राजीव गांधी राष्ट्रीय युवा विकास संस्थान

Rajiv Gandhi National Institute of Youth Development (RGNIYD)

(राष्ट्रीय महत्व का संस्थान / An Institute of National Importance)

युवा कार्यक्रम विभाग / Department of Youth Affairs

युवा कार्यक्रम और खेल मंत्रालय, भारत सरकार

Ministry of Youth Affairs & Sports, Government of India

श्रीपेरुम्बुदूर Sriperumbudur – 602105 / तमिल नाडु /Tamil Nadu

---

Regulations, Curriculum, and Syllabus

For

Master of Science in Computer Science – Data Science

Under Choice Based Credit System (CBCS)

(Regulations effective from the academic year 2024-25)

Department of Computer Science (Data Science)

REGULATIONS AND SYLLABUS FOR MASTER OF SCIENCE

COMPUTER SCIENCE (DATA SCIENCE)

(For CBCS System in RGNIYD)

(Effective from the Academic Year 2024-2025)

Eligibility for Admission

Any Computer Science Graduate / Any Information Technology Graduate/ Any UG degree with atleast 4 mathematics Papers or any other equivalent UG Computer Science Degree.

Duration of the Course

The course duration shall be for two years spread over four semesters. The maximum duration to complete the course shall be 3 years.

Medium

The medium of instruction shall be English.

Passing & Classification

Passing Eligibility & Classification for the award of the Degree are as per the RGNIYD Norms

REGULATIONS AND SYLLABUS FOR MASTER OF SCIENCE

COMPUTER SCIENCE (DATA SCIENCE)

(For CBCS System in RGNIYD)

CURRICULUM

(Effective from the Academic Year 2024 - 2025)

COURSE STRUCTURE

Course Category (Minimum credits)	**Number of Credits**
DATA SCIENCE - Core	**25**
DATA SCIENCE - Specialization	**16**
Department Electives	**06**
Hands-On Mini Projects	**12**
Project Work and Viva-Voce	**12**
Soft core course (English for Communication / Buddhist Philosophy / Ethics/IPR/Open Electives)	**02**
Course from Online Platform (SWAYAM, MOOC, NPTEL) (Courses are suggested by the Department)	**05**
Total	**78**

Important Note: If the candidates are undergoing a dedicated project during their fourth semester course of study in any of the Industries / MNCs / Organizations, then they should take an additional course from SWAYAM / MOOC / NPTEL etc., instead of their regular course offered by RGNIYD during the fourth semester

M.Sc. – COMPUTER SCIENCE (DATA SCIENCE)

Vision

Department of Computer science aims to become the an internationally recognized center of excellence in computer science with special focus on emerging areas like Artificial Intelligence, Machine Learning, Internet of Things, Data science and Computer Vision.

Mission

The mission of the department is

- To mentor students by providing them an environment that is supportive in fostering intellectual skills
- To produce employable graduates who are trained in cutting edge technologies of Computer Science
- To train the graduates with the skillsets required for collaborative multidisciplinary research
- To develop core competencies in the field of Artificial Intelligence, Machine Learning, Internet of Things, Data science and computer vision

Course Description

M.Sc., in Data Science is a two-years masters' programme which imparts the technical skills and knowledge in the field of data science and its application to the graduates. This programme will equip the students with skills in data analytics, data mining, and business intelligence. The graduands will experience an industry-like environment wherein the learnt techniques will be employed to design and develop innovative & productive solutions. The course has a range of electives that kindle the thought process of the students and helps in selecting a focus area in the applications of Data Science.

Programme Objective

1. To impart technical knowledge, and skills in the field of computer science with specific focus on Data analytics
2. Transform the students in to technically competent industry ready graduands who can adapt to the pressing demands of the industry.
3. Encourage the students to pursue higher studies in Data Analytics

Course Learning Objective

1. To impart fundamental knowledge, thinking skills and technical skills for superior mastery in the core areas of data science and industry 4.0
2. Promote the students in placement in leading business organizations globally
3. Demonstrate global standards to the students in the field of Data Science by providing hands-on training in the hot topics such as decision-making, statistical learning and so on.
4. Enable the students to go to higher studies in Data science or allied subjects like Business Analytics, Big Data and so on.

Course Duration

1. The course duration is of 24 months (2 years) spread over four semesters with credit hours as per the RGNIYD norms
2. The course has sufficient emphasis on mathematical skills as well along with its science and management parts

Course Curriculum Plan

The Course Curriculum is based on comparative analysis of existing MSc. Data Science and Information System curriculums of other Universities, ISI, IITs and NITs. The curriculum has sufficient exposure to hands-on skills and is much more directed towards higher employability. It is also well suited for upward accommodation of science graduates and Mathematics graduates.

Eligibility

Any Computer Science Graduate / Any Information Technology Graduate/ Any UG degree with atleast 4 mathematics Papers or any other equivalent UG Computer Science Degree.

FIRST SEMESTER

Course Code	Course Title	L	T	P	Credits	Core/Specialization
MSDS 101	Data Structures & Algorithms	3	0	0	3	Core
MSDS 102	Problem Solving using Python	3	0	2	4	Core
MSMA104	Probability & Statistics	3	1	0	4	Core
MSDS 103	Database Management System (DBMS)	3	0	2	4	Core
MSDS 104	Foundations of Data Science	3	0	0	3	Specialization
MSDS 105	Data Structures & Algorithms using C Lab	0	0	4	2	Core
	Total Credits	**12**	**1**	**8**	**20**	

SECOND SEMESTER

Course Code	Course Title	L	T	P	Credits	Core/Specialization
MSDS 201	Web Technology and Internet Applications	3	0	0	3	Core
MSDS 202	Machine Learning	3	0	0	3	Core
	Elective -I	3	0	0	3	Specialization
MSDS 203	Data Visualization Techniques	3	0	0	3	Specialization
MSDS 204	Web Technology and JAVA Lab	0	0	4	2	Core
MSDS 205	Mini Project – Use Case Development (Industry Oriented – Banking/Retail/ other core areas)	0	0	12	6	Hands-On
	Total Credits	**9**	**0**	**4**	**20**	

THIRD SEMESTER

Course Code	Course Title	L	T	P	Credits	Core/Specialization
	Elective -II	3	0	0	3	Core
MSDS 301	Deep Learning	3	0	0	3	Specialization
MSDS 302	Big Data Technology & Analytics	2	1	2	4	Specialization
MSDS 303	Data Ethics and Privacy	3	0	0	3	Specialization
MSDS 304	Mini Project - Big Data Analytics/ Business Analytics/Healthcare Analytics	0	0	12	6	Hands-On
	Total Credits	**12**	**4**	**4**	**19**	

FOURTH SEMESTER

Course Code	Course Title	L	T	P	Credits
MSDS 401	(MOOC/ SWAYAM/NPTEL/Elective as decided by the department)	2	0	0	2
MSDS 402	Project Work and Viva-Voce	0	0	24	12
	Total Credits	**2**	**0**	**24**	**14**

SOFTCORE & MOOC'S

Course Code	Course Title	L	T	P	Credits
	Softcore (English Communication / Buddhist Philosophy / Ethics & IPR/Open Elective)	2	0	0	2
MSDS SC02	MOOC / SWAYAM / NPTEL online courses (Minimum 12 weeks, select any one course out of a pool of subjects which are related to the branch study and the list is provided by the Head of the Department)	3	0	0	3
	Total Credits	**5**	**0**	**0**	**5**

