B. K. Birla College of Arts, Science and Commerce (Autonomous), Kalyan

(Affiliated to University of Mumbai)

Syllabus for M.Sc. Program as per

National Education Policy (NEP) 2020 framework:

M.Sc. Data Science and Big Data Analytics

In association with

Program Code: Information Technology

(Credit Based Semester and Grading System
Academic year 2023–2024)

Preamble:

In today's world there is data available in abundance from variety of sources like web server logs, social media, and large databases and from diverse domains like Ecommerce, Medical, Scientific etc. Big data analytics is the process of examining these data to uncover hidden patterns, unknown correlations and other useful information that can be used to make better decisions. Engineers, Business personnel, Doctors, Scientists etc. can use this to improve their services.

This course is designed to give students a comprehensive understanding of data science principles and techniques and the ability to analyze and interpret large volumes of complex data. In this era of digital transformation, organizations across industries are seeking skilled professionals who can leverage the power of data to drive insights, make informed decisions, and gain a competitive edge.

Throughout this program, students will delve into the world of data science and big data analytics, exploring various methodologies, tools, and technologies used to extract valuable knowledge from massive datasets. The curriculum is designed to equip students with both theoretical knowledge and practical skills, enabling them to navigate the entire data science lifecycle – from data collection and preprocessing to analysis, modeling, and visualization.

The program emphasizes a hands-on approach, allowing students to gain practical experience through real-world case studies, industry projects, and internships. They will have the opportunity to work with state-of-the-art tools and technologies, such as Python, R, Hadoop, Spark, and machine learning algorithms, to solve complex data challenges and uncover meaningful insights.

In addition to technical expertise, the program also focuses on developing critical thinking, problem-solving, and communication skills. Students will learn to effectively communicate their findings and insights to diverse stakeholders, making data-driven recommendations that drive business success.

Upon successful completion of this program, students will be equipped with the knowledge and skills to pursue various career paths in data science and big data analytics. They will be prepared to work in industries such as finance, healthcare, marketing, e-commerce, and more, where the ability to harness the power of data is crucial for making strategic decisions and driving innovation.

We are excited to embark on this data science journey with you, and we look forward to equipping you with the skills and knowledge needed to thrive in the dynamic world of data analytics. Get ready to explore the vast landscape of data, uncover hidden insights, and make a meaningful impact with your data-driven expertise.

Eligibility:

To secure entry into the MSc DSBDA program in any college, one needs to meet the following criteria:

- Applicants for the MSc DSBDA program must have earned a bachelor's degree.
- Candidates must have passed their Bachelor's degree in B.Sc. (I.T. / C.S. / Mathematics / Statistics) / B.E. (I.T. /C.S.) / B.Tech. (I.T. /C.S.) / BCA and any other equivalent degree
- MSc DSBDA admission will be based on merit score and Personal Interview.

Duration:

Qualification Title	Credit Requirement		Semester	Year
	Minimum	Maximum		
PG Diploma	40	44	2	1
PG Degree	40	44	4	2

Program Outcome

- **Proficiency in Data Analysis and Interpretation:** Graduates will possess advanced skills in analyzing and interpreting large and complex datasets. They will be proficient in applying statistical methods, data mining techniques, and machine learning algorithms to extract meaningful insights and patterns from data.
- **Expertise in Data Visualization:** Graduates will demonstrate expertise in visualizing data effectively to communicate insights and findings. They will be skilled in using data visualization tools and techniques to present complex information in a clear and visually appealing manner.
- **Proficiency in Machine Learning and Predictive Analytics:** Graduates will have a solid foundation in machine learning algorithms and predictive analytics techniques. They will be able to apply these methods to build models that can make accurate predictions and support data-driven decision-making.
- **Knowledge of Big Data Technologies and Tools:** Graduates will be familiar with the tools and technologies used for handling big data. They will have practical experience working with distributed computing frameworks (such as Hadoop and Spark), NoSQL databases, and cloud-based data platforms.
- **Skills in Data Preprocessing and Feature Engineering:** Graduates will possess skills in data preprocessing and feature engineering. They will be able to clean and transform raw data, handle missing values, and engineer relevant features for machine learning models.
- **Proficiency in Programming and Data Manipulation:** Graduates will be proficient in programming languages commonly used in data science, such as Python or R. They will have expertise in data manipulation, including data cleaning, data integration, and data transformation.
- **Understanding of Data Privacy and Ethical Considerations:** Graduates will have a sound understanding of data privacy regulations and ethical considerations in data science. They will be aware of the importance of handling data responsibly, ensuring data privacy and security, and adhering to ethical guidelines in data analysis.
- **Effective Communication and Collaboration:** Graduates will possess effective communication and collaboration skills. They will effectively communicate their findings and insights to technical and non-technical stakeholders. They will also be adept at working in interdisciplinary teams and collaborating with professionals from different domains.
- **Problem-Solving and Critical Thinking:** Graduates will demonstrate strong problem-

solving and critical thinking skills in the context of data science and big data analytics. They will be able to analyze complex problems, identify appropriate methodologies, and propose effective solutions based on data-driven insights.

- **Lifelong Learning and Adaptability:** Graduates will recognize the importance of lifelong learning and adaptability in a rapidly evolving field. They will be equipped with the skills and mindset to stay updated with emerging trends, technologies, and methodologies in data science and big data analytics.

