalls.	8
V.	Advanced topics: Dynamic parallelism, Unified Virtual Memory, Multi-GPU processing, Peer access, Heterogeneous processing.	7

Course Outcomes as per Bloom's Taxonomy	
At the end of the course the students should be able to:	
CO 1	**Understand**[2] concepts in parallel programming.
CO 2	**Implement**[6] efficient algorithms in GPUs for common application kernels.
CO 3	**Demonstrate**[3] debugging programs.
CO 4	**Understand**[2] profiling parallel programs.
CO 5	**Identify**[4] efficient parallel programming patterns to solve problem
Text Books	• Cook Shane (2012), **CUDA Programming: —A Developer's Guide to Parallel Computing with GPUs (Applications of GPU Computing)",** Ist Edition, Morgan Kaufmann. • Kaeli David R., Perhaad, Mistry, Dana Schaa, Dong Ping Zhang (2015), **Heterogeneous computing with OpenCL**, IIIrd Edition, Morgan Kauffman.
Reference Books	• David Kirk, Wen-meiHwu (2010), **Programming Massively Parallel Processors: A Hands-on Approach**, Morgan Kaufman, (ISBN: 978-0123814722). • Cook Shane (2012), **CUDA Programming: —A Developer's Guide to Parallel Computing with GPUs (Applications of GPU Computing)",** Ist Edition, Morgan Kaufmann.

Code	DSE-III	Total Lecture:45 Practical:15
DS20M208	BIOINFORMATICS	3 – 0–2 = 4

Course Objectives

- To learn how Bioinformatics is storing, extracting, organizing, analyzing, interpreting and using information.
- The approaches to the discipline of bioinformatics incorporate expertise from the biological sciences, computer science and mathematics.
- The major in bioinformatics is designed for students interested in molecular biology and genetics, information technologies and computer science.
- To learn how Bioinformatics are involved in the analysis of the human genome
- tudy of structural and functional relationships, and molecular evolution.

Unit	Contents	Hours
I.	Introduction, chronological history of Bioinformatics, evolution of Bioinformatics, Objectives of Bioinformatics, Importance of bioinformatics, Bioinformatics in business, future scope of Bioinformatics.	10
II.	Bioinformatician and bioinformaticist, role, need and importance of Biology, Computer Science, mathematics and information technology in bioinformatics, biological classification and nomenclature, life in space and time.	10
III.	Introduction, information networks, protein and genome information resources, DNA sequence analysis, pairwise alignment techniques, multiple alignment techniques, secondary databases, analysis packages.	10
IV.	The dawn of sequencing, the biological sequence or structure deficit, human genome project and its status, homology and analogy, web browsers, European molecular biology network, national centre for biotechnological information, specialized genomic resources.	8
V.	Building a sequence search protocol, practical approach for structural and functional interpretation, introduction to analysis package, commercial databases, softwares and comprehensive packages, internet packages specializing in DNA and protein analysis.	7

Course Outcomes as per Bloom's Taxonomy

At the end of the course the students should be able to:

CO 1	**Understand**[2] the basic principles and concepts of biology, computer science and mathematics.
CO 2	**Use**[3] existing software effectively to extract information from large databases and to use this information in computer modeling.
CO 3	**Develop**[6] problem-solving skills, including the ability to develop new algorithms and analysis methods.
CO 4	**Develop**[6] an understanding of the intersection of life and information sciences.
CO 5	**Understand**[2] how the bioinformatics problems can be solved by computer science.
Text Books	• T.K. Attwood and Parry Smith (1999), **Introduction to Bioinformatics**, Prentice Hall. • Arthur M. Lesk, 2013, "**Introduction to Bioinformatics**", IVth Edition, OUP Oxford. • Krane and Raymer, 2005, **Fundamental Concepts in Bioinformatics** , San Val.

Code	DSE-III	Total Lecture:45 Tutorial: 15
DS20M209	OPTIMIZATION TECHNIQUES	3-1-0 = 4

Course Objectives

- Learn basic concepts of linear programming
- Introduction to optimization techniques using both linear and non-linear programming.
- Understand the concepts of queuing theory
- The focus of the course is on convex optimization though some techniques will be covered for non-convex function optimization too.
- After an adequate introduction to linear algebra and probability theory, students will learn to frame engineering minima maxima problems in the framework of optimization problems.

