SIES College of Arts, Science & Commerce
RISE WITH EDUCATION
Sion (West), Mumbai – 400022.
(Autonomous)

Program: B.Sc.
Subject: Data Science
Academic year: 2025-2026

Syllabus for Semester- V and Semester – VI

Choice Based Credit System Syllabi (as per NEP)

approved by Board of Studies in Data Science

To be brought into effect from June 2025.

PREAMBLE

Data has become the most important factor in this era of digital transformation. The technological advancements are seen in all walks of life and therefore we are flooded with enormous data. Every business relies on data to deliver better products as well as services. All data are stored in cloud, and so accessed and processed easily. Data analytics has helped in better decision making with sufficient data insights.

Predictive Analysis has played a crucial role in making businesses smarter with improvised strategies. Machine Learning and Artificial Intelligence are used together to optimize business operations and data management. Augmented analytics uses machine learning and natural language processing to automate the process of data analysis. Global data is predicted to grow due to data generated by the Internet of Things(IoT) and cloud computing advancements. These developments have given rise to a new area of study, called Data Science.

Data Science as an area has evolved out of the applications of various tools and techniques in the field of Computer Science, Mathematics and Statistics. There is an increasing demand to capture, analyse the enormous data present in a number of application domains. The data in these applications then needs to be converted into actionable strategies for effective decision making. So, the study of data science has become essential to cater to the growing need of data scientists and data analysts.

This course focuses on educating the students about the essentials of computer science, applied mathematics and applied statistics with respect to the data science applications.

PROGRAM OUTCOMES AND PROGRAM SPECIFIC OUTCOMES

NO.	Details
PO1.	**Solving Complex Problems**:- Apply the knowledge gained in breaking down complex problems into simple components; and to design processes required for problem solving.
PO2.	**Critical Thinking: -** Ability to apply the acquired knowledge to identify assumptions and evaluate their accuracy and validity.
PO3.	**Reasoning ability and Rational thinking: -** Ability to analyse, interpret data and draw logical conclusions; to evaluate ideas rationally.
PO4.	**Research Aptitude: -** Ability to ask relevant questions to identify and define the problem, applying research tools for analysis and interpretation of data. Understand and comply with research ethics.
PO5.	**Effective Communication skill: -** Demonstrate the ability to listen and to clearly express ideas verbally. Equip to write reports, make presentations effectively.
PO6.	**Information and Digital Literacy: -** Equip to use appropriate tools and techniques inclusive of internet and electronic media for acquiring, assessing and analysing data from diverse resources.
PO7.	**Social Interactive Skills and team work: -** Exhibit networking and social interactive skills; function effectively as an individual and as a member in diverse groups; demonstrate leadership quality useful for employability
PO8.	**Self-directed and Lifelong Learning:** Ability to explore and gain knowledge in independent and self-reliant ways. Demonstrate ability to adapt and upgrade with the global , social and technological changes.
PO9.	**Awareness towards Environment and Sustainable Development:** Exhibit awareness and a concern for environmental issues; understand and realize the significance of co-habitation and co-evolution in attaining the needs of sustainable development.
PSO1.	**Sound Knowledge:** Demonstrate the knowledge of core data science concepts and apply them to develop a user- friendly, scalable, and robust applications
PSO2.	**Critical and Rational Thinking:** Exhibit higher order skills to adapt to the everchanging technological environment
PSO3.	**Logic Building and Programming Skills:** The ability to apply logic to problem solving and acquire proficiency in various programming languages.
PSO4.	**Data Analysis :** Apply quantitative modelling and data analysis techniques to solve real world business problems, Learn tools and techniques for transformation of data and statistical data analysis

PSO5.	**Pursue Higher Education:** Make students competent to take up advanced degree courses like MSc(Data Science),MCA, MSc(CS), MSc(IT) and MBA etc.