Syllabus as per NEP 2020

<table>
<tr><th>Sem</th><th>Major
(Credits-14)</th><th>Electives
(Credits- 4)</th><th>Minor
(Credits – 4)</th><th>OJT
(Credits – 4)</th><th>Total</th></tr>
<tr><td rowspan="4">I</td><td>Course- I
Applied Statistics with Excel and R</td><td rowspan="4">Basics of Data Science with Python

OR

Design Thinking</td><td rowspan="4">Research Methodology</td><td rowspan="4">-</td><td rowspan="4">22</td></tr>
<tr><td>Course -II
Data Modelling and Visualization</td></tr>
<tr><td>Course -III
Data on Cloud</td></tr>
<tr><td>Course -III
Entrepreneurial Skills and Scientific Writing</td></tr>
<tr><td rowspan="3">II</td><td>Course -I
Practical Approach to Data Mining & Analytics</td><td rowspan="3">Advance Data Structures and algorithm

OR

Sentiment, Web and Text Analytics</td><td rowspan="3">-</td><td rowspan="3">Internship with Project</td><td>22</td></tr>
<tr><td>Course -II
Machine Learning</td><td></td></tr>
<tr><td>Course -III
Optimization and Simulation for Data Science</td><td></td></tr>
</table>

Course Code	Applied Statistics with Excel and R		L	T	P	C
			2	-	2	4
Pre-requisites		**Semester**	**I**			

Course Objectives:

- Develop Proficiency in Data Analysis
- Understand Data Summarization and Visualization
- Apply Descriptive and Diagnostic Analytics
- Explore Predictive and Prescriptive Analytics
- Enhance Data Analysis Skills
- Gain Practical Experience
- Develop Data-Driven Decision-Making Skills
- Foster Critical Thinking and Problem-Solving

Course Outcomes (CO):

CO 1	Recall and identify key statistical concepts and definitions.
CO 2	Explain the fundamental principles and theories of statistics.
CO 3	Apply statistical techniques to analyze and interpret real-world data.
CO 4	Analyze and critically evaluate the validity and reliability of statistical results.
CO 5	Evaluate the quality and accuracy of statistical reports and research findings.
CO 6	Design and conduct statistical experiments or surveys.

UNIT-1	**Introduction to Statistics and Data Analysis**	5	CO1, CO2

Overview of statistics and its applications in decision-making, Introduction to Excel and R for statistical analysis, Understanding data types and data structures in Excel and R, Data import and export in Excel and R

UNIT-2	**Data Summarization and Data Visualization**	6	CO3

Data cleaning and preprocessing techniques, summarizing data using measures of central tendency and dispersion, Creating charts and graphs for data visualization in Excel and R, Customizing visualizations and enhancing data presentation, Exploratory data analysis (EDA) techniques

UNIT-3	**Descriptive and Diagnostic Analytics**	6	CO3, CO4

Data distribution analysis and probability concepts, Hypothesis testing and confidence intervals, Correlation and regression analysis, Outlier detection and handling, Diagnostic techniques for model evaluation and validation

UNIT-4	**Predictive Analytics**	6	CO3, CO4

Introduction to predictive modeling and machine learning, Classification and regression techniques using Excel and R, Model evaluation and selection, Time series forecasting methods, Cross-validation and model performance assessment

UNIT-5	**Prescriptive Analytics**	7	CO4, CO6

Introduction to optimization and decision-making, Linear programming and optimization models in Excel and R, Solver add-in for optimization problems, Decision analysis and sensitivity analysis, Introduction to decision tree analysis and simulation modeling

Course Code	Data Modeling and Visualization		L	T	P	C
			2	-	2	4
Pre-requisites	Basic knowledge of programming fundamentals	**Semester**	**I**			

Course Objectives:

- Develop Proficiency in Data Modeling
- Understand Data Visualization Principles
- Apply Statistical Techniques for Data Analysis
- Gain Proficiency in Data Visualization Tools
- Communicate Insights Effectively

Course Outcomes (CO):

CO 1	Recall the concepts of data-analytic thinking, including understanding data, data preprocessing, and storytelling with data.
CO 2	Understand the importance of data-analytic thinking and its application in data analysis and visualization.
CO 3	Apply programming skills to create data visualizations.
CO 4	Analyze and interpret data visualizations and the performance of modeling algorithms using appropriate performance measures
CO 5	Evaluate the concepts of dynamic/interactive data visualization and generate reports
CO 6	Apply data modeling techniques to create visualizations from structured data sources like databases (SQL and NoSQL), semi-structured data like CSV files, XML, JSON, and live streaming data.

UNIT-1	**Data-Analytic Thinking**		CO1, CO2
Knowing your data, Data preprocessing, Storytelling with data			
UNIT-2	**Data Visualization using R**		CO3, CO4
Introduction to R programming, Visualization using R, Transformation using R, Exploratory data analysis			
UNIT-3	**Data Modeling**		CO3, CO4, CO5
Linear regression, Logistic regression ,K-means clustering, Performance measure, Implementation of some modeling algorithms using R			
UNIT-4	**Data Visualization using Tableau**		CO3, CO5
Introduction to Tableau, data import and management, data type and operations, Different types of data visualizations, dashboards, storytelling, Understanding the concepts of dynamic/interactive data visualization and report generation			
UNIT-5	**Data Modeling from Different Data Sources for Visualization**		CO2, CO3
Understanding structured, unstructured and semi-structured data sources,Data modeling and creating visualization charts/dashboards from structured data like databases (SQL and NoSQL),Data modeling and creating visualization charts/dashboards from semi-structured data like CSV files, XML, JSON and others ,Data modeling and creating visualization charts/dashboards from live streaming data.			

Course Code	Data on Cloud		L	T	P	C
			2	-	2	4
Pre-requisites		**Semester**	**I**			

Course Objectives:

- Understand Cloud Computing Concepts
- Explore Cloud-based Data Storage Solutions
- Master Data Processing on the Cloud
- Ensure Data Security and Privacy on the Cloud
- Implement Cloud-based Data Integration and Migration

Course Outcomes (CO):

CO 1	Recall the key concepts and principles of cloud computing and its application in data management.
CO 2	Understand the distinct types of cloud-based data storage and processing services and their functionalities. Comprehend the security and privacy challenges associated with data on the cloud.
CO 3	Apply cloud-based data storage solutions to effectively store and retrieve large volumes of data. Utilize cloud-based data processing tools and technologies to analyze and transform data.
CO 4	Analyze the performance and scalability of cloud-based data solutions. Evaluate data security measures and propose strategies to protect data on the cloud.
CO 5	Design data integration and migration strategies for seamless transfer of data to and from cloud platforms.
CO 6	Create innovative data solutions using cloud resources that meet specific business requirements

UNIT-1	**Introduction to Cloud Computing and Data Management**		CO1, CO3

- Overview of cloud computing and its relevance in data science and big data analytics
- Cloud service models (IaaS, PaaS, SaaS) and deployment models (public, private, hybrid)
- Cloud-based data storage options and architectures
- Cloud-based data management tools and technologies
- Data security and privacy considerations in cloud environments