Unit	Contents	Hours
I.	Linear Programming, Mathematical Model, Assumptions of Linear Programming, simplex Method, Degeneracy, Applications, Duality, Dual Simplex Method, & Algorithm Assignment Problem. Hungarian Method & its Algorithm.	10
II.	Transportation Problem, Integer Programming: - Gomorra's method, Branch and Bound techniques. Integer Programming Algorithm, Dynamic Programming:- Bellman's Principle of optimality, Dynamic Programming Approach, optimal subdivision problem, Decomposition, Applications in linear programming.DPP Algorithms.	10
III.	Queuing Theory : Queuing problem and system, Transient and steady state distributions in queuing system, Poisson process, Exponential process, classification of queuing models, Model I (M/M/1) : (∝/FCFS), Model -II General Erlang Queuing model, Model - III (M/M/1) : (N/FCFS), Model - IV (M/M/S) : (∝/FCFS), Algorithms.	10
IV.	Inventory Theory :- Basic concepts, classification of Inventory systems & models, Economic order quantity, Deterministic Inventory models :- EoQ Models without shortages, EoQ Models with shortages, Probabilistic Inventory Models with instantaneous demand, no set up cost model, Discrete and continuous cases.	8
V.	Graph Theory: Introduction, Digraphs, Paths and Subgraphs, Bipartite Graphs. Planar Graphs. Euler's Formula. Platonic Graphs. Coloring. Connectivity of graphs, Tours and Matchings, graphs on surfaces, directed graphs and undirected graphs, Flows, Random Graphs.	7

Course Outcomes	
At the end of the course the students should be able to:	
CO 1	**Understand**[2] and apply the concept of optimality criteria for various type of optimization problems.
CO 2	**Apply**[3] basic concepts of mathematics to formulate an optimization problem.
CO 3	**Analyze**[4] and appreciate variety of performance measures for various optimization problems.
CO 4	**Apply**[3] the methods of optimization in real life situation.
CO 5	**Solve**[3] various constrained and unconstrained problems in Single variable as well as multivariable.
Text Books	• S.D. Sharma, 2012, "Operations Research," 2020th Edition, Kedarnath Ramnath & Co. Meerut. • P.K.Gupta & D.S.Hira, 1976, "Operations Research",Vth Edition, S.Chand & Co. • Kanti swarup , 2014, "Operations Research",Ist Edition, S.Chand & Sons. • Gillet, 1979," Introduction to operations Research - A Computer Algorithm Approach", New Ist Edition, McGraw Hill. • Hillier, 2017 "Introduction to operations Research, Xth Edition", McGraw Hill Education, TMH. • Reinhard Diestel, 2010, "Graph Theory (Graduate Text in Mathematics)", IVth Edition, Springer.

Code	DSE-IV	Total Lecture:45 Tutorial: 15
DS20M210	GEOSPATIAL INFORMATION MANAGEMENT	3-1-0 = 4

Course Objectives

The student should be made to:

- Get the knowledge of Geospatial Analysis. Basic concept of spatial analysis.
- Connectivity of network. Surface Analysis.

Unit	Contents	Hours
I.	Introduction to Spatial Analysis: Significance of Spatial Analysis. Overview of Tools For Analysis Spatial Analysis - Vector Based: Overlay Operations: Point-In-Polygon, Line-In-Polygon, Polygon-In-Polygon. Single Layer Operations: Feature Identification, Extraction, Classification Manipulation. Multilayer Operation: Union, Intersection, Symmetrical Difference, Update, Merge, Append and Dissolve.	10
II.	Spatial Analysis - Raster Based: Map Algebra, Grid Based Operations, Local, Focal, Zonal and Global Functions, Cost Surface Analysis, Optimal Path and ProximitySearch, Network Analysis: Concepts, Evaluation of Network Complexity Using Alpha-Gamma Indices. C-Matrices for Evaluating.	10
III.	Connectivity of the Network. Network Data Model. Path Analysis. Linear Referencing and Segmentation. Types of Network Analysis: Optimum Cyclic Path, Vehicle Routing, Path Determination and Cost-Path Analysis. Geocoding	10
IV.	Surface Analysis: Interpolation Methods: Trend Surface Analysis, IDW, Kriging, Measures of Arrangement and Dispersion, Autocorrelation, Semi-Variogram, DEM, TIN, Slope, Aspect, Hillshade and Viewshed.	8
V.	Spatial Modeling: Role of Spatial Model, Explanative, Predictive and Normative Models. Correlation-Regression Analysis in Model Building. Handling Complex Spatial Query and Case Studies Introduction to Spatial Analysis using ‘RUnit 2	7