Total Credits: 73+5 = 78 Credits

LIST OF ELECTIVES

Course Code	Course Title	L	T	P	Credits
	Elective-I				
MSCY 301	Data Communication and Networking	3	0	0	3
MSDS EL01	Data Mining and Warehousing	3	0	0	3
MSDS EL02	Natural Language Processing	3	0	0	3
	Elective-II				
MSDS EL03	Cloud Computing	3	0	0	3
MSDS EL04	Business Intelligence & Decision Modelling	3	0	0	3
MSDS EL05	Exploratory Data Analysis	3	0	0	3

MSDS101	DATA STRUCTURES AND ALGORITHMS	3	0	0	3

Course Outcomes:

CO1: Understand the concepts of algorithm analysis and solving recurrence relations
CO2: Learn linear and non-linear data structures and their usage in applications
CO3: Understand and Analyze various searching and sorting algorithms
CO4: Applying greedy and dynamic approaches to solve challenging problems
CO5: Understand computational complexity classes and their importance in designing algorithms

Unit – 1	INTRODUCTION TO ALGORITHMS AND ANALYSIS	9 hrs

Overview of algorithms and data structures: Fundamentals of algorithm analysis – Space and time complexity of an algorithm – Asymptotic Notations – Order of growth – Algorithm Efficiency – Best case, Worst Case, Average Case. **Recurrence Relations**: Solving recurrence relations using substitution method, recurrence tree method and Master Method.

Unit – 2	LINEAR DATA STRUCTURES	9 hrs

Linear Data Structures: Array: Types, Applications. **Stacks:** Types-Applications. **Queues:** Types-Applications. **Linked Lists**: Types-Applications.

Unit – 3	NON-LINEAR DATA STRUCTURES	9 hrs

Non-linear Data Structures: Graphs: Types of Graphs-Breadth First Search Algorithm-Depth First Search Algorithm. **Trees:** Binary Trees – Binary Search Tree– AVL Trees-**Tree Traversal Techniques**: Pre-order, In-order and Post -Order Traversal.

Unit – 4	SEARCHING AND SORTING ALGORITHMS &	9 hrs

Searching: Linear Search – Binary Search-Hashing-Interpolation Search -Jump Search-Ternary Search. **Sorting:** Bubble Sort – Insertion Sort – Heap Sort – Merge Sort – Quick Sort – Selection Sort. **Complexity Analysis:** Searching and Sorting algorithms.

Unit –5	DYNAMIC PROGRAMMING APPROACH & ALGORITHM PARADIGMS	9 hrs

Dynamic Programming Paradigm: Knapsack problem – Matrix Chain Multiplication. **All Pair Shortest Path:** Floyd Warshall Algorithm. **Single Source Shortest Path:** Dijkstra's- Bellman Ford Algorithm.

Backtracking: 8-Queens problem – Graph coloring. **Branch and Bound:** Least Cost 0/1 Knapsack problem- Travelling Salesman Problem. Cook's Theorem (without proof).

Text Books:

1. Thomas H. Cormen, C.E. Leiserson, R L.Rivest and C. Stein, Introduction to Algorithms, Third edition, MIT Press February 2010

Reference Books:

1. Sanjoy Dasgupta, C.Papadimitriou and U.Vazirani , Algorithms , Tata McGraw-Hill, 2008.
2. A. V. Aho, J.E. Hopcroft and J. D. Ullman, The Design and Analysis of Computer Algorithms, Pearson, 1st edition, 2006.
3. Fundamentals of Data Structures in C , Horowitz, Sahni and Anderson-Freed (Silicon Press 2007).

MSDS102	PROBLEM SOLVING USING PYTHON	3	0	2	4

Course Outcomes:

CO1: Understand the fundamental concepts of python and its main components.

CO2: Develop (Read and Write) python programs using variables, assignments, and conditional statements using functions.

CO3: Illustrate and implement different data structures.

CO4: Demonstrate Object-oriented concepts and file handling.

CO5: Analyze and plot data using python visualization libraries.

Unit – 1	Introduction to Python	6 hrs

Introduction to a programming language – Features of python – Variables – Expressions and statements- Order of Operators - Type Conversions - Control Flow statements and Loops – Functions – Test and Debug simple Python Programs – Different datatypes in python(variables constants and strings)-Programs on different operators , Control statements, Loops and Functions.

Unit – 2	Data Structures in Python	6 hrs

List Basics- List Indexing and Slicing – Appending - Sorting and Ranging – Tuples – Creation – Deletion - Converting tuple to list – Assignment – Dictionaries – Adding - Modifying and Retrieving Values - Traversing all keys in the dictionary - Operations and methods – Sets - manipulating and accessing sets – Implementing Data Structures in Python

Unit – 3	Exception handling and File I/O	6 hrs

Exception handling - Catching and Handling Exceptions - Object Orientated Concepts (Basics) – Creating python class and Objects - Object properties and methods- Inheritance- Operator overloading- Polymorphism- File handling - Opening, Reading, Writing and Deleting files – Implementing Objects and Classes manipulation using python. Open, Read and write data from/to files in Python

Unit – 4	Graph Plotting	6 hrs

Introduction to plotting python Libraries-Plots and Graphs-Applied Visualizations –Seaborn – Different plots using Matplotlib- Rolling Dice with Pygal- Making a Histogram

Unit –5	Advanced Python Features	6 hrs

CSV File Format. Parsing CSV File Headers, Extracting and Reading Data, Plotting Data in a Temperature Chart, the datetime Module, Plotting Date.

Text Books:

1. Python: The Complete Reference , 2018
2. Python in easy steps, McGraw Hill, 2nd Reprint , 2014

References:

1. Python 3 Documentation, https://docs.python.org/
2. https://www.coursera.org/learn/python
3. Eric Matthes: Python Crash Course: A Hands-On ,Project-Based Introduction to Programming. No Starch Press

MSMA104	PROBABILITY & STATISTICS	3	1	0	4

Pre-requisite: Basic knowledge in definitions and preliminaries of Mathematical Statistics

Course Objectives:

CO1: To provide a thorough treatment of probability ideas and techniques necessary for a firm understanding of the subject.

CO2: Understanding of the ideas in their proofs, and ability to make direct application of those results to related problems.

CO3: As evidence of that understanding, students should be able to demonstrate mastery of all relevant vocabulary, familiarity with common examples and counter examples, knowledge of the content of the major theorems

Unit-1	RANDOM VARIABLES	12 hrs

Random variables - Conditional probability- Probability density function - Distribution function- Marginal and conditional distributions. Two Dimensional Random Variables: Joint distributions – Marginal and conditional distributions – Transformations of random variables of thecontinuous type.

Unit-2	REGRESSION AND CORRELATION	12 hrs

Mathematical expectation I & II: Expectations of functions of random variables. Moment generating function. Regression curve & lines – Correlation.

Unit-3	SOME SPECIAL DISTRIBUTIONS	12 hrs

Some special distributions: The Normal distribution - The Gamma and Chi-Square Distributions and related problem. Chebyshev's inequality and related problem. Central-limit theorem (statement only).