UNIT-2	**Cloud-based Data Storage and Retrieval**		CO2

- Object storage systems and services (e.g., Amazon S3, Google Cloud Storage)
- Cloud-based file systems (e.g., Hadoop Distributed File System, Azure Blob Storage)
- Cloud-based databases (e.g., Amazon RDS, Google Cloud SQL, Microsoft Azure Cosmos DB)
- Data migration and synchronization techniques between on-premises and cloud storage

UNIT-3	**Data Processing and Analytics on the Cloud**		CO3, CO2

- Distributed computing frameworks (e.g., Apache Hadoop, Apache Spark) for big data processing
- Cloud-based data processing services (e.g., Amazon EMR, Google Cloud Dataproc)
- Data preprocessing and feature engineering techniques on the cloud
- Introduction to distributed data processing languages (e.g., Apache Pig, Apache Hive)

UNIT-4	**Virtualization**		CO4

- Virtualization system-specific attacks Guest hopping, attacks on the VM (delete the VM, attack on the control of the VM
- Code or file injection into the virtualized file structure), VM migration attack, hyper jacking.
- Technologies for virtualization-based security enhancement IBM security virtual server protection, virtualization-based sandboxing.
- Storage Security- HIDPS, log management, Data Loss Prevention, Location of the Perimeter.

UNIT-5	**Data Visualization and Reporting in the Cloud**		CO5, CO6

- Cloud-based data visualization tools (e.g., Tableau, Power BI, Google Data Studio)
- Creating interactive dashboards and reports on the cloud
- Integrating data visualizations with cloud-based data sources
- Collaborative data sharing and storytelling using cloud platforms

Books:

1. "Big Data: A Revolution That Will Transform How We Live, Work, and Think" by Viktor Mayer-Schönberger and Kenneth Cukier
2. "Cloud Computing: Concepts, Technology, and Architecture" by Thomas Erl, Ricardo Puttini, and Zaigham Mahmood
3. "Cloud Computing: Principles and Paradigms" by Rajkumar Buyya, James Broberg, and Andrzej Goscinski
4. "Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking" by Foster Provost and Tom Fawcett
5. Data Analysis in the Cloud by Fabrizio Marozzo, Paolo Trunfio, Domenico Talia
6. "Big Data Analytics: Turning Big Data into Big Money" by Frank J. Ohlhorst

Course Code	**Entrepreneurial Skills and Scientific Writing**		**L**	**T**	**P**	**C**
			2	-	2	4
Pre-requisites		**Semester**	**I**			

Course Objectives:

- Understand the fundamentals of entrepreneurship in the context of data science
- Develop effective scientific writing skills
- Cultivate critical thinking and analysis skills
- Foster innovation and creativity
- Develop entrepreneurial mindset and business acumen

Course Outcomes (CO):

CO 1	Recall the fundamental concepts and principles of entrepreneurship and scientific writing in the context of data science and big data analytics.
CO 2	Understand the importance of entrepreneurial skills in the context of data science and big data analytics, including identifying opportunities, risk assessment, and innovation.
CO 3	Apply entrepreneurial skills to identify business opportunities and develop innovative data-driven solutions.
CO 4	Analyze market trends, customer needs, and competition to make informed business decisions in the field of data science and big data analytics.
CO 5	Design and develop business plans and strategies for data-driven ventures,

	considering market dynamics and potential challenges.
CO 6	Create well-structured and concise scientific reports, research papers, and presentations that effectively communicate research methodologies, results, and implications.

UNIT-1	**Introduction**		CO1, CO3

Introduction to Technical Communication:
What Is Technical Communication? The Challenges of Producing Technical Communication, Characteristics of a Technical Document, Measures of Excellence in Technical Documents, Skills and Qualities Shared by
Successful Workplace Communicators, How Communication Skills and Qualities Affect Your Career?
Understanding Ethical and Legal Considerations: A Brief Introduction to Ethics, Your Ethical Obligations, Your Legal Obligations, The Role of Corporate Culture in Ethical and Legal Conduct, Understanding Ethical and
Legal Issues Related to social media, Communicating Ethically Across Cultures, Principles for Ethical Communication
Writing Technical Documents:
Planning, Drafting, Revising, Editing, Proofreading
Writing Collaboratively: Advantages and Disadvantages of Collaboration, Managing Projects, Conducting Meetings, Using social media and Other
Electronic Tools in Collaboration, Importance of WordPress Website, Gender and Collaboration, Culture and Collaboration

UNIT-2	**Content Writing and Organization**		CO2, CO3

Introduction to Content Writing: Types of Content (Article, Blog, E-Books, Press Release, Newsletters etc.), Exploring Content Publication Channels. Distribution of your content across various channels.
Blog Creation:
Understand the psychology behind your web traffic, creating killing landing pages which attract users, Using Landing Page Creators, setting up Accelerated Mobile Pages, Identifying UI UX Experience of your website or
blog.
Organizing Your Information: Understanding Three Principles for Organizing Technical Information, Understanding Conventional Organizational Patterns,
Emphasizing Important Information: Writing Clear, Informative Titles, Writing Clear, Informative Headings, Writing Clear Informative Lists, Writing Clear Informative Paragraphs.

UNIT-3			CO4, CO5

Creating Graphics: The Functions of Graphics, The Characteristics of an Effective Graphic, Understanding the Process of Creating Graphics, Using Color Effectively, Choosing the Appropriate Kind of Graphic, Creating Effective Graphics for Multicultural Readers.
Researching Your Subject: Understanding the Differences Between Academic and Workplace Research, Understanding the Research Process, Conducting Secondary Research, Conducting Primary Research,
Research and Documentation: Literature Reviews, Interviewing for Information, Documenting Sources, Copyright, Paraphrasing, Questionnaires.
Report Components: Abstracts, Introductions, Tables of Contents, Executive Summaries, Feasibility Reports, Investigative Reports, Laboratory Reports, Test Reports, Trip Reports, Trouble Reports

UNIT-4			CO 4

Writing Proposals, The Logistics of Proposals, The ―Deliverables‖ of Proposals, Persuasion

and Proposals, Writing a Proposal, The Structure of the Proposal.
Writing Informational Reports: Understanding the Process of Writing Informational Reports, Writing Directives, Writing Field Reports, Writing Progress and Status Reports, Writing Incident Reports, Writing Meeting Minutes.
Writing Recommendation Reports:
Understanding the Role of Recommendation Reports, using a Problem-Solving Model for Preparing Recommendation Reports, Writing Recommendation Reports.
Reviewing, Evaluating, and Testing Documents and Websites: Understanding Reviewing, Evaluating, and Testing, Reviewing Documents and Websites, Conducting Usability Evaluations, Conducting Usability Tests, Using Internet tools to check writing Quality, Duplicate Content Detector, what is Plagiarism? How to avoid writing plagiarism content?
Innovation management: an introduction: The importance of innovation, Models of innovation, Innovation as a management process.
Market adoption and technology diffusion: Time lag between innovation and usable product, Innovation and the market, Innovation and market vision, Analyzing internet search data to help adoption and forecasting sales, Innovative new products and consumption patterns, Crowd sourcing for new product ideas, Frugal innovation and ideas from everywhere, Innovation diffusion theories.