Course Code	Course Type	Course Title	Credits
SIUDSMJ311	Core Subject	Introduction to AI and ML	**3**
SIUDSMJ312	Core Subject	NoSQL and Data Security	**3**
SIUDSMJ313	Core Subject	Research Methodology	**3**
SIUDSEL311	Core Subject (Elective)	Social Media Analytics	**3**
SIUDSMN311	Minor	Advanced Web Technology	**1**
SIUDSVS311	Vocational Skill Course (VSC)	Data Analysis & Visualization with Microsoft Power BI	**1**
	FP/CEP	Field Project/Community Engagement Program	**2**
SIUDSMJP311	Core Subject Practical	Introduction to AI and ML Practical (BigMl, rapid minor)	**1**
SIUDSMJP312	Core Subject Practical	NoSQL and Data Security Practical	**1**
SIUDSMJP313	Core Subject Practical	Research Project	**1**
SIUDSELP311	Core Subject Practical	Social Media Analytics Practical	**1**
SIUDSMNP311	Minor Practical	Advanced Web Technology	**1**
SIUDSVS311	Vocational Skill Course Practical (VSC)	Data Analysis & Visualization with Microsoft Power BI Practical	**1**
		TOTAL CREDITS	**22**

T.Y. B.Sc. (DS) – Semester V

SIUDSMJP31: Introduction to Artificial Intelligence and Machine Learning

B.Sc. (Data Science)	Semester – V
Course Name: Artificial Intelligence and Machine Learning	**Course Code: SIUDSMJ311**
Credits	**3**

Course Objective:

- To provide the foundations for AI problem-solving techniques and knowledge representation formalisms
- Understanding Human learning aspects.
- Understanding primitives in learning process by computer.
- Understanding nature of problems solved with Machine Learning

Course Outcomes:

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

- **CO1:** Get acquainted with AI techniques
- **CO2:** Identify problem formulation and problem-solving techniques
- **CO3:** Learn Heuristic Search to find solution in proper time
- **CO4:** Optimize mathematical problems in AI
- **CO5:** Apply the planning techniques and machine learning strategies

Unit	Contents	No. of Lectures
I	**Introduction to AI:** The AI problems, AI technique, philosophy and development of Artificial intelligence. Minimax algorithm, alpha-beta pruning, stochastic games, Constraint satisfaction problems. Knowledge and Reasoning: Logical agents, Propositional logic, First-order logic, **Inference in FoL:** forward chaining, backward chaining, resolution, Knowledge representation: Frames, Ontologies, Semantic web and RDF. **Introduction to PROLOG:** Facts and predicates, data types, goal finding, backtracking, simple object, compound objects, use of cut and fail predicates, recursion, lists, simple input/output, dynamic database. **Basics of Artificial Neural Networks:** Characteristics of Neural Networks, Historical Development of Neural Network Principles, Artificial Neural Networks: Terminology, Models of Neuron,	15

	Topology, Basic Learning Laws, Feedback Neural Networks: Introduction, Analysis of Linear Auto associative FF Networks, Analysis of Pattern Storage Networks.	
II	**Machine Learning:** Machine learning, Examples of Machine Learning Problems, Structure of Learning, learning versus Designing, Training versus Testing, Characteristics of Machine learning tasks, Predictive and descriptive tasks, Machine learning Models: Geometric Models, Logical Models, Probabilistic Models. **Features:** Feature types, Feature Construction and Transformation, Feature Selection. **Classification and Regression: Classification:** Binary Classification- Assessing Classification performance, Class probability Estimation Assessing class probability Estimates, Multiclass Classification. **Regression:** Assessing performance of Regression- Error measures, Overfitting- Catalysts for Overfitting, **Theory of Generalization:** Effective number of hypotheses, Bounding the Growth function, VC Dimensions, Regularization theory.	15
III	**Linear Models:** Least Squares method, Multivariate Linear Regression, Regularized Regression, Using Least Square regression for Classification. Perceptron, Support Vector Machines, Soft Margin SVM, Obtaining probabilities from Linear classifiers, Kernel methods for non-Linearity. **Logic Based and Algebraic Model: Distance Based Models:** Neighbours and Examples, Nearest Neighbours Classification, Distance based clustering-K means Algorithm, Hierarchical clustering, **Rule Based Models:** Rule learning for subgroup discovery, Association rule mining. **Tree Based Models:** Decision Trees, Ranking and Probability estimation Trees, Regression trees, Clustering Trees. Probabilistic Model: Normal Distribution and Its Geometric Interpretations, Naïve Bayes Classifier.	15

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Artificial Intelligence	Elaine Rich, Kevin Knight	Tata McGraw Hill	Third Edition	2017
2	Machine Learning: The Art and Science of Algorithms that Make Sense of Data	Peter Flach	Cambridge University Press	First Edition	2012
3	Introduction to Statistical Machine Learning with Applications in R	Hastie, Tibshirani, Friedman	Springer	Second Edition	2012
4	Introduction to Machine Learning	EthemAlpaydin	PHI	Second Edition	2012