Unit-4	SAMPLING THEORY	12 hrs

Sampling theory- Introduction to statistical inference: Point Estimation –Unbiased estimator- Maximum Likelihood estimate- Confidence intervals for means & variances of Normal population - Curve fitting by principle of least squares: Straight-line & Parabola – Regression lines- Goodness fit.

Unit-5	TESTING OF HYPOTHESIS	12 hrs

Testing Of Hypothesis: Type I and Type II errors- Small and large samples – Tests based on Normal, t and Chi square distributions for testing of mean, variance

Text Book: 1. Fundamental of Mathematical Statistics, Gupta & Kapoor. Sultan Chand & Sons 2002 2. Mathematical Probability, Banerjee, De and Sen. U. N Dhur & Sons Private LTD 3. Mathematical Statistics, De and Sen. U. N Dhur & Sons Private LTD

Reference Books: 1. S. Ross, A first Course in Probability, 6th Edition, Pearson Education, 2006 2. Oliver C Ibe, Fundamental of applied probability & Random Process. 3. "Probability theory and Mathematical statistics" by Marek Fisz, John Wiley andSons, Third Edition, New York, 1963.

MSDS103	DATABASE MANAGEMENT SYSTEMS	3	0	2	4

Course Outcomes:

CO1: Identify the basic concepts and apply relational database theory and recognize and identify the use of normalization and functional dependency, indexing technique used in database design

CO2: Describe relational algebra expression, tuple and domain relation expression for queries and query processing concepts.

CO3: Apply and relate the concept of transaction, concurrency control and recovery in database.

CO4: Familiar with basic database storage structures and access techniques: file and page organizations, indexing methods including B tree, and hashing.

CO5: Understand various concepts like Parallel, distributed and object-oriented databases.

Unit – 1	FUNDAMENTALS OF DBMS	9 hrs

File System versus DBMS – Advantages -Views – Data models – Database languages – Architecture – ER Model – Entities, Attributes and Entity Sets, Relationships and Relationship Sets – Features of ER Model – Conceptual design with ER – Extended E-R. Relational Model - Codd's rule – Integrity Constraints over Relations – Relational database design – Anomalies - Functional dependencies – Normalization – Normal Forms – Decomposition – Denormalization.

Unit – 2	RELATIONAL ALGEBRA AND RELATIONAL CALCULUS	9 hrs

Relational Query Languages – Relational Algebra – Tuple and Domain Relational Calculus – SQL – Query processing and optimization – Transformation of relational expressions – Evaluation Plans

Unit – 3	TRANSACTION CONCURRENCY CONTROL AND SECURITY	9 hrs

Transaction – Properties – Concurrent execution – Serializability – Concurrency control – Protocols – Recovery System – Database Security

Unit – 4	FILE ORGANIZATION	9 hrs

File organization – Organization of records in files – Indexing – B tree and B+ tree index files – Static hashing – Dynamic hashing

Unit –5	ADVANCED CONCEPTS	9 hrs

Parallel and distributed databases – Object-based databases - Mobile databases - XML and Web databases – Intelligent databases – Mongo DB – NOSQL – PostgreSQL

Text Books:-

1. A. Silberchatz, F. Korth, and S. Sudarshan, "Database System Concepts", 7th Edition, McGraw Hill, 2021.
2. R. Elmasri and S. B. Navathe, "Fundamentals of Database Systems", 7th Edition, Pearson 2016

References:-

1. Coronel, Carlos, and Steven Morris. Database systems: design, implementation, & management. Cengage Learning, 2016.
2. Ramakrishnan, R., Gehrke, J., & Gehrke, J. Database management systems , 3rd Edition. McGraw Hill 2006.

MSDS104	FOUNDATION OF DATA SCIENCE	3	0	0	3

Course Outcomes:

CO1: To understand the data science fundamentals and process.
CO2: learn to describe the data for the data science process.
CO3: To learn to describe the relationship between data.
CO4: To utilize the Python libraries for Data Wrangling.
CO5: To present and interpret data using visualization libraries in Python

Unit – 1	INTRODUCTION	9 hrs

Data Science: Benefits and uses – facets of data – Data Science Process: Overview – Defining research goals – Retrieving data – Data preparation – Exploratory Data analysis – Data Mining – Data Warehousing – Basic Statistical descriptions of Data

Unit – 2	DESCRIBING DATA	9 hrs

Types of Data – Types of Variables -Describing Data with Tables and Graphs –Describing Data with Averages – Describing Variability – Normal Distributions and Standard (z) Scores

Unit – 3	DESCRIBING RELATIONSHIPS	9 hrs

Correlation –Scatter plots –correlation coefficient for quantitative data –computational formula for correlation coefficient – Regression –regression line –least squares regression line – Standard error of estimate – interpretation of r2 –multiple regression equations –regression towards the mean

Unit – 4	PYTHON LIBRARIES FOR DATA WRANGLING	9 hrs

Basics of Numpy arrays –aggregations –computations on arrays –comparisons, masks, boolean logic – fancy indexing – structured arrays – Data manipulation with Pandas – data indexing and selection – operating on data – missing data – Hierarchical indexing – combining datasets – aggregation and grouping – pivot tables

Unit –5	DATA VISUALIZATION	9 hrs

Introduction to Data Science Tools – SAS-APACHE FLINK-BigML-Excel-Tableau-Matplotlib-TensorFlow-Weka-Applications: Hands-on with Solving Data Problems-Introduction-Collecting and Analyzing Twitter Data-Collecting and Analyzing YouTube Data.

Text Books:-

1. David Cielen, Arno D. B. Meysman, and Mohamed Ali, “Introducing Data Science”, Manning Publications, 2016. (Unit I)
2. Robert S. Witte and John S. Witte, “Statistics”, Eleventh Edition, Wiley Publications, 2017. (Units II and III)
3. Jake VanderPlas, “Python Data Science Handbook”, O’Reilly, 2016. (Units IV and V)

References:-

1. Allen B. Downey, “Think Stats: Exploratory Data Analysis in Python”, Green Tea Press,2014.
2. Cielen, Davy, Arno DB Meysman, Mohamed Ali, Introducing DataScience:Big Data,Machine Learning and more using Python Tools,Manning Publications Co., 2016

MSDS105	DATA STRUCTURES AND ALGORITHMS USING C LAB	0	0	4	2

Course Outcomes:

CO1: Understand the fundamental concepts of python and its main components.

CO2: Develop (Read and Write) python programs using variables, assignments, and conditional statements using functions.

CO3: Illustrate and implement different data structures.

CO4: Demonstrate Object-oriented concepts and file handling.

CO5: Analyze and plot data using python visualization libraries.

Programs should include but not limited to:

Implementation of

i. Stacks
ii. Queues
iii. Linked Lists
iv. Linear Search and Binary Search
v. Sorting Algorithms
vi. Graph Traversal Algorithms
vii. Tree Traversal Algorithms
viii. Shortest Path Algorithms
ix. Travelling Salesman Problem
x. N-Queens's Problem

MSDS 201	WEB TECHNOLOGY AND INTERNET APPLICATIONS	3	0	0	3

Course Outcomes:

CO1: Understand the principles of www and concepts of web clients and web servers
CO2: Interpret the key responsibilities and functionalities of different internet technologies
CO3: Demonstrate Markup languages and illustrate the working of it
CO4: Analyze the different client side scripting/programming languages
CO5: Explain and demonstrate server side , presentation and database technologies

Unit – 1	WEB ESSENTIALS	9 hrs

Internet Principles – basic web concepts – Client/ server model – Retrieving data from Internet –Internet Protocols and applications

Unit – 2	INTERNET TECHNOLOGIES	9 hrs

Streaming – Networking Principles – Sockets for Clients – Sockets for Servers – Protocol Handlers – Content Handlers – Multicast sockets – Remote method Invocation.