UNIT-5	**Managing Operations**		CO5, CO6

Managing innovation within firms: Organizations and innovation, the dilemma of innovation management, Innovation dilemma in low technology sectors, Dynamic capabilities, managing uncertainty, Managing innovation projects
Operations and process innovation:
Operations management, The nature of design and innovation in the context of operations, Process design, Process design and innovation
Managing intellectual property: Intellectual property, Trade secrets, An introduction to patents, Trademarks, Brand names, Copyright
Management of research and development: What is research and development? R&D management and the industrial context, R&D investment and company success, Classifying R&D, R&D management and its link with business strategy, Strategic pressures on R&D, Which business to support and how? Allocation of funds to R&D, Level of R&D expenditure
Managing R&D projects: Successful technology management, the changing nature of R&D management, the acquisition of external technology, Effective R&D management, the link with the product innovation process, Evaluating R&D projects.

Course Code	**Basics of Data Science with Python**		**L**	**T**	**P**	**C**
			2	-	2	4
Pre-requisites		**Semester**	**II**			

Course Objectives:

- Gain knowledge of the data science workflow and the role of Python in data analysis.
- Explore data preprocessing methods, including data cleaning, data transformation, and handling missing values.
- Master the skills to perform EDA using Python libraries such as Pandas, NumPy, and Matplotlib.
- Learn to implement supervised and unsupervised learning techniques using Python libraries like scikit-learn.
- Develop proficiency in visualizing data using Python libraries

Course Outcomes (CO):			
CO 1	Recall the fundamental concepts and principles of data science and its applications in various domains.		
CO 2	Understand the core concepts of data acquisition, preprocessing and exploratory data analysis using Python.		
CO 3	Implement functions, handle files, and manage exceptions in Python to create efficient and error-handling programs.		
CO 4	Analyze and interpret data using Python libraries like Pandas, NumPy, and Matplotlib for exploratory data analysis and visualization.		
CO 5	Design and develop Python programs for real-world case studies, applying data science techniques to solve complex problems.		
CO 6	Create visualizations and reports to effectively communicate data findings and insights, highlighting data-driven decision-making skills.		
UNIT-1	**Introduction**		CO1
Introduction to Data Science Current Landscape, Big data and Data Science, Importance of data science, life cycle of data science, Data Collection and management, tools and technologies used in data science, business intelligence vs data science, role of data scientist, Basic tools (plots, graphs and summary statistics) of Exploratory Data Analysis, Philosophy of EDA, The Data Science Process **Introduction to Python** Python identifiers, operators, operator precedence, getting input from user, decision making statements like if, if-else, if-elif, while loop, for loop, continue and break statement, Python Lists, Tuples, Dictionaries, Accessing Values, Basic Operations Indexing, Slicing, and Matrixes, Built-in Functions & Methods			
UNIT-2	**Functions, Files I/O and Exceptional Handling**		CO2
Defining Functions, Calling Functions, Functions with Multiple Arguments, Anonymous Functions – Lambda, Using Built-In Modules, User-Defined Modules, Module Namespaces, Iterators and Generators Opening and Closing Files, open Function, file Object Attributes, close () Method, Read, write, seek. Exception Handling, the try-finally Clause, Raising an Exception, User- Defined Exceptions, Regular Expression- Search and Replace, Regular Expression Modifiers, Regular Expression Patterns			
UNIT-3	**NumPy**		CO3, CO4
NumPy Introduction to NumPy, Array Creation, Printing Arrays, Basic Operations- Indexing, Slicing and Iterating, Shape Manipulation - Changing shape, stacking and splitting of array, Vector stacking,			
UNIT-4	**Pandas and Matplotlib**		CO5
Introduction to Pandas, importing data into Python, Pandas Data Frames, Indexing Data Frames, Basic Operations with Data frame, Renaming Columns, Subletting and filtering a data frame, Combining and Merging Data Frames, Removing Duplicates, Discretization and Binning, String Manipulation, Matplotlib - Introduction, plot (), Controlling Line Properties, Working with Multiple Figures, Histograms			
UNIT-5	**Case studies**		CO6
Default Modeling using Logistic Regression in Python Credit Risk Analytics using SVM in Python Intrusion Detection using Decision Trees & Ensemble Learning in Python etc.			

Books –

1. Introduction to Data Science by Jeffrey Stanton
2. Python: The Complete Reference by Martin Brown
3. Mastering Python for Data Science by Samir Madhavan
4. Python for Data Analysis by W Mckinney, O'Reilly Publications
5. Pandas for Everyone: Python Data Analysis by Daniel Chen
6. Python Data Science Handbook by Jake Vander Plas, O'Reilly Publications

Course Code	**Design Thinking**		**L**	**T**	**P**	**C**
			2	-	2	4
Pre-requisites		**Semester**	**II**			

Course Objectives:

- Understand the principles of design thinking
- To develop customer-centric perspectives
- To Foster creativity and ideation
- Acquire skills to create prototypes and models to visualize and communicate design concepts.
- Apply design thinking to real-world data challenges

Course Outcomes (CO):

CO 1	Recall the fundamental concepts and principles of design thinking and its applications in problem-solving.
CO 2	Understand the importance of mapping business hypotheses and defining the customer perspective in the design thinking process.
CO 3	Apply design thinking methods to frame and articulate design challenges.
CO 4	Create prototypes and models to represent design concepts and ideas.
CO 5	Develop strategies and plans for business launch and implementation based on design thinking principles.
CO 6	Evaluate the feasibility and viability of design solutions from a business perspective, considering market dynamics and potential impact.