Course Outcomes

At the end of the course the students should be able to:

CO 1	**Demonstrate**[6] Introduction of Spatial Analysis.
CO 2	**Understand**[2] significance of Spatial Analysis.
CO 3	**Demonstrate**[6] connectivity of the Network.
CO 4	**Perform**[3] surface Analysis.
CO 5	**Perform**[3] spatial Modeling.
Text Books	• Demers, M. N. (2000), **Fundamentals of Geographic Information Systems**, John Wiley and Sons, New Delhi. • Burrough, P. A. and McDonnell, R. A. (2000), **Principles of Geographical Information Systems**, Oxford University Press, New York. • Makrewski, J. (1999), **GIS Multi-criteria Analysis**, John Wiley and Sons, New York.
Reference Books	• Chang, K. T. (2008), **Introduction to Geographic Information Systems**, Avenue of the Americas, McGraw-Hill, New York. • Longley, P. A., Good child, M. F., Maguire, D. J. Rhind, D. W. (2002), **Geographical Information Systems and Science**, John Wiley & Sons, Chichester. • Lo, C. P. Yeung, A. W. (2002), **Concepts Techniques of Geographical Information Systems**, Prentice-Hall of India, New Delhi.

Code	DSE-IV	Total Lecture:45 Tutorial: 15
DS20M211	INFORMATION RETRIEVAL	3 – 1-0 = 4

Course Objectives

The Student should be made to:

- Learn the information retrieval models. Be familiar with Web Search Engine. Be exposed to Link Analysis.
- Understand Hadoop and Map Reduce.
- Learn document text mining techniques.

Unit	Contents	Hours
I.	Introduction: Introduction -History of IR- Components of IR - Issues –Open source Search engine Frameworks - The impact of the web on IR - The role of artificial intelligence (AI) in IR – IR Versus Web Search - Components of a Search engine-Characterizing the web...	10
II.	Information Retrieval: Boolean and vector-space retrieval models- Term weighting - TF-IDF weighting- cosine similarity – Preprocessing - Inverted indices - efficient processing with sparse vectors – Language Model based IR - Probabilistic IR –Latent Semantic Indexing - Relevance feedback and query expansion.	10
III.	Web Search Engine – Introduction And Crawling Web search overview, web structure, the user, paid placement, search engine optimization/ spam. Web size measurement - search engine optimization/spam – Web Search Architectures - crawling - meta-crawlers- Focused Crawling - web indexes –- Near-duplicate detection - Index Compression - XML retrieval. Cross Lingual Retrieval	10
IV.	Web Search – Link Analysis And Specialized Search : Link Analysis –hubs and authorities – Page Rank and HITS algorithms -Searching and Ranking – Relevance Scoring and ranking for Web – Similarity - Hadoop & Map Reduce - Evaluation - Personalized search - Collaborative filtering and content-based recommendation of documents and products – handling “invisible” Web - Snippet generation, Summarization, Question Answering.	8
V.	Document Text Mining: Information filtering; organization and relevance feedback – Text Mining -Text classification and clustering - Categorization algorithms: naive Bayes; decision trees; and nearest neighbor - Clustering algorithms: agglomerative clustering; k-means; expectation maximization (EM)	7