Unit – 3	MARKUP LANGUAGES AND STYLE SHEETS	9 hrs

HTML Elements – HTML Lists – HTML Tables – HTML Forms – Links and addressing – HTML Frames and Images – DHTML – Document Object model (DOM) – XML Schemas – CSS – Text Properties – Advanced CSS

Unit – 4	CLIENT SIDE TECHNOLOGY	9 hrs

Scripts – Javascript – VB Script – JQuery and Ajax

Unit – 5	SERVER SIDE TECHNOLOGY AND DATABASE	9 hrs

Server Scripts – Servlets – Sessions – Cookies – PHP– Node.js – Java Server Pages (JSP) – Active Server pages (ASP) – Simple Applications – On-line databases – Monitoring user events – Plug-ins – Database connectivity

Text Books:

1. Web Technologies: A Computer Science Perspective , Jeffrey Jackson.
2. Harvey M. Deitel and Paul J. Deitel, “Internet & World Wide Web How to Program”, 4th edition, 2008.

Reference:-

3. https://nptel.ac.in/courses/106/105/106105084/

MSDS 202	MACHINE LEARNING	3	0	0	3

Course Outcomes:

CO1: Understand the fundamentals of machine learning, model preparation and evaluation
CO2: Learn ability theory and its application in data understanding
CO3: Apply supervised learning algorithms and unsupervised learning algorithms on real data
CO4: Apply ensemble classifiers on data and compare the model performance
CO5: Explore data repositories and apply Machine Learning Algorithms on case studies

Unit – 1	INTRODUCTION – MODEL PREPARATION – EVALUATION	10 hrs

Introduction to Machine Learning: Human Learning and its Types; Machine Learning and its types; Well-Posed Learning Problem; Applications of Machine Learning; Issues in Machine Learning. Preparing To Model: Basic Data Types; Exploring Numerical Data; Exploring Categorical Data; Exploring Relationship Between Variables; Data Issues and Remediation; Data Pre-Processing. Modelling And Evaluation: Selecting A Model; Training Model – Holdout, K-Fold Cross-Validation, Bootstrap Sampling; Model Representation and Interpretability – Under-Fitting, Over-Fitting, Bias-Variance Tradeoff – Model Performance Evaluation – Classification, Regression, Clustering – Performance Improvement.

Unit – 2	FEATURE ENGINEERING – REVIEW OF PROBABILITY	6 hrs

Feature Engineering: Feature Construction – Feature Extraction – Feature Selection. Brief Review of Probability: Basic Concept Of Probability – Random Variables – Discrete Distributions – Binomial, Poisson – Bernoulli - Continuous Distribution – Uniform, Normal, Laplace; Central Theorem; Monte Carlo Approximation.

Unit – 3	CONCEPT LEARNING – SUPERVISED AND UNSUPERVISED LEARNING	12 hrs

Bayesian Concept Learning: Bayes Theorem – Prior and Posterior Probability, Likelihood; Concept Learning; Bayesian Belief Network. Supervised Learning – Regression: Simple Linear Regression; Other Regression Techniques. Supervised Learning – Classification: Basics of Supervised Learning – Classification; Logistic Regression – k-Nearestneighbour – Decision Tree – Support Vector Machine. Unsupervised Learning: Basics of Unsupervised Learning – Clustering Techniques; Association Rules.

Unit – 4	ENSEMBLE LEARNING	10 hrs

Concept of Ensemble Learning – Bagging and Boosting and its impact on bias and variance – Random Forest – Adaboost Classifier – Gradient Boosting Machines – XG Boost

Unit – 5	PUBLIC DATASETS AND CASE STUDIES	7 hrs

Exploring UCI Machine Learning Repository – Kaggle Data Sets – Analysing Data Sets – Case Studies

Text Books:

1. Saikat Dutt, Subramanian Chandramouli, Amit Kumar Das, “Machine Learning”, Pearson Publications, 2019

Reference Books:-

1. Tom Mitchell, ―Machine Learning‖, McGraw Hill, 3rd Edition,1997
2. Christopher M. Bishop, ―Pattern Recognition and Machine Learning‖, Springer 2011 Edition.
3. Ethem Alpaydin,"Introduction to Machine Learning‖, MIT Press, Prentice Hall of India, Third Edition 2014.

MSDS 203	DATA VISUALIZATION TECHNIQUES	3	0	0	3

Course Outcome:

CO1: To develop skills to both design and critique visualizations.

CO2: To introduce visual perception and core skills for visual analysis.

CO3: To understand visualization for time-series, ranking, deviation, distribution, correlation, multivariate analysis.

CO4: To understand issues and best practices in information dashboard design.

UNIT I	INTRODUCTION	9 hrs

Core Skills for Visual Analysis: Information visualization – effective data analysis – traits of meaningful data – visual perception –making abstract data visible – building blocks of information visualization – analytical interaction – analytical navigation – optimal quantitative scales – reference lines and regions – trellises and crosstabs – multiple concurrent views – focus and context – details on demand – over-plotting reduction – analytical patterns – pattern examples.

UNIT II	KNOWLEDGE REPRESENTATION AND REASONING	9 hrs

Time-Series, Ranking, and Deviation Analysis: Time-series analysis – time-series patterns – time-series displays – time-series best practices – part-to-whole and ranking patterns – part-to-whole and ranking displays – best practices – deviation analysis – deviation analysis displays – deviation analysis best practices.

UNIT III	UNCERTAINTY	9 hrs

Distribution, Correlation, and Multivariate Analysis: Distribution analysis – describing distributions – distribution patterns – distribution displays – distribution analysis best practices – correlation analysis – describing correlations – correlation patterns – correlation displays – correlation analysis techniques and best practices – multivariate analysis – multivariate patterns – multivariate displays – multivariate analysis techniques and best practices.

UNIT IV	LEARNING	9 hrs

Information Dashboard Design: Information dashboard – Introduction– dashboard design issues and assessment of needs – Considerations for designing dashboard-visual perception – Achieving eloquence.

UNIT V	INTELLIGENCE AND APPLICATIONS	9 hrs

Information Dashboard Design: Advantages of Graphics _Library of Graphs – Designing Bullet Graphs – Designing Sparklines – Dashboard Display Media –Critical Design Practices – Putting it all togetherUnveiling the dashboard.

Text Book:

1. Ben Fry, "Visualizing data: Exploring and explaining data with the processingenvironment", O'Reilly, 2008.

References:

1. Edward R. Tufte, "The visual display of quantitative information", Second Edition,Graphics Press, 2001.
2. Evan Stubbs, "The value of business analytics: Identifying the path to profitability",Wiley, 2011.
3. Gert H. N. Laursen and Jesper Thorlund, "Business Analytics for Managers: Takingbusiness intelligence beyond reporting", Wiley, 2010.