SIUDSMJP311: Artificial Intelligence and Machine Learning Practical

B.Sc. (Data Science)	Semester – V
Course Name: Artificial Intelligence and Machine Learning Practical	**Course Code: SIUDSMJP311**
Credits	**1**

List of Practical: (Use python)

1.	Write a program to implement depth first search algorithm.
2.	Write a program to implement breadth first search algorithm.
3.	Write a prolog program to implement simple diagnostic tool for medical condition (eg. define a set of rules and facts about different medical conditions and their symptoms, and then use those rules and facts to diagnose a patient's condition based on their reported symptoms).
4.	Write a program to implement K-Nearest Neighbor (KNN) algorithm.
5.	Write a program to implement Naïve Bayes algorithm (Make use of titanic.csv).
6.	Write a program to implement Linear Regression (eg. to predict home prices).
7.	Write a program to implement Logistic Regression (eg. to predict whether customer will buy insurance or not).
8.	Write a program to build a decision tree algorithm (e.g. to predict whether employee get more than 100k salary).
9.	Write a program to develop a Support Vector Machine (SVM) model (e.g. to classify whether a user purchased a particular product or not, make use of Social_Network_Ads.csv).
10.	Write a program to implement K-Means Clustering algorithm (Make use of salary_data.csv).
11.	Write a program to implement Apriori algorithm.

SIUDSMJP312: NoSQL Database and Security

B.Sc. (Data Science)	**Semester – V**
Course Name: Introduction to NoSQL Database and Security	**Course Code: SIUDSMJ312**
Credits	**3**

Course Objective:

- Understand the installation and operation of NOSQL databases
- Have an overview of software like MongoDb, Cassandra, HBASE, Neo4
- Acquire knowledge about replication and sharding
- Understand basic details about HBASE and Riak Operations
- Explain graph NOSQL databases

Course Outcomes:

After completion of this course, student will be able to:

CO1: Understand the basic concepts of NoSql databases
CO2: Execute replication and sharding in document databases using MongoDB
CO3: Implement the Key-Value databases using Riak
CO4: Apply the Column databases using Apache Cassandra
CO5: Evaluate the Graph databases using Neo4j

Unit	Contents	No. of Lectures
I	**Introduction:** Why NoSQL? The Value of Relational Databases, Getting at Persistent Data, Concurrency, Integration, A (Mostly) Standard Model, Impedance Mismatch, Application and Integration Databases, Attack of the Clusters, The Emergence of NoSQL. **Aggregate Data Models:** Aggregates, Example of Relations and Aggregates, Consequences of Aggregate Orientation, Key-Value and Document Data Models, Column-Family Stores, Summarizing Aggregate-Oriented Databases. More Details on Data Models; Relationships, Graph Databases, Schemaless Databases, Materialized Views, Modeling for Data Access. **Distribution Models:** Single Server, Sharding, Master-Slave Replication, Peer-to-Peer Replication, Update Consistency, Read	15

	Consistency, Relaxing Consistency, The CAP Theorem, Relaxing Durability, Quorums. Version Stamps, Business and System Transactions, Version Stamps on Multiple Nodes	
II	**Map-Reduce:** Basic Map-Reduce, Partitioning and Combining, Composing Map-Reduce Calculations, A Two Stage Map-Reduce Example, Incremental Map-Reduce Key-Value Databases, What Is a Key-Value Store, Key-Value Store Features, Consistency, Transactions, Query Features, Structure of Data, Scaling, Suitable Use Cases, Storing Session Information, User Profiles, Preference, Shopping Cart Data, When Not to Use, Relationships among Data, Multioperation Transactions, Query by Data, Operations by Sets. **Document Databases:** What Is a Document Database? Features, Consistency, Transactions, Availability, Query Features, Scaling, Suitable Use Cases, Event Logging, Content Management Systems, Blogging Platforms, Web Analytics or Real-Time Analytics, E-Commerce Applications, When Not to Use, Complex Transactions Spanning Different Operations, Queries against Varying Aggregate Structure **Graph Databases:** What Is a Graph Database? Features, Consistency, Transactions, Availability, Query Features, Scaling, Suitable Use Cases, Connected Data, Routing, Dispatch, and Location-Based Services, Recommendation Engines, When Not to Use.	15
III	**Security:** The Importance of Information Protection, The Evolution of Information Security, Security Methodology, Risk Analysis: Threat Definition, Types of Attacks, Risk Analysis. Secure Design Principles: The CIA Triad and Other Models, Defense Models, Zones of Trust, Best Practices for Network Defense. **Authentication and Authorization:** Authentication, Authorization Encryption: A Brief History of Encryption, Symmetric-Key Cryptography, Public Key Cryptography, Public Key Infrastructure **Web and Database security techniques:** SQL Injection, Cross Site Scripting, Database privileges, Multilevel databases, Query modification, social engineering and Phishing Attacks	15