Course Outcomes	
At the end of the course the students should be able to:	
CO 1	**Understand**[2] statistical models of text can be used to solve problems in IR.
CO 2	**Understand**[2] statistical models of text can be used for other IR applications.
CO 3	**Understand**[2] Web search engine and crawling.
CO 4	**Understand**[2] the link analysis and specialized search.
CO 5	**Understand**[2] Text classification and clustering.
Text Books	• C. Manning, P. Raghavan, and H. Schütze (2008), **Introduction to Information Retrieval** , Cambridge University Press. • Bruce Croft, Donald Metzler and Trevor Strohman (2009), "Search Engines: Information Retrieval in Practice", Ist Edition Addison Wesley. • Mark Levene (2010) ,**An Introduction to Search Engines and Web Navigation**, IInd, Edition Wiley.
Reference Books	• Stefan Buettcher, Charles L. A. Clarke, Gordon V. Cormack (2010), **Information Retrieval: Implementing and Evaluating Search Engines**, The MIT Press. • Ophir Frieder (2004), **Information Retrieval: Algorithms and Heuristics: The Information Retrieval Series**, 2nd Edition, Springer. • Manu Konchady (2008), **Building Search Applications: Lucene, Ling Pipe**, First Edition, Gate Mustru Publishing,.

COURSE CODE	DSE-IV SEMANTIC WEB	Total Lecture : 60 Theory : 45 Tutorial : 15
AI20M212		(LTP= 3 – 1– 0 = 4)

Course Objectives:

- To learn Web Intelligence.
- To learn Knowledge Representation for the Semantic Web.
- To learn Ontology Engineering.
- To learn Semantic Web Applications, Services and Technology.
- To learn Social Network Analysis and semantic web.

UNIT	CONTENTS	HOURS
I.	Web Intelligence: Thinking and Intelligent Web Applications, The Information Age ,The World Wide Web, Limitations of Today's Web, The Next Generation Web, Machine Intelligence, Artificial Intelligence, Ontology, Inference engines, Software Agents, Berners-Lee www, Semantic Road Map, Logic on the semantic Web.	10
II.	Knowledge Representation for the Semantic Web: Ontologies and their role in the semantic web, Ontologies Languages for the Semantic Web –Resource Description Framework(RDF) / RDF Schema, Ontology Web Language(OWL), UML,XML/XML Schema.	10
III.	Ontology Engineering: Ontology Engineering, Constructing Ontology, Ontology Development Tools, Ontology Methods, Ontology Sharing and Merging, Ontology Libraries and Ontology Mapping, Logic, Rule and Inference Engines.	10
IV.	Semantic Web Applications, Services and Technology: Semantic Web applications and services, Semantic Search, e-learning, Semantic Bioinformatics, Knowledge Base ,XML Based Web Services, Creating an OWL-S Ontology for Web Services, Semantic Search Technology, Web Search Agents and Semantic Methods.	8
V.	Social Network Analysis and semantic web: What is social Networks analysis, development of the social networks analysis, Electronic Sources for Network Analysis – Electronic Discussion networks, Blogs and Online Communities, Web Based Networks. Building Semantic Web Applications with social network features.	7

Course Outcomes as per Bloom's Taxonomy

At the end of the course the students should be able to:	
CO 1	**Understand**[2] the concept structure of the semantic web technology.
CO 2	**Understand**[2] the concepts of Web Science, semantics of knowledge and resource, ontology.
CO 3	**Understand**[2] logic semantics and inference with OWL.
CO 4	**Applying**[3] ontology engineering approaches in semantic applications
CO 5	**Experimenting**[3] Web graph processing for various applications.
Text Books	• Lee Berners, Godel and Turing. (2008): ***Thinking on the Web***, Wiley inter science. • Mika Peter (2007): ***Social Networks and the Semantic Web***, Springer.
Reference Books	• Davies J, Studer R, Warren P: ***Semantic Web Technologies, Trends and Research in Ontology Based Systems***, John Wiley & Sons. • Lu. Liyang (2017): ***Semantic Web and Semantic Web Services***, Chapman & Hall/CRC Publishers (Taylor & Francis Group). • Stucken schmidt Heiner, Van Harmelen Frank: **Information sharing on the semantic Web**, Springer Publications. • Segaran T, Evans C, Taylor J: ***Programming the Semantic Web***, O'Reilly, SPD.

Code	SOFTWARE LAB-II R PROGRAMMING	Practical: 30
AI0M204		0– 0–4 = 2

Course Objective

- To learn how to program in R
- To learn how to use R for effective data analysis.
- You will learn how to install and configure software necessary for a statistical programming environment.
- The course covers practical issues in statistical computing which includes programming in R, reading data into R, accessing R packages, writing R functions, debugging, and organizing and commenting R code.