MSDS 204	WEB TECHNOLOGY LAB	0	0	4	2

Course Outcomes:

CO1: Understand the principles of www and concepts of web clients and web servers
CO2: Interpret the key responsibilities and functionalities of different internet technologies
CO3: Demonstrate Markup languages and illustrate the working of it
CO4: Analyze the different client side scripting/programming languages
CO5: Explain and demonstrate server side , presentation and database technologies

Programs should include but not be limited to:
Implementations of

i. Demonstration of client server programming
ii. Static and Dynamic web pages creation using HTML
iii. Inline, Internal and external style sheet using CSS
iv. Creation and manipulation of XML schema
v. Scripts – Simple programs
vi. Front end development using JavaScript, VBScript
vii. Programs based on sessions and cookies
viii. Web applications (Backend) using PHP, Servlets
ix. Exercises based on database connectivity
x. Web application for event tracking and monitoring

MSDS 205	Mini Project – Use Case Development (Industry oriented – Banking/Retail/other core areas)	1	2	6	6

General Outcome: On completion of the study of this course the students will be able:
1 To create an Industrial environment and culture within the institution.
2 To set up production lab utilizing the infrastructure of the institution.
3 To standardize laboratories to industrial standard, thereby giving exposure to industrial housekeeping standards.
4 To provide students hands on experience on, troubleshooting, maintenance, fabrication, innovation, record keeping, documentation etc thereby enhancing the skill and competency part of technical education.
5 To promote the concept of entrepreneurship.
6 To inculcate innovative thinking and thereby preparing students for main project.
7 To set up self -maintenance cell within departments to ensure optimal usage of infrastructure facilities.

Guidelines: Mini Project shall be assessed based on following criteria;
1. Quality of survey/ need identification
2. Clarity of Problem definition based on need.
3. Innovativeness in solutions
4. Feasibility of proposed problem solutions and selection of best solution
5. Cost effectiveness
6. Societal impact
7. Innovativeness
8. Cost effectiveness and Societal impact
9. Full functioning of working model as per stated requirements
10. Effective use of skill sets
11. Effective use of standard engineering norms
12. Contribution of an individual's as member or leader
13. Clarity in written and oral communication

Evaluation: Standardization (30%) –(Group (15%) , Individual (15%), Problem identification and solving (50%) or collaborative work (50%) or involvement in production centre (50%) & Documentation (20%).

Innovative ideas of commercial values should be encouraged to be continued as main project for the forth coming semester.

MSDS 301	DEEP LEARNING	3	0	0	3

Course Outcomes:

CO1: Recognize characteristics of deep learning models to solve real world problems
CO2: Understand the concepts of convolutional neural networks and its usage in classification problem
CO3: Learn deep learning supporting environments
CO4: Understand the concept of recurrent neural networks and apply them in solving real world problems
CO5: Generate auto encoders and generative models for real time applications

Unit – 1	INTRODUCTION TO DEEP LEARNING & ARCHITECTURES	9 hrs

Machine Learning Vs. Deep Learning – Representation Learning – Width Vs. Depth of Neural Networks - Activation Functions: RELU – LRELU – ERELU –Unsupervised Training of Neural Networks – Regularization– Dropout - Drop connect – Optimization methods for neural networks – Adagrad – Adadelta – Rmsprop – Adam – NAG.

Unit – 2	CONVOLUTIONAL NEURAL NETWORKS & TRANSFER LEARNING	9 hrs

Architectural Overview – Motivation - Layers – Filters – Parameter sharing – Regularization - Popular CNNArchitectures: LeNet – ResNet – Vggnet – AlexNet. Transfer learning Techniques –DenseNet – PixelNet.

Unit – 3	TRAINING NEURAL NETWORKS	9 hrs

Deep Learning Hardware and Software – CPUs – GPUs – TPUs – PyTorch – TensorFlow – Dynamic vs Static computation graphs – Data Preprocessing – Data Augmentation – Batch Normalization – Transfer Learning – Deep Transfer Learning Strategies – Update Rules - Hyperparameter Tuning – Learning Rate Scheduling –Variants of CNN – ResNet – GoogleNet – Xception etc

Unit – 4	RECURRENT NEURAL NETWORKS	9 hrs

Recurrent Neural Networks – Bidirectional RNNs – Encoder-decoder sequence to sequence Architectures – Backpropagation Through Time for training RNN – Long Short Term Memory Networks.

Unit – 5	AUTO ENCODERS AND DEEP GENERATIVE MODELS	9 hrs

Under complete Autoencoders – Regularized Autoencoders – Sparse Autoencoders – Denoising Autoencoders –Representational Power – Layer – Size and Depth of Autoencoders – Stochastic Encoders and Decoders – Contractive Encoders – Deep Belief networks – Boltzmann Machines – Deep Boltzmann Machine – Generative Adversarial Networks

Text Books:

1. Ian Goodfellow, Yoshua Bengio and Aaron Courville, " Deep Learning", MIT Press, 2017.
2. Josh Patterson, Adam Gibson "Deep Learning: A Practitioner's Approach", O'Reilly Media, 2017

Reference Books:

1. Kevin P. Murphy "Machine Learning: A Probabilistic Perspective", The MIT Press, 2012.
2. Ethem Alpaydin,"Introduction to Machine Learning", MIT Press, Prentice Hall of India, Third Edition 2014.
3. Giancarlo Zaccone, Md. Rezaul Karim, Ahmed Menshawy "Deep Learning with TensorFlow: Exploreneural networks with Python", Packt Publisher, 2017.

CSDS 302	BIG DATA TECHNOLOGY & ANALYTICS	3	1	2	4

Course Outcome:

CO1: To understand the competitive advantages of big data analytics
CO2: To understand the big data frameworks
CO3: To learn data analysis methods
CO4: To learn stream computing
CO5: To gain knowledge on Hadoop related tools such as HBase, Cassandra, Pig, and Hive for big data analytics

UNIT I	INTRODUCTION TO BIG DATA	9 Periods

Big Data – Definition, Characteristic Features – Big Data Applications – Big Data vs Traditional Data – Risks of Big Data- Structure of Big Data – Challenges of conventional Systems – Web Data – Evolution of Analytics Scalability – Evolution of Analytic Processes, Tools and methods – Analysis vs Reporting – Modern Data Analytics Tools.

UNIT II	HADOOP FRAMEWORK	9 Periods

Distributed File Systems – Large-Scale File System Organization – HDFS concepts- MapReduce Execution, Algorithms using MapReduce, Matrix-Vector Multiplication – Hadoop YARN

UNIT III	DATA ANALYSIS	9 Periods

Statistical Methods: Regression modeling, Multivariate Analysis – Classification: SVM & Kernal Methods – Rule Mining – Cluster Analysis, Types of Data in Cluster Analysis, Partitioning Methods, Hierarchical Methods, Density Based Methods, Grid Based Methods, Model Based Clustering Methods, Clustering High Dimensional Data – Predictive Analytics – Data analysis using R

UNIT IV	MINING DATA STREAMS	9 Periods

Streams: Concepts – Stream Data Model and Architecture – Sampling data in a stream – Mining Data Streams and Mining Time – series data- Real Time Analytics Platform (RTAP) Applications – Case Studies – Real Time Sentiment Analysis, Stock Market Predictions

UNIT V	BIG DATA FRAMEWORKS	9 Periods

Introduction to NoSQL – Aggregate Data Models – Hbase: Data Model and Implementations – Hbase Clients – Examples – Cassandra: Data Model – Examples – Cassandra Clients – Hadoop Integration. Pig-Grunt – Pig Data Model – Pig Latin – Developing and testing Pig Latin scripts. Hive- Data Types and File Formats – HiveQL Data Definition – HiveQL Data Manipulation – HiveQL Queries.