UNIT-1	**Introduction**		CO1

Onboarding process: Welcome and Course Resources, what is Design Thinking: Introduction, Process, Modes, Importance in socio-economic context: WHY - Challenges, Awareness and Impact, Design thinking broader business picture: Broader aspects and impact, Multiple points of Interactions, The Product Form and the content

UNIT-2	**Business Hypothesis Mapping and Customer Perspective**		CO2, CO3

Need Analysis, Business Goals, Design Vision & Stakeholder mapping, what is hypothesis: Business Context and market analysis, Archetype Creation: Persona and Customer Journey mapping questionnaire, Market research vs. Design research, Types of research, Research scenario (Business Hypothesis Mapping)Identifying Customer need: Empathizing, what is Empathy, Difference between Sympathy & Empathy, Customer Perspectives, Recruitment process, Research (Ethnographic) methods: Observe. Immerse. Interact, Research Synthesis/Field work: observation & interview techniques, Archetype Creation: Persona,

Customer Journey Mapping preparation, Various observation & empathy frameworks, Supporting conceptual Models. User Models			
UNIT-3	**Design Challenge and Ideation**		CO3
Analysis & Synthesis: Research data prioritization/mapping, Data mapping (root cause) tools & techniques, Data interpretation. Developing insights, reframe challenge based on customer need and hypothesis validation, Design Challenge Summary: Final challenge, SCOPE and HMW, developing contextual conclusions, developing design response. **Ideation**: Creativity, Invention, Innovation, Various Thinking approaches for enhancing creativity, Ideation tools, Transformation, Brainwriting Methods, Conceptualization: Prioritizing ideas, Product Goals and Profile. User Experience Goals. Parameters and Weightage Perceptual Appropriation of Design Solution. Relevance and Validity, Design implications, product positioning, Sustainable design solution, standards, heuristics, affordance, principles			
UNIT-4	**Prototyping and Testing**		CO4
Prototyping Introduction, Iteration - Mindset for prototyping, Types of prototyping, Prototyping tools and techniques, Information architecture and design, Low and high-fidelity prototypes, handling complexity with simplicity Testing methods, Testing mindset: Planning and conducting User Testing, Heuristic evaluation, Expert usability testing, Feedback analysis and iteration, Revisiting Design Criterion, Preparing Guidelines, Recommendations			
UNIT -5	**Business Launch**		CO5, CO6
Impact Delivery: Revisiting entire process and project, Business goals and impact delivery, KPIs and Risk Prediction, Change Management, devising a preliminary Implementation Plan, What and How are we Delivering: Product, Service, Experience.			

Course Code	**Research Methodology**		**L**	**T**	**P**	**C**
			2	-	2	4
Pre-requisites		**Semester**	**II**			

Course Objectives:

1. identify and discuss the role and importance of research in the social sciences.
2. identify and discuss the issues and concepts salient to the research process.
3. identify and discuss the complex issues inherent in selecting a research problem, selecting an appropriate research design, and implementing a research project.
4. identify and discuss the concepts and procedures of sampling, data collection, analysis and reporting.

Course Outcomes (CO):

CO 1	Students who complete this course will be able to understand and comprehend the basics in research methodology and applying them in research/ project work.
CO 2	This course will help them to select an appropriate research design.
CO 3	With the help of this course, students will be able to take up and implement a research project/ study.
CO 4	The course will also enable them to collect the data, edit it properly and analyze it accordingly. Thus, it will facilitate students' prosperity in higher education.

CO 5	The Students will develop skills in qualitative and quantitative data analysis and presentation.
CO 6	Students will be able to demonstrate the ability to choose methods appropriate to research objectives.

UNIT-1	**Foundation of Research**		CO1

Foundations of Research: Meaning, Objectives, Motivation, Utility. Concept of theory, empiricism, deductive and inductive theory. Characteristics of scientific method – Understanding the language of research – Concept, Construct, Definition, Variable. Research Process,

Problem Identification & Formulation – Research Question – Investigation Question – Measurement Issues – Hypothesis – Qualities of a good Hypothesis –Null Hypothesis & Alternative Hypothesis. Hypothesis Testing – Logic & Importance

UNIT-2	**Research Design**		CO2

Research Design: Concept and Importance in Research – Features of a good research design – Exploratory Research Design – concept, types and uses, Descriptive Research Designs – concept, types and uses. Experimental Design: Concept of Independent & Dependent variables.

Qualitative and Quantitative Research: Qualitative research – Quantitative research – Concept of measurement, causality, generalization, replication. Merging the two approaches.

UNIT-3	**Measurements and Sampling**		CO3, CO4

Measurement: Concept of measurement– what is measured? Problems in measurement in research – Validity and Reliability. Levels of measurement – Nominal, Ordinal, Interval, Ratio.

Sampling: Concepts of Statistical Population, Sample, Sampling Frame, Sampling Error, Sample Size, Non-Response. Characteristics of a good sample. Probability Sample – Simple Random Sample, Systematic Sample, Stratified Random Sample & Multi-stage sampling. Determining size of the sample – Practical considerations in sampling and sample size

UNIT-4	**Data Analysis and Interpretation**		CO5

Data Analysis: Data Preparation – Univariate analysis (frequency tables, bar charts, pie charts, percentages), Bivariate analysis – Cross tabulations and Chi-square test including testing hypothesis of association.

Interpretation of Data and Paper Writing – Layout of a Research Paper, Journals in Computer Science, Impact factor of Journals, When and where to publish? Ethical issues related to publishing, Plagiarism and Self-Plagiarism.