Unit	Contents	Hours
I.	Introduction: Introducing to R, R Data Structures, Help functions in R, Vectors, Scalars, Declarations, recycling, Common Vector operations, Using all and any, Vectorized operations, NA and NULL values, Filtering, Vectorized if-then else, Vector Equality, Vector Element names	5
II.	Matrices, Arrays And Lists: Creating matrices, Matrix operations, Applying Functions to Matrix Rows and Columns, Adding and deleting rows and columns, Vector/Matrix Distinction, Avoiding Dimension Reduction, Higher Dimensional arrays, lists, Creating lists, General list operations, Accessing list components and values, applying functions to lists, recursive lists	10
III.	Data Frames: Creating Data Frames, Matrix-like operations in frames, Merging Data Frames, Applying functions to Data frames, Factors and Tables, factors and levels, Common functions used with factors, Working with tables, Other factors and table related functions, Control statements, Arithmetic and Boolean operators and values, Default values for arguments, Returning Boolean values, functions are objects, Environment and Scope issues, Writing Upstairs, Recursion, Replacement functions, Tools for composing function code, Math and Simulations in R	10
IV.	OOP: S3 Classes, S4 Classes, Managing your objects, Input/Output, accessing keyboard and monitor, reading and writing files, accessing the internet, String Manipulation, Graphics, Creating Graphs, Customizing Graphs, Saving graphs to files, Creating three-dimensional plots	10
V.	Interfacing: Interfacing R to other languages, Parallel R, Basic Statistics, Linear Model, Generalized Linear models, Non-linear models, Time Series and Auto-correlation, Clustering	10

Course Outcomes as per Bloom's Taxonomy	
At the end of the course the students should be able to:	
CO 1	**Understand** the basics in R programming in terms of constructs, control statements, string functions
CO 2	**Understand** the use of R for Big Data analytics
CO 3	**Create** applications using R programming
CO 4	**Learn** to apply R programming for Text processing
CO 5	**Apply** the R programming from a statistical perspective
Text Books	• Norman Matloff , "The Art of R Programming: A Tour of Statistical Software Design", No Starch Press, 2011 • Jared P. Lander, "R for Everyone: Advanced Analytics and Graphics", Addison-Wesley Data & Analytics Series, 2013.
Reference Books	• Mark Gardener, " Beginning R – The Statistical Programming Language", Wiley, 2013 • Robert Knell, "Introductory R: A Beginner's Guide to Data Visualisation, Statistical Analysis and Programming in R", Amazon Digital South Asia Services Inc, 2013.

COURSE CODE	PROJECT BASED LEARNING-II	Total Lecture:30 Practical:30
PB20M201		(LTP=0-0-4=2)
Learning Objectives:	• Integrating the knowledge and skills of various courses on the basis of multidisciplinary projects • Develop the skill of critical thinking and evaluation. • To develop 21st century success skills such as critical thinking, problem solving, communication, collaboration and creativity/innovation among the students. • To enhance deep understanding of academic, personal and social development in students. • Employ the specialized vocabularies and methodologies.	
Course Outcome		
At the end of the course the students will be able to:		
Course Outcomes:	• **Apply**[3]a sound knowledge/skills to select and develop their topicand project respectively. • **Develop**[6] plans and allocate roles with clear lines of responsibilityand accountability. • **Design**[6] solutions to complex problems following a systematic approach like problem identification, formulation and solution. • **Collaborate**[6] with professionals and the community at large in written and in oral forms. • **Correlate**[4]the knowledge, skills and attitudes of a professional.	
General Guidelines:	• PBL will be an integral part of UG/PG Programs at different levels. • Each semester offering PBL will provide a separate Course Code, twocredits will be allotted to it. • Faculty will be assigned as mentor to a group of 30 students minimum byHoS. • Faculty mentor will have 4 hours/week to conduct PBL for assigned students. • Student will select a topic of their choice from syllabus of any course offered in respective semester (in-lines with sustainable development goals). • Student may work as a team maximum 3 or minimum 2 members for single topic. • For MSE, student's performance will be assessed by panel of three experts either from other department/school, or from same department/school based on chosen topic. This will be comprised of apresentation by student followed by viva-voce. It will be evaluated for 30 marks. • 20 marks would be allotted for continuous performance assessment by concerned guide/mentor. For ESE, student will need to submit a project report in prescribed format,duly signed by concerned guide/mentor and head of the school. The report should be	