Text Books:-

1. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, Second Edition, 2007.
2. Michael Minelli, Michelle Chambers and Ambiga Dhiraj, "Big Data, Big Analytics:Emerging

References:

1. Business Intelligence and Analytics Trends for Today's Businesses", Wiley, 2013.
2. P.J Sadalage and M.Fowler, "NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence", Addison – Wesley Professional, 2012.
3. Richard Cotton, "Learning R – A Step-by-Step Function Guide to Data Analysis", O'Reilly Media, 2013.

MSDS 303	DATA ETHICS & PRIVACY	3	0	0	3

Course Objectives:

The main objectives of this course are to:

1. To understand Data source evolution, data exploration, data format and structure.
2. To understand the importance of Data privacy, ethics and access
3. To analyse data for bias and credibility

Unit:1	Introduction to Data Exploration	9 hrs

Data foundation with data types and structures – Data Collection - Data formats – Types of data – Structured vs Unstructured – The structure of data – Data pre-processing - Check data on bias – credibility – privacy – ethics – pipeline of data access – extract – filter – sort data – Data organizing – protecting data - Data modelling techniques - Data Transformation - Data for exploration — Data analysis - Data design - Data governance

Unit:2	Data Risks and Privacy	9 hrs

Data access and analysis – Risk mitigation – Risks, Harms and Benefit assessment – Sensitive data – Sensitive contexts – Data security – Data Retention – Data Minimization – Data Quality – Open data transparency – Data accountability – Introduction to Data privacy– History of privacy – Degrees of privacy – Modern privacy risks – Anonymity – Data validity – Choice of Attributes and Measures – Errors in Data Processing – Errors in Model Design – Algorithmic Fairness

Unit:3	Data Ethics with Unbiased and Objective Data	9 hrs

Data Ethics – the importance of data ethics — Data anonymization – The ethical use of data - Data Science needs ethics – Data ownership – Data Integrity – Biased and Unbiased data – Fairness – Accountability – Transparency - Data credibility – Data ethics and privacy – Data anonymization – The ethical use of data – Ownership - key issues in Data ethics - Open data usage – Features and characteristics - Legal compatibility of fairness

Unit:4	Database Security and Analytics	9 hrs

Relational databases - Database features – Metadata – Importance – Descriptive and structural metadata – Schemas - Metadata management – Internal and External sources – combine data – Data Integration - Access of different data sources – sorting – filtering – Large datasets – Big Query – organize and secure data

Unit:5	Ethics and Data Protection	9 hrs

Personal Data definition – Transparency – Anonymization – Physical and IT security – Procedures – Passing data to third party – Receiving data – organizing and protecting data – balancing security and analytics – Data protection – Privacy laws – Design privacy – Principles – Compliance with laws and standards – Data sharing

Text Book(s):

1. G.E. Kennedy, Data Privacy Law A Practical Guide to the GDPR, 2019
2. Mike Loukides, Hilary Mason, DJ Patil, “Ethics and Data Science”, O’REILLY Media, Inc., 2018.
3. Journal Joseph, Data & Analytics 4.0, The future of work, Privacy and Trust in the Age of Artificial Intelligence, 2019.

Reference Books:

1. Data Privacy, Ethics and Protection Guidance note on Big Data For Achievement of the 2030 Agenda, United Nations Development Group

MSDS 301	Mini Project – Big Data Analytics/ Business Analytics/Healthcare Analytics	1	2	6	6

General Outcome: On completion of the study of this course the students will be able:
1 To create an Industrial environment and culture within the institution.
2 To set up production lab utilizing the infrastructure of the institution.
3 To standardize laboratories to industrial standard, thereby giving exposure to industrial housekeeping standards.
4 To provide students hands on experience on, troubleshooting, maintenance, fabrication, innovation, record keeping, documentation etc thereby enhancing the skill and competency part of technical education.
5 To promote the concept of entrepreneurship.
6 To inculcate innovative thinking and thereby preparing students for main project.
7 To set up self -maintenance cell within departments to ensure optimal usage of infrastructure facilities.

Guidelines: Mini Project shall be assessed based on following criteria;
1. Quality of survey/ need identification
2. Clarity of Problem definition based on need.
3. Innovativeness in solutions
4. Feasibility of proposed problem solutions and selection of best solution
5. Cost effectiveness
6. Societal impact
7. Innovativeness
8. Cost effectiveness and Societal impact
9. Full functioning of working model as per stated requirements
10. Effective use of skill sets
11. Effective use of standard engineering norms
12. Contribution of an individual's as member or leader
13. Clarity in written and oral communication

Evaluation: Standardization (30%) –(Group (15%) , Individual (15%), Problem identification and solving (50%) or collaborative work (50%) or involvement in production centre (50%) & Documentation (20%).

Innovative ideas of commercial values should be encouraged to be continued as main project for the forth coming semester.

MSCY 301	DATA COMMUNICATION AND NETWORKING	3	0	0	3

Course Outcomes:

CO1: Describe the fundamental concepts of Networking and Physical layer
CO2: Understand details and functionality of Data link layer.
CO3: Analyze switching protocols and routing algorithms
CO4: Analyze features, services and operations of various protocols of TCP/IP suite
CO5: Identify various application layer protocols and its functions

Unit – 1	FUNDAMENTAL CONCEPTS AND PHYSICAL LAYER	9 hrs

Introduction: Basic Definitions - Basic Communication Models – Network Types – Protocol Layers and Service Models – OSI Model. **Multiplexing**: FDM–WDM–Synchronous TDM (time slots & frames, interleaving, data rate management). **Spread Spectrum**: FHSS–DSSS. **Switching**: Switching–Circuit Switched Networks–Datagram networks–Concept of virtual circuit networks–Structure of circuit and packet switch. Concepts of DSL and ADSL.

Unit – 2	PHYSICAL LAYER & MEDIA AND DATA LINK LAYER	9 hrs

Physical Layer & Media: Data and Signal - Analog and Digital Data - Digital Signals - Transmission of Digital Signals - Transmission Impairments - Digital and Analog Transmission - Digital to Digital– Digital to Analog–Analog to Digital and Analog to Analog Communication– Guided and Unguided Transmission media.

Data Link Layer: Types of Errors– Error Detection and Correction– Hamming Codes– Cyclic Redundancy Check– Checksum– Framing– Flow and Error Control– Multiple Access– Aloha, CSMA, CSMA/CD, CSMA/CA, FDMA, TDMA, CDMA, Ethernet– Standard Ethernet, Frame format, Fast and Gigabit Ethernet.

Unit – 3	NETWORK LAYER	9 hrs

Logical Addressing - IPv4 addressing– Classfull and Classless addressing– IPv4 Datagram Structure, Fragmentation and Checksum– ARP–RARP–ICMP– IGMP–Unicast routing Protocols– Distance Vector and Link State Routing–IPv6 addressing.