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence	Pramod J. Sadalage, Martin Fowler	Pearson Education		2012

2	NoSQL for Mere Mortals	Dan Sullivan	Addison-Wesley		2015
3	NoSQL for Dummies	Adam Fowler	John Wiley & Sons		2015
4	Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement	Luc Perkins, Eric Redmond, Jim Wilson	O′Reilly	2nd	2018

SIUDSMJP312: Introduction to NoSQL Database Practical

B.Sc. (Data Science)	**Semester – V**
Course Name: Introduction to NoSQL Database Practical	**Course Code: SIUDSMJP312**
Credits	**1**

List of Practical:

1	Demonstrate Installation of MongoDB and working with CLI and MongoDB Compass
2	Demonstrate creation of database, collection, listing existing databases and collections
3	Demonstrate Insertion of simple and complex datatypes
4	Demonstrate simple and complex Read operations
5	Demonstrate Update, Delete and Drop Operation
6	Demonstrate Aggregation pipeline and regression expressions
7	Demonstrate Simple, Compound and Geo Spatial Indexing
8	Demonstrate Cursors
9	Demonstrate Map Reduce Operations
10	Demonstrate importing data from CSV, EXCEL and TSV files

SIUDSMJ313: Research Methodology

B.Sc. (Data Science)	**Semester – V**
Course Name: Research Methodology	**Course Code: SIUDSMJ313**
Credits	**3**

Course Objective:

- To impart analytical skill in solving complex problems.
- To foster the ability to critically think in developing robust, extensible and highly maintainable solutions to simple and complex problems.
- To explore the unknown and unlock new possibilities in different dimensions of the system.
- To portray accurately the characteristics of a particular individual, situation or a group under study

Course Outcomes:

Upon completion of this course, student will be able to:

CO1: Learner understands the reasons for doing research, the applications of research, characteristics and requirements of the research process, types of research and Research paradigms.
CO2: Learner is applying major approaches to information gathering, the relationship between attitudinal and measurement scales Methods for exploring attitudes in research.
CO3: Learner is able to analyze data in qualitative and quantitative studies, application of IT in data analysis.
CO4: Learner is able to write a research report and use Information Technology in Research
CO5: Learner is practicing ethical codes and practices of conducting research.

Unit	Contents	No. of Lectures
I	**Research Methodology:** An Introduction: Meaning of Research, Objectives of Research, Motivation in Research, Types of Research, Research Approaches, Significance of Research, Research Methods versus Methodology, Research and Scientific Method, Importance of Knowing How Research is Done, Research Process, Criteria of Good Research.	15