	comprised of following components: 1. Introduction 2. Review of literature 3. Methodology 4. Result and Discussion 5. Conclusion and Project Outcomes 6. References • Student will need to submit three copies for 1. Concerned School 2.Central Library 3. Self • The integrity of the report should be maintained by student. Any malpractice will not be entertained. • Writing Ethics to be followed by student, a limit of 10 % plagiarism is permissible. Plagiarism report is to be attached along with the report. • Project could be a case study/ analytical work /field work/ experimentalwork/ programming or as per the suitability of the program.

SANJEEV AGRAWAL GLOBAL EDUCATIONAL (SAGE) UNIVERSITY, BHOPAL

Syllabus

for

Master of Technology (Data Science)

III & IV Semester

School of Advanced Computing

COURSE CODE	MOOC-1/ MOOC-2	Total Lecture: Practical:60
		(LTP=0-0-8=4)
Learning Objective:	• Integrating the knowledge and skills of various courses available in online mode. • Develop the skills of critical thinking and evaluation. • To make students to learn themselves by choosing the course as per there area of interest.	
	CONTENTS	**HOURS**
General Guidelines:	• This course creates an excellent opportunity for students to acquire the necessary skill set for research, employability through massive open online courses (MOOCs) where the rare expertise of world famous experts from academics and industry are available. • The basket for MOOCs will be a dynamic one, as courses keep on updating with time. • In this semester 8 credits will have to be acquired with online courses (MOOCs). Students will have to complete 2 MOOC's of their choice in the third semester. • The MOOC-1 and MOOC-2 each carries internal marks of 50, which will be attained after he/she gets the MOOC certificate for which he/she got himself/herself enrolled. For end sem evaluation a Viva-Voce examination shall be conducted and it will carrie 50 marks.	60

GUIDELINES FOR M. TECH. DISSERTATION/ THESIS

Phase-1

- Every candidate shall be required to submit a thesis or dissertation on a topic approved by the Dissertation Review Committee (DRC).
- A Dissertation Review Committee shall be constituted with the Head of the Department as Chairperson, Dissertation Supervisor and one senior faculty member of the Department offering the M. Tech. programme.
- Candidate has to present in Dissertation Work Review I, in consultation with his Dissertation Supervisor, the title, objective and plan of action of his dissertation work to the Dissertation Work Review Committee (DRC) for approval within four weeks from the commencement of Second year First Semester. The Dissertation Work Review I carries internal marks of 100. Evaluation should be done by the DRC for 50 marks and the Supervisor will evaluate the review for the other 50 marks. Only after obtaining the approval of the DRC can the student initiate the Dissertation work.
- If a candidate wishes to change his/her supervisor or topic of the dissertation, he/she can do so with the approval of the DRC. However, the DRC shall examine whether or not the change of topic/supervisor leads to a major change of his initial plans of dissertation proposal. If yes, his/her date of registration for the dissertation work starts from the date of change of Supervisor or topic as the case may be.
- A candidate shall submit his dissertation progress report in two stages at least with a gap of three months between them.
- The work on the dissertation shall be initiated at the beginning of the II year and the duration of the dissertation is two semesters. A candidate is permitted to submit thesis only after successful completion of all theory and practical courses with the approval of DRC not earlier than 40 weeks from the date of approval of the dissertation work. For the approval of DRC the candidate shall submit the draft copy of thesis to the Head of the Department and make an oral presentation before the DRC.
- The Dissertation Work Review II in II Year III Sem. carries internal marks of 100. Evaluation should be done by the DRC for 50 marks and the Supervisor will evaluate the work for the other 50 marks. The Supervisor and DRC will examine the Problem Definition, Objectives, Scope of Work, Literature Survey in the same domain and progress of the Dissertation Work. A candidate has to secure a minimum of 70% of marks to be declared successful in Dissertation Work Review II. If he fails to obtain the minimum required marks, he has to reappear for Dissertation Work Review-II as and when conducted.
- One paper in third semester has to be published in any one journal of UGC care, SCOPUS or SCI.