UNIT-5	**Tools for Research**		CO6

Use of Encyclopedias, Research Guides, Handbook etc., Academic Databases for Computer Science Discipline. (5%) 10. Use of tools / techniques for Research: methods to search required information effectively, Reference Management Software like Zotero/Mendeley, Software for paper formatting like LaTeX/MS Office, Software for detection of Plagiarism

Books:

1. K. S. Bordens, and B. B.Abbott, , "Research Design and Methods – A Process Approach", 8th Edition, McGraw Hill, 2011
2. C. R. Kothari, "Research Methodology – Methods and Techniques", 2nd Edition, New Age International Publishers Second Semester
3. Douglas C. Montgomary & George C. Runger, Applied Statistics & probability for Engineers, 3rd edition,2007, Wiley
4. Robert P. Merges, Peter S. Menell, Mark A. Lemley, "Intellectual Property in the New Technological Age". Aspen Law & Business; 6th edition July 2012
5. A Beginners Guide to Latex, Chetan Shirore, 5 July 2015.
6. Business Research Methods – Donald Cooper & Pamela Schindler, TMGH, 9th edition

7. Business Research Methods – Alan Bryman & Emma Bell, Oxford University Press. 8. Research Methodology – C.R.Kothari 4. Select references from the Internet

Course Code	**Practical Approach to Data Mining & Analytics**		**L**	**T**	**P**	**C**
			2	-	2	4
Pre-requisites		**Semester**	**II**			

Course Objectives:

Data Mining studies algorithms and computational paradigms that allow computers to find patterns and regularities in databases, perform prediction and forecasting, and generally improve their performance through interaction with data. It is currently regarded as the key element of a more general process called Knowledge Discovery that deals with extracting useful knowledge from raw data. The knowledge discovery process includes data selection, cleaning, coding, using different statistical and machine learning techniques, and visualization of the generated structures.

Course Outcomes (CO):

CO 1	Compare several models and select and interpret the most appropriate for the data.
CO 2	Apply appropriate Data Mining techniques to analyze data.
CO 3	Critically evaluate a technical report.
CO 4	To introduce students to the basic concepts and techniques of Data Mining.
CO 5	To develop skills of using recent data mining software for solving practical problems.
CO 6	To gain experience of doing independent study and research.

UNIT-1	**Data Warehousing and Business Analysis**		CO1

Data warehousing Components –Building a Data warehouse –- Mapping the Data Warehouse to a Multiprocessor Architecture – DBMS Schemas for Decision Support – Data Extraction, Cleanup, and Transformation Tools –Metadata, Data Warehouse and OLAP, Data Warehouse and DBMS, Multidimensional data model, OLAP operations

UNIT-2	**Business Analysis**		CO2

Reporting and Query tools and Applications – Tool Categories – The Need for Applications – Cognos Impromptu – Online Analytical Processing (OLAP) – Need – Multidimensional Data Model – OLAP Guidelines – Multidimensional versus Multi-relational OLAP – Categories of Tools – OLAP Tools and the Internet.

UNIT-3	**Data mining, clustering and applications and trends in data mining**		CO3, CO4

Introduction – Data – Types of Data – Data Mining Functionalities – Interestingness of Patterns – Classification of Data Mining Systems – Data Mining Task Primitives – Integration of a Data Mining System with a Data Warehouse – Issues –Data Preprocessing, Cluster Analysis - Types of Data – Categorization of Major Clustering Methods – Kmeans – Partitioning Methods – Hierarchical Methods - Density-Based Methods –Grid Based Methods – Model-Based Clustering Methods – Clustering High Dimensional Data - Constraint – Based Cluster Analysis – Outlier Analysis – Data Mining Applications.

UNIT-4	**Association rule mining and classification**		CO5

Mining Frequent Patterns, Associations and Correlations – Mining Methods – Mining

Various Kinds of Association Rules – Correlation Analysis – Constraint Based Association Mining – Classification and Prediction - Basic Concepts - Decision Tree Induction - Bayesian Classification – Rule Based Classification – Classification by Backpropagation – Support Vector Machines – Associative Classification – Lazy Learners – Other Classification Methods – Prediction			
UNIT-5	**Advanced techniques, Data Mining software and application**		CO6
Text mining: extracting attributes (keywords), structural approaches (parsing, soft parsing), Bayesian approach to classifying text, Web mining: classifying web pages, extracting knowledge from the web, Data Mining software and applications			

Course Code	**Machine Learning**	**L**	**T**	**P**	**C**
		2	-	2	4
Pre-requisites	**Semester**	**I**			

Course Objectives:

The main objective of this course is to enable the student with basic knowledge on the techniques to build an intellectual machine for making decisions on behalf of humans. This course covers the techniques on how to make learning by a model, how it can be evaluated, what are all different algorithms to construct a learning model.

Course Outcomes (CO):

CO 1	Develop an appreciation for what is involved in learning from data.
CO 2	Understand a wide variety of learning algorithms.
CO 3	Understand how to apply a variety of learning algorithms to data.
CO 4	Understand how to perform evaluation of learning algorithms and model
CO 5	Understand detail topics of Machine Learning

UNIT-1	**Foundations for ML**		CO1
Introduction to Machine Learning: Definition, types of learning, applications. Supervised Learning: Classification, regression, over fitting, bias-variance trade-off. Unsupervised Learning: Clustering, dimensionality reduction. Evaluation Metrics: Accuracy, precision, recall, F1-score, ROC curves.			
UNIT-2	**Data Preprocessing and Feature Engineering**		CO2, CO4
Data Cleaning: Handling missing values, outliers. Feature Selection: Filter, wrapper, and embedded methods. Feature Transformation: Scaling, normalization, encoding categorical variables. Handling Imbalanced Data: Techniques for handling imbalanced datasets.			
UNIT-3	**Supervised Learning Algorithms**		CO3
Linear Regression: Simple and multiple linear regressions, regularization. Logistic Regression: Binary and multi-class classification. K-Nearest Neighbors (k-NN): Distance metrics, choosing k. Decision Trees and Random Forests: Entropy, Gini impurity, ensemble methods. Support Vector Machines (SVM): Linear and non-linear kernels, hyper parameters. Naive Bayes: Probability theory, Bayes' theorem, text classification.			
UNIT-4	**Unsupervised Learning Algorithms**		CO5
K-Means Clustering: Distance metrics, initialization methods. Hierarchical Clustering: Agglomerative and divisive methods. Principal Component Analysis (PCA): Dimensionality reduction, eigenvectors. Recommender Systems: Collaborative			

filtering, content-based filtering.			
UNIT-5	Advanced Topics in Machine Learning		CO2
Neural Networks and Deep Learning: Feed forward networks, back propagation. Convolutional Neural Networks (CNN): Image classification, convolution layers. Recurrent Neural Networks (RNN): Sequence modeling, LSTM, GRU. Ensemble Learning: Bagging, boosting, stacking. Model Selection and Hyper parameter Tuning: Cross-validation, grid search.			