	Defining the Research Problem: What is a Research Problem? Selecting the Problem, Necessity of Defining the Problem, Technique Involved in Defining a Problem, An Illustration. **Research Design:** Meaning of Research Design, Need for Research Design, features of a Good Design, Important Concepts Relating to Research Design, Different Research Designs, Basic Principles of Experimental Designs.	
II	**Sampling Design:** Sampling Design, steps in sample design, criteria of selecting a sampling procedure, characteristics of a good sample design, different types of sample design. **Measurement and Scaling Techniques:** Measurements in Research, Measurement Scales, Sources of errors in measurement, Collection of primary data, Observation Method, Interview Method, through questionnaires, through schedules, difference between questionnaire and schedule, Collection of secondary data, Selection of appropriate methods for data collection, Case study method. **Processing and Analysis of Data:** Data processing, processing operations: editing, coding, classification, tabulation, graphical representation, types of analysis, Statistics in research, Dispersion and Asymmetry, Measures of Relationship	15
III	**Sampling Fundamentals:** Need for Sampling, Some Fundamental Definitions, Important Sampling Distributions, Central Limit Theorem, Sampling Theory, Sandler's A-test, Concept of Standard Error, Estimation, Estimating the Population Mean (m), Estimating Population Proportion, Sample Size and its Determination, Determination of Sample Size through the Approach Based on Precision Rate and Confidence Level, Determination of Sample Size through the Approach, Based on Bayesian Statistics **Testing of Hypotheses:** What is a Hypothesis? Basic Concepts Concerning Testing of Hypotheses, Procedure and Flow diagram for Hypothesis Testing, Measuring the Power of a Hypothesis Test, Tests of Hypotheses, Hypothesis Testing of Correlation Coefficients and Limitations of the Tests of Hypotheses. **Chi-Square Test:** Chi-Square Test for Comparing Variance, Chi-square as a Non-parametric Test, Conditions for the Application of Chi-Square Test, Steps Involved in Applying Chi-square Test, Important Characteristics of Chi-Square Test and caution in using Chi-Square test.	15

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Research Methodology – Methods and techniques	C. R. Kothari	New Age International (P) Ltd., Publishers	---	---
2	Business Research Methods	Donald R. Cooper Pamela Schindler	Mc Graw Hill/ Irwin	12th Ed	---
3	RESEARCH METHODOLOGY - a step-by-step guide for beginners	Ranjit Kumar	SAGE Publication Ltd	---	---

SIUDSMJP313: Research Methodology Practical

B.Sc. (Data Science)	**Semester – V**
Course Name: Research Methodology	**Course Code: SIUDSMJP313**
Credits	**1**

List of Practical:

1	**Introduction to LaTex**
a.	**Report Writing:** report style having chapter, section and subsection, article style having section, subsection and subsubsection, Automatic generation of table of contents, toc file to store the information that goes into the table of contents, Automatic numbering of section numbers.
b.	**Equations and Numbering Equations:** Creating an equation, writing multiple equations, Aligning multiple equations, creating matrices in Latex, label command, Cross referencing with ref command.
c.	**Tables and Figures:** Tables and Figures Creating tables and figures in LaTeX
d.	**Bibliography:** Bibliography Creating Bibliography in LaTeX
2	**Introduction to EndNote, Zotero or Mendeley**
a.	Integration with Word and adding citation and creating bibliographies
b.	Creating your own library
c.	Creating references from website
d.	Creating references manually
3	Visit the college library or nearby research center or from the internet collect 5 titles of research papers/thesis and classify them according to types of research, discuss how the problems are delineated, how they are relevant to scientific method etc.
4	Identify 2 researchable problems relevant to your context and knowledge disciplines and justify the significance of their study.
5	Preparation of a review article
6	Identification of variables of a research study and their classification in terms of functions and level of measurement
7	Preparation of a sampling design given the objectives and research questions/hypotheses of a research study
8	Preparation of questionnaire for micro-level educational survey

9	Prepare 1 proposal on an identified research problem
10	Checking and removing plagiarism using Plagiarism Detection Software

SIUDSEL311: Social Media Analytics

B.Sc. (Data Science)	**Semester – V**
Course Name: Social Media Analytics	**Course Code: SIUDSEL311**
Credits	**3**

Course Objective:

- Familiarize the learners with the concept of social media.
- Familiarize the learners with the concept of social media analytics and understand its significance.
- Enable the learners to develop skills required for analyzing the effectiveness of social media.
- Familiarize the learners with different tools of social media analytics.
- Familiarize the learner with different visualization techniques for Social media analytics.

Course Outcomes:

After completion of this course, student will be able to:
CO1: Understand the concept of social media Analytics and its significance.
CO2: Learners will be able to analyze the effectiveness of social media
CO3: Learners will be able to use different Social media analytics tools effectively and efficiently.
CO4: Learners will be able to use different effective Visualization techniques to represent social media analytics.
CO5: Acquire the fundamental perspectives and hands-on skills needed to work with social media data.