- After successful completion of Dissertation Work Review II, it will be further adjudicated by an external examiner selected by the University. For this, the Principal of the College/School/Institute shall submit name of examiners from among the list of experts in the relevant specialization as submitted by the supervisor concerned and Head of the Department. It will carries external marks of 200.

Phase-2

- The Dissertation Work Review III in II Year IV Sem. carries 250 internal marks. Evaluation should be done by the DRC for 125 marks and the Supervisor will evaluate it for the other 125 marks. The DRC will examine the overall progress of the Dissertation Work and decide whether or not the Dissertation is eligible for final submission. A candidate has to secure a minimum of 70% of marks to be declared successful in Dissertation Work Review III. If he fails to obtain the required minimum marks, he has to reappear for Dissertation Work Review III as and when conducted. For Dissertation Evaluation (Viva Voce) in II Year II Sem. there are external marks of 250 and it is evaluated by the external examiner. The candidate has to secure a minimum of 50% marks in Dissertation Evaluation (VivaVoce) examination.
- One paper in fourth semester has to be published in any one journal of UGC care, SCOPUS or SCI.
- Dissertation Work Reviews II and III shall be conducted in phase I (Regular) and Phase II (Supplementary). Phase II will be conducted only for unsuccessful students in Phase I. The unsuccessful students in Dissertation Work Review II (Phase II) shall reappear for it at the time of Dissertation Work Review III (Phase I). These students shall reappear for Dissertation Work Review III in the next academic year at the time of Dissertation Work Review II only after completion of Dissertation Work Review II, and then Dissertation Work Review III follows. The unsuccessful students in Dissertation Work Review III (Phase II) shall reappear for Dissertation Work Review III in the next academic year only at the time of Dissertation Work Review II (Phase I).
- After approval from the DRC, a soft copy of the thesis should be submitted for ANTIPLAGIARISM check and the plagiarism report should be submitted to the University and be included in the final thesis. The Thesis will be accepted for submission, if the similarity index is less than 30%. If the similarity index has more than the required percentage, the student is advised to modify accordingly and re-submit the soft copy of the thesis after one month. The maximum number of re-submissions of thesis after plagiarism check is limited to TWO. The candidate has to register for the Dissertation work and work for two semesters. After three attempts, the admission is liable to be cancelled. The college authorities are advised to make plagiarism check of every soft copy of theses before submissions.
- Three copies of the Dissertation thesis certified by the supervisor shall be submitted to the College/School/Institute, after submission of a 2 research paper related to the

dissertation work in a UGC care, SCOPUS or SCI journal. A copy of the submitted research paper shall be attached to thesis.

- The thesis shall be adjudicated by an external examiner selected by the University. For this, the Principal of the College/School/Institute shall submit a panel of three examiners from among the list of experts in the relevant specialization as submitted by the supervisor concerned and Head of the Department.
- If the report of the external examiner is unsatisfactory, the candidate shall revise and resubmit the Thesis. If the report of the examiner is unsatisfactory again, the thesis shall be summarily rejected. Subsequent actions for such dissertations may be considered, only on the specific recommendations of the external examiner and /or Dissertation work Review Committee. No further correspondence in this matter will be entertained, if there is no specific recommendation for resubmission.
- If the report of the examiner is satisfactory, the Head of the Department shall coordinate and make arrangements for the conduct of Dissertation Viva- Voce examination. The Dissertation VivaVoce examination shall be conducted by a board consisting of the Supervisor, Head of the Department and the external examiner who adjudicated the Thesis, with an external marks of 250. The candidate has to secure a minimum of 50% of marks in Dissertation Evaluation (Viva-Voce) examination.
- If he fails to fulfill the requirements as specified in previous point he will reappear for the Viva-Voce examination only after three months. In the reappeared examination also, if he fails to fulfill the requirements, he will not be eligible for the award of the degree, unless he is asked to revise and resubmit his dissertation work by the board within a specified time period (within four years from the date of commencement of his first year first semester).
- The Dissertation Viva-Voce External examination marks must be submitted to the University on the day of the examination.