Unit – 4	TRANSPORT LAYER	9 hrs

Overview of Transport Layer Protocol: Services – Port Numbers – Protocols – User Datagram Protocol (UDP) – Transmission Control Protocol (TCP) – Details - Flow Control. **Congestion Control & Quality of Service**: Data traffic–Congestion–Congestion Control (Open Loop, Closed Loop & Congestion control in TCP)–QoS.

Unit – 5	APPLICATION LAYER AND INTERNET APPLICATIONS	9 hrs

Network Architecture – Layers - HTTP – DNS – Remote Logging (Telnet)– E–Mail (SMTP, MIME, POP3, IMAP, Web Mail) – FTP – Telnet – SNMP.

Text Books:

1. Behrouz A. Forouzan, "Computer Networks - A top-down approach", Tata McGraw-Hill,2012.
2. Computer Networks -- Andrew S Tanenbaum, 6th Edition, Pearson Education

References:-

1. James F. Kurose, Keith W. Ross, "Computer Networking, A Top-Down Approach Featuring theInternet", Sixth Edition, Pearson Education, 2012.
2. Larry L. Peterson, Bruce S. Davie, "Computer Networks: A Systems Approach", Fifth
3. Edition, Morgan Kaufmann Publishers Inc., 2012.
4. Ying-Dar Lin, Ren-Hung Hwang and Fred Baker, Computer Networks: An Open Source
5. Approach ", McGraw Hill Publisher, 2011.

MSDS EL01	DATA MINING & WAREHOUSING	3	0	0	3

Course Outcome:
CO1: This course will allow learners to discuss diverse modalities of data by outlining the core difference in the nature of data and to demonstrate computational methods for outlier detection, classification, and clustering that explores interesting patterns

UNIT I	Introduction	9 Periods

Introduction: What is data; How to Mine Data; Principles of Data Mining; Web content mining; Features and Weighted Features; Class Labels – Binary and Multistage; Data Visualization; Data Discretization and Concept Hierarchy Generation; Feature Interaction; Pattern & Analysis; Noise & Outliers

UNIT II	Outlier Analysis	9 Periods

Outlier Analysis: What is an Outlier; Outlier in Text Data; Outlier in Higher Multimedia Context; Principal Component Analysis; Singular Value Decomposition; Z-score method; Extreme Value Method; Clustering method; Bayesian Outlier Detection.

UNIT III	Data Pre-processing	9 Periods

Data Pre-processing: Data Transformation; Missing Value Detection & Imputation – Types of Imputation; Binning Methods; Regression Models; Geo-Imputation; Rare Pattern Approach.

UNIT IV	Classification	9 Periods

Classification: Introduction to classifiers, Naïve Bayes algorithm, Decision tree, Random Forest, KNN algorithm, Support Vector Machine, Kernel Functions, Back propagation neural network, Symbolic dynamic filter, Extreme learning machines, Metrics for classification

UNIT V	Clustering	9 Periods

Clustering: Introduction to clustering, Categorization of Major Clustering Methods, Partion clustering, Hierarchial clustering, Density-based clustering, Grid-based clustering, Metric for clustering

Text Book:-

1. J. Han and M. Kamber, Data Mining – Concepts & Techniques, Morgan Kaufmann Publisher, 2011.

References:-

2. C. W. Hsu, C. C. Chang & C. J. Liu, Practical Guide to SVM , Taiwan Publishers, 2021
3. F. Sun, K. A. Toh, M. G. Romay & K. Mao, Extreme Learning Machines Algorithms & Applications, Springer Publisher, 2013

MSDS EL02	NATURAL LANGUAGE PROCESSING	3	0	0	3

Course Outcomes:

- To learn about the concepts and principles of natural language processing.
- To explore both theoretical and practical issues of natural language processing.
- To develop skills of finding solutions and building software using natural language processingtechniques.

Unit – 1	Introduction to Natural Language Processing	9 hrs

Introduction: Natural Language Processing tasks in syntax, semantics, and pragmatics – Issues - Applications - The role of machine learning - Probability Basics –Information theory – Collocations -N-gram Language Models - Estimating parameters and smoothing - Evaluating language models.

Unit – 2	Morphology and part of speech tagging	9 hrs

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

Unit – 3	Syntax Parsing	9 hrs

Syntax Parsing - Grammar formalisms and treebanks - Parsing with Context Free Grammars - Features and Unification -Statistical parsing and probabilistic CFGs (PCFGs)-Lexicalized PCFGs.

Unit – 4	Semantic analysis	9 hrs

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.

Unit – 5	Applications	9 hrs

Applications- Named entity recognition and relation extraction- IE using sequence labeling- Machine Translation (MT) - Basic issues in MT-Statistical translation-word alignment- phrase- based translation – Question Answering

Text / Reference Books:

1. Daniel Jurafsky and James H. Martin Speech and Language Processing (2nd Edition), Prentice Hall; 2 edition, 2008
2. Foundations of Statistical Natural Language Processing by Christopher D. Manning and Hinrich Schuetze, MIT Press, 1999
3. Steven Bird, Ewan Klein and Edward Loper Natural Language Processing with Python, O'Reilly Media; 1 edition, 2009
4. Roland R. Hausser, Foundations of Computational Linguistics: Human- C o m p u t e r Communication in Natural Language, Paperback, MIT Press, 2011
5. Pierre M. Nugues, An Introduction to Language Processing with Perl and Prolog: An Outline of Theories, Implementation, and Application with Special Consideration of English, French, and German (Cognitive Technologies) Softcover reprint, 2010
6. James Allen, Natural Language Understanding, Addison Wesley; 2 edition 1994

MSDS EL03	Cloud Computing	3	0	0	3

Course Objectives:

1. To provide students with the fundamentals and essentials of Cloud Computing.
2. To provide students a sound foundation of the Cloud Computing so that they are able to start using and adopting Cloud Computing services and tools in their real life scenarios.
3. To enable students exploring some important cloud computing driven commercial systems such as Google Apps, Microsoft Azure and Amazon Web Services and other businesses cloud applications.
4. To impart knowledge in applications of cloud computing

UNIT I	Introduction to Cloud Computing	9 hrs

Cloud Computing Overview: Characteristics – challenges, benefits, limitations, Evolution of Cloud Computing, Cloud computing architecture, Cloud Reference Model (NIST Architecture)

UNIT II	Infrastructure, Platform & Software as a Service	9 hrs

Infrastructure as a Service: Service Model, Characteristics, Benefits, Enabling Technologies Case Study : AWS, OpenStack
Platform as a Service: Service Model, Characteristics, Benefits, Enabling Technologies Case Studies : IBM Bluemix, GAE, Microsoft Azure
Software as a Service: Service Model, Characteristics, Benefits, Enabling Technologies Case Study : Salesforce.com, CRM, Online Collaboration Services

UNIT III	Data Analytics as a Service	9 hrs

Hadoop as a service, MapReduce on Cloud, Chubby locking Service

UNIT IV	Introduction to Public and Private Clouds	9 hrs

Shared Resources – Resource Pool – Usage and Administration Portal – Usage Monitor – Resource Management– Cloud Security – Workload Distribution – Dynamic provisioning.