Books –

1. Machine Learning: A Probabilistic Perspective: Kevin P Murphy, The MIT Press Cambridge (2012).
2. Introducing Monte Carlo Methods with R, Christian P. Robert, George Casella, Springer, 2010
3. Introduction to Machine Learning (Third Edition): Ethem Alpaydın, The MIT Press (2015).
4. Pattern Recognition and Machine Learning: Christopher M. Bishop, Springer (2006)
5. Bayesian Reasoning and Machine Learning: David Barber, Cambridge University Press (2012).
6. Statistical And Machine Learning Approaches For Network Analysis, Edited By Matthias Dehmer, Subhash C. Basak: John Wiley & Sons, Inc (2012)

Course Code	**Optimization and Simulation for Data Science**		**L**	**T**	**P**	**C**
			2	-	2	4
Pre-requisites		**Semester**	**II**			

Course Objectives:

This course offers an introduction to mathematical nonlinear optimization with applications in data science. The theoretical foundation and the fundamental algorithms for nonlinear optimization are studied and applied to supervised learning models, including nonlinear regression, logistic regression, support vector machines, and deep neural networks.

Course Outcomes (CO):

CO 1	Understand the techniques of modeling in the context of hierarchy of knowledge about a system and develop the capability to apply the same to study systems through available software.
CO 2	Students will learn different types of simulation techniques.
CO 3	Students will learn to simulate the models for the purpose of optimum control by using software

UNIT-1			CO1

Linear Programming

Central Problem of linear Programming,Various definitions including Statements of basic theorem and also their properties ,Simplex methods ,Primal and dual simplex method Transport problem,Tic-Tac problem and its solution Assignment problem and its solution Graphical Method Formulation Linear Programming Problem

Linear Programming Applications

Use of software for solving linear optimization problems using graphical and simplex methods Examples for transportation, assignment, water resources, structural and other

optimization problems

UNIT-2			CO2

Introduction to Simulation, Need of Simulation, Time to simulate, Inside simulation software: Modeling the progress of Time, Modeling Variability, Conceptual Modeling: Introduction to Conceptual modeling, Defining conceptual model, Requirements of the conceptual model, Communicating the conceptual model, Developing the Conceptual Model: Introduction, A framework for conceptual modeling, methods of model simplification.

UNIT-3			CO3

Model Verification and Validation

Data Collection and Analysis: Introduction, Data requirements, Obtaining data, Representing unpredictable variability, Selecting statistical distributions. Obtaining Accurate Results: Introduction, The nature of simulation models and simulation output, Issues in obtaining accurate simulation results, example model, dealing with initialization bias: warm-up and initial conditions, Selecting the number of replications and run-length. Searching the Solution Space: Introduction, The nature of simulation experimentation, Analysis of results from a single scenario, Comparing alternatives, Search experimentation, and Sensitive analysis. Verification, Validation and Confidence: Introduction, Defining Verification and Validation, The difficulties of verification and validation, Methods of verification and validation, Independent verification and validation.

UNIT-4			CO3

Modeling and simulation modeling

Types of models, Analytical vs Simulation modeling, Application of simulation modeling, Level of abstraction, Simulation Modeling. Methods, System Dynamics, Discrete Event Modeling, Agent Based modeling: Introduction to Agent, Agent-based modeling**,** Time in agent based models, Space in agent based models, Discrete space, Continuous space movement in continuous space, Communication between agents, Dynamic creation and destruction of agents, Statics on agent population, Condition triggered events and transition in agents. Building agents-based models: The problem statement, Phases of modeling, Assumptions, 3 D animation. Dynamics Systems: Stock and flow diagrams, examples of stock and flow diagrams. Multi- method modeling: Architecture, Technical aspects of combining modeling methods, Examples.

UNIT-5			CO2

Design and behavior of models

Designing state-based behavior: State charts, State transitions, Viewing and debugging State charts at runtime, State charts for dynamic objects. Discrete events and Event model object: Discrete event, Event-the simplest low level model object, Dynamic events, and Exchanging data with external world. Presentation and animation: Working with shapes, groups and colors, designing interactive models: using controls, Dynamic properties of controls, 3D Animation. Randomness in Models: Probability distributions, sources of randomness in the model, randomness in system dynamics model, random number generators, Model time, date and calendar: Virtual and real time: The model time, date and calendar, Virtual and real-time execution modes.

Books –

1. Simulation: The Practice of Model Development and Use by Stewart Robinson, John Wiley and Sons, Ltd, 2004.
2. The Big Book of Simulation Modeling: Multi Method Modeling by Andrei Borshchev, 2013.
3. Agent Based Modeling and Simulation, Taylor S, 2014.
4. Simulation Modeling Handbook: A Practical Approach, Christopher A. Chung, 2003.

5. Object Oriented Simulation: A Modeling and Programming Perspective, Garrido, José M, 2009.
6. Simulation, Modeling and Analysis, Averill M Law and W. David Kelton, "Tata McGraw Hill, Third Edition, 2003.
7. Process Control: Modeling, Design and Simulation, Wayne Bequette W, Prentice Hall of India, 2003.

Course Code	Advance Data Structures and algorithm		L	T	P	C
			2	-	2	4
Pre-requisites		**Semester**	**II**			

Course Objectives:

• Learn asymptotic notations and analyze the performance of different algorithms.
• Understand and implement various data structures.
• Learn and implement greedy, divide and conquer, dynamic programming and backtracking algorithms using relevant data structures.
• Understand non-deterministic algorithms, polynomial and non-polynomial problems.

Course Outcomes (CO):

CO 1	Analyze the complexity of algorithms and apply asymptotic notations.
CO 2	Apply non-linear data structures and their operations.
CO 3	Understand and apply greedy, divide and conquer algorithms.
CO 4	Develop dynamic programming algorithms for various real-time applications.
CO 5	Illustrate Backtracking algorithms for various applications.

UNIT-1	**Introduction to Algorithms**		CO1

Algorithms, Pseudo code for expressing algorithms, Performance Analysis-Space complexity, Time complexity, Asymptotic Notation- Big oh, Omega, Theta notation and Little oh notation, Polynomial Vs Exponential Algorithms, Average, Best and Worst Case Complexities, Analyzing Recursive Programs.