UNIT V	Storage as a service	9 hrs

Historical Perspective, Datacenter Components, Design Considerations, Power Calculations, Evolution of Data Centers, Cloud data storage - CloudTM

Text Book/References:

1) Kai Hwang, Geoffrey Fox, Jack J. Dongarra, Morgan Kaufmann, "Distributed and Cloud Computing: From Parallel Processing to the Internet of Things," 1st Edition, 2011.
2) Gautham Shroff, "Enterprise Cloud Computing: Technology, Architecture, Applications", Cambridge press, 2010. Kris Jamsa, "Cloud Computing", Jones & Barlett Learning, 2013.
3) Rajkumar Buyya, James Broberg, Andrzej Goscinski, "Cloud Computing Principles and Paradigms", John Wiley & Sons, 2011.
4) John Rhoton and Risto Haukiojal, "Cloud Computing Architectured : Solution Design Handbook", Recursive Press, 2013.
5) George Recse, "Cloud Application Architectures: Building Application and Infrastructure in the Cloud" , O' Reilly Media, First Edition, 2009.
6) Dinkar Sitaram, Geetha Manjunathan, "Moving to the Cloud: Developing Apps in the new world of Cloud Computing", Syngress, 2012.
7) Samee. U. Khan, Albert. Y. Zomaya, "Handbook on Data Centers", Springer, 2015.

MSDS EL04	BUSINESS INTELLIGENCE AND DECISION MODELLING	3	0	0	3

OBJECTIVES

- Be exposed with the basic rudiments of business intelligence system
- Understand the modeling aspects behind Business Intelligence
- Be exposed with different data analysis tools and techniques
- Apply business intelligence in different domain

UNIT-1	DECISION SUPPORT AND BUSINESS INTELLIGENCE	9Hrs

Definition, concept and need for Business Intelligence – Changing Business Environments and Computerized Decision Support, Managerial Decision Making, Computerized Support for Decision Making, An Early framework for Computerized decision support, Concept of Decision Support Systems, A framework for Business Intelligence, Major Tools and Techniques, Decision Making – Introduction and Definitions, Models, Phases of the decision-Making process, Decision Support System – Concepts, Methodologies and Technologies, Classifications, Components of Decision Support Systems

UNIT-2	KNOWLEDGE DELIVERY	9Hrs

The business intelligence user types, Standard reports, Interactive Analysis and Ad Hoc Querying, Automated reports and Self-Service Reporting, dimensional analysis, Alerts/Notifications, Visualization: Charts, Graphs, Widgets, Scorecards and Dashboards, Heatmap, Geographic Visualization, Integrated Analytics, Considerations: Optimizing the Presentation for the Right Message

UNIT-3	DATA STORAGE FOR BUSINESS INTELLIGENCE	9Hrs

Data Mining Concepts and Applications, Data Mining Applications, Data Mining Process, Data Mining Methods, Artificial Neural Networks for Data Mining, Data Warehousing – Definitions and Concepts, Process Overview, Architectures, ETL Process, Data warehouse development, Real-Time Data Warehousing

UNIT-4	BUSINESS INTELLIGENCE TOOLS FOR APPLICATIONS	9Hrs

Text and Web Mining – Text mining applications and tools, Web Mining – Overview, Web content mining and web structure mining, Web usage Mining. Data Visualization – Using R, Case Studies – Boston Housing Data, Iris Dataset

UNIT-5	FUTURE OF BUSINESS INTELLIGENCE	9Hrs

Future of business intelligence – Emerging Technologies, Machine Learning, Predicting the Future, BI Search & Text Analytics – Advanced Visualization – Rich Report, Future beyond Technology

TEXT BOOKS:

1. Efraim Turban, Ramesh Sharda, DursunDelen, “Decision Support and Business Intelligence Systems”, 9th Edition, Pearson 2013.
2. GalitShmueli, Nitin R. Patel and Peter C. Bruce, ―Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner, Wiley, 2007

REFERENCES:

1. Larissa T. Moss, S. Atre, “Business Intelligence Roadmap: The Complete Project Lifecycle of Decision Making”, Addison Wesley, 2003.

2. David Loshin Morgan, Kaufman, “Business Intelligence: The Savvy Manager’s Guide”, Second Edition, 2012.
3. Cindi Howson, “Successful Business Intelligence: Secrets to Making BI a Killer App”, McGraw-Hill, 2007.
4. . Ralph Kimball , Margy Ross , Warren Thornthwaite, Joy Mundy, Bob Becker, “The Data Warehouse Lifecycle Toolkit”, Wiley Publication Inc.,2007.

MSDS EL05	EXPLORATORY DATA ANALYSIS	3	0	0	3

Course Objectives:

1.This course introduces the methods for data preparation and data understanding.
2.It covers essential exploratory techniques for understanding multivariate data by summarizing it through statistical methods and graphical methods.
3.Supports to Summarize the insurers use of predictive analytics, data science and Data Visualization

Unit – 1	Introduction To Exploratory Data Analysis	9 Hrs

Data Analytics lifecycle, Exploratory DataAnalysis(EDA)– Motivation,Stepsindataexploration, The basic data types Data Type Portability

Unit – 2	Preprocessing-Traditional Methods And Maximum Likelihood Estimation	9 Hrs

Introduction to Missing data, Traditional methods for dealing with missing data, Maximum Likelihood Estimation – Basics, Missing data handling, Improving the accuracy of analysis

Unit – 3	Preprocessing Bayesian Estimation	9 Hrs

Multiple Imputation-Imputation Phase, Analysis and Pooling Phase, Practical Issues in Multiple Imputation, Models for Missing Notation Random Data

Unit – 4	Data Summarization , Visualization & Outlier Analysis	9 Hrs

Statistical data elaboration, 1-D Statistical data analysis, 2-D Statistical data Analysis, ND Statistical data analysis, Introduction, Extreme Value Analysis, Clustering based, Distance Based and Density Based outlier analysis, Outlier Detection in Categorical Data,

Unit – 5	Feature Subset Selection & Dimensionality Reduction	9 Hrs

Feature selection algorithms: filter methods, wrapper methods and embedded methods, Forward selection backward elimination, Relief, greedy selection, genetic algorithms for features election, Principal Component Analysis(PCA), Kernel PCA, Canonical Correlation Analysis, Factor Analysis, Multi dimensional scaling, Correspondence Analysis

Text Book(s) & Reference Books:

1. Charu C. Aggarwal ,“Data Mining The Text book”, Springer, 2015.
2. Craig K. Enders, “Applied Missing Data Analysis”, The Guilford Press, 2010.
3. Inge Koch, “Analysis of Multivariate and High dimensional data”, Cambridge University Press, 2014.
4. Michael Jambu, “Exploratory and multivariate data analysis”, Academic Press Inc. , 1990.
5. Charu C. Aggarwal, “Data Classification Algorithms and Applications”, CRC press, 2015

SOFTCORE (ENGLISH)	2	0	0	2
This is a Spoken English Training course offered by the department of English. There are many people who know reading and writing of English but find it difficult to speak while communicating with others. As we know, without speaking practice it's impossible to learn and speak any language.				
English is the most used language in this entire world. Speaking English immediately opens doors to opportunities regardless of your race, color, and caste anywhere in this world. Not only has that, speaking fluent English had its major advantages in terms of employability.				
Softcore has carefully taken into account the general mistakes and errors most commonly used by non-English speaking speakers and also aim to reduce the heavy mother tongue influence of non-native speakers.				