UNIT-2	**Trees Part I**		CO2

Binary Search Trees: Definition and Operations, AVL Trees: Definition and Operations, Applications. B Trees: Definition and Operations.

UNIT-3	**Trees Part II**		CO4

Red-Black Trees, Splay Trees, Applications.
Hash Tables: Introduction, Hash Structure, Hash functions, Linear Open Addressing, Chaining and Applications.

UNIT-4	**Divide and conquer, Greedy method**		CO3

Divide and conquer: General method, applications-Binary search, Finding Maximum and minimum, Quick sort, Merge sort, Strassen's matrix multiplication. Greedy method: General method, applications-Job sequencing with deadlines, knapsack problem, Minimum cost spanning trees, Single source shortest path problem.

UNIT-5	**Dynamic Programming & Backtracking**		CO5

Dynamic Programming: General method, applications- 0/1 knapsack problem, All pairs shortest path problem, Traveling salesperson problem, Reliability design.
Backtracking: General method, applications-n-queen problem, sum of subsets problem, graph coloring, Hamiltonian cycles.

Introduction to NP-Hard and NP-Complete problems: Basic Concepts.
REFERENCES: 1. Data Structures and algorithms: Concepts, Techniques and Applications, G A V Pai. 2. Fundamentals of Computer Algorithms, Ellis Horowitz, Sartaj Sahni and Rajasekharam, Galgotia publications Pvt. Ltd. 3. Classic Data Structures by D. Samanta, 2005, PHI 4. Design and Analysis of Computer Algorithms by Aho, Hopcraft, Ullman 1998, PEA. 5. Introduction to the Design and Analysis of Algorithms by Goodman, Hedetniemi, TMG. 6. Online Learning Resources: https://www.tutorialspoint.com/advanced_data_structures/index.asp http://peterindia.net/Algorithms.htm

Course Code	**Sentiment, Web and Text Analytics**		**L**	**T**	**P**	**C**
			2	-	2	4
Pre-requisites		**Semester**	**II**			

Course Objectives:

With the rapid growth of unstructured natural language data, such as web pages, blogs, product reviews, news articles and enterprise data, there is an increasing need for systems that would retrieve relevant documents, extract specific information from them, mine opinions, summarize and categorize texts. This course provides students with an understanding of the major methods for retrieving, mining and analyzing textual data, with the emphasis on algorithms, techniques and their evaluation.

Course Outcomes (CO):

CO 1	To Understand Explain the text analytics framework.
CO 2	Analyze various sources of text data.
CO 3	Measure machine learning model performance with appropriate metrics.
CO 4	Interpret the results, gain insights, and recommend possible actions from analytics performed on text data.
CO 5	Text analytics are the methods and techniques used to extract useful knowledge from text to support decision making.

UNIT-1	**Introduction**		CO1

Introduction, Sentiment analysis applications, Sentiment analysis research, Sentiment analysis as mini-NLP, The Problem of Sentiment Analysis, Definition of opinion, Definition of opinion summary, different types of opinions, Document Sentiment Classification, Supervised sentiment classification, Unsupervised sentiment classification, Sentiment rating prediction

UNIT-2	**Sentence Subjectivity and Sentiment Classification**		CO3

Sentence Subjectivity and Sentiment Classification, Subjectivity, Sentence Subjectivity Classification, Sentence Sentiment Classification, Aspect Sentiment Classification, Rules of Sentiment composition, Negation and Sentiment, Aspect and Entity Extraction, Frequency based aspect extraction, exploring syntactic relations, Using supervised learning

UNIT 3	**Sentiment Lexicon Generation**		CO4

Sentiment Lexicon Generation, Dictionary based approach, Corpus based approach, Sentiment word embedding, Analysis of Comparative Opinions, Problem definition, identifying comparative sentences, Identifying the preferred entity set, Special types of comparison, Opinion Summarization and Search, Aspect based opinion summarization, enhancements to aspect based summaries, Traditional summarization			
UNIT 4	**Analysis of Debates and Comments**		CO2
Analysis of Debates and Comments, recognizing stances in debates, Modeling debates/Discussions, Modeling comments, Mining Intents, Problem of intent mining, Intent classification, Fine grained mining of intent, Detecting Fake or Deceptive Opinions, Different types of Spam, Supervised fake review detection, Automated discovery of abnormal patterns, Model based behavioral analysis, Group spam detection, Quality of Reviews, Quality prediction as a regression problem			
UNIT 5	**Applications**		CO5
Autoencoders and their NLP applications. Language modeling with autoencoders, Sequence-to-sequence models. Machine translation. Dialogue systems. Adversarial and multi-task learning for NLP.			
REFERENCES: The NLTK book: http://www.nltk.org/book/ Kumar, A., and Paul, A. (2016). Mastering Text Mining with R. Packt Publishing. (E-book). Ravindran, S. K., and Garg, V. (2015). Mastering Social Media Mining with R. Packt Publishing (E-book).			

Course Code	**Internship with Project**		**L**	**T**	**P**	**C**
Pre-requisites		**Semester**	**II**			

Course Objectives:

1. Will expose students to the industrial environment and hence creating competent professionals for the industry.
2. Provide opportunities to learn and sharpen the real time technical skills required in niche area of Data Analytics.
3. Exposure to the latest technological developments relevant to the subject area of training.
4. Create conditions conducive to quest for knowledge and its applicability on the job
5. Learn to apply the technical knowledge in real industrial situations.
6. Gain experience in writing technical reports/projects.

Course Outcomes (CO):

CO 1	Interact with people working in the same area
CO 2	Improve communication and critical thinking ability
CO 3	Meet and resolve risks involved in the specified area

The syllabus proposes an internship for about 10 weeks to 12 weeks to be done by a student. It is expected that a student chooses an IT or IT-related industry providing Internship in the field of Data Science and Big Data Analytics and formally works as a

full-time intern during the period. The student should subject oneself with an internship evaluation with proper documentation of the attendance and the type of work he or she has done in the chosen organization. Proper certification by the organisation to whom the student was reporting, with Organization's seal should be attached as part of the documentation.

*****