Unit	Contents	No. of Lectures
I	**Social Media Analytics: An Overview** Core Characteristics of Social Media, Types of Social Media, Social media landscape, Need for Social Media Analytics (SMA), SMA in small & large organizations. Purpose of Social Media Analytics, Social Media vs. Traditional Business Analytics, Seven Layers of Social Media Analytics, Types of Social Media Analytics, Social Media Analytics	15

	Cycle, Challenges to Social Media Analytics, Social Media Analytics Tools **Social Network Structure, Measures & Visualization** **Basics of Social Network Structure** - Nodes, Edges & Tie, **Describing the Networks Measures** - Degree Distribution, Density, Connectivity, Centralization, Tie Strength & Trust **Network Visualization** - Graph Layout, Visualizing Network features, Scale Issues,	
II	**Social Media Network Analytics** - Common Network Terms, Common Social Media Network Types, Types of Networks, Common Network Terminologies, Network Analytics Tools. **Social Media Text, Action & Hyperlink Analytics** **Social Media Text Analytics** - Types of Social Media Text, Purpose of Text Analytics, Steps in Text Analytics, Social Media Text Analysis Tools **Social Media Action Analytics** - What Is Actions Analytics? Common Social Media Actions, Actions Analytics Tools **Social Media Hyperlink Analytics** - Types of Hyperlinks, Types of Hyperlink Analytics, Hyperlink Analytics Tools	15
III	**Social Media Location & Search Engine Analytics** **Location Analytics** - Sources of Location Data, Categories of Location Analytics, Location Analytics and Privacy Concerns, Location Analytics Tools, **Search Engine Analytics** - Types of Search Engines, Search Engine Analytics, Search Engine Analytics Tools **Social Information Filtering** **Social Information Filtering** - Social Sharing and filtering, Automated Recommendation systems, Traditional Vs social Recommendation Systems, Understanding Social Media and Business Alignment, Social Media KPI, Formulating a Social Media Strategy, Managing Social Media Risks **Social Media Analytics Applications and Privacy** **Social media in public sector** - Analyzing public sector social media, analyzing individual users, case study. **Business use of Social Media** - Measuring success, Interaction and monitoring, case study. **Privacy** - Privacy policies, data ownership and maintaining privacy online	15

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Seven Layers of Social Media Analytics: Mining Business Insights from Social Media Text, Actions, Networks, Hyperlinks, Apps, Search Engine, and Location Data	Gohar F. Khan	CreateSpace Independent Publishing Platform		2015
2	Analyzing the Social Web	Jennifer Golbeck	Morgan Kaufmann Publishers	1st	2013
	Mining the Social Web	Matthew A. Russell,	O′Reilly	3rd	2019

3		Mikhail Klassen			
4	Social Media Data Mining and Analytics	Gabor Szabo, Gungor Polatkan, P. Oscar Boykin, Antonios Chalkiopoulos	Wiley		2018

SIUDSELP311: Social Media Analytics Practical

B.Sc. (Data Science)	**Semester – V**
Course Name: Social Media Analytics Practical	**Course Code: SIUDSELP311**
Credits	**1**

List of Practical:

1	Study various - Social Media platforms (Facebook, Twitter, YouTube, etc) Social Media analytics tools (Facebook insights, google analytics netlytic, etc) Social Media Analytics techniques and engagement metrics (page level, post level, member level) Applications of Social media analytics for business. e.g. Google Analytics https://marketingplatform.google.com/about/analytics/ https://netlytic.org/
2	Data Collection-Select the social media platforms of your choice (Twitter, Facebook, LinkedIn, YouTube, Web blogs, etc), connect to and capture social media data for business (scraping, crawling, parsing) using web Scraper tool
3	You are tasked with scraping data from a social media platform (e.g. any e commerce web site) using Python. Your objective is to extract the following details • Product name • product price • Description of the product • Link of the product • Product is in stock or not
4	Data Cleaning and Storage- Preprocess, filter and store social media data for business (Using Python, MongoDB, R, etc).
5	Exploratory Data Analysis and Visualization of Social Media Data for business.
6	Location analytics using power BI tool. 1.take dataset of transactions that include location information (city, region, or coordinates). 2. Create a map visualization (e.g., filled map or bubble map) in Power BI to display sales by region.

	3. Use the map to analyze which regions are generating the most sales and identify any patterns or trends.
7	Text analytics using python • In text analytics, analyze the different types of emotion present in the text and also count number of times particular emotion appearing in the text.
8	Develop a dashboard and reporting tool based on real time social media data.
10	Design the creative content for promotion of your business on social media platform.
11	Analyze competitor activities using social media data.

SIUDSMN311: Advanced Web Technology

B.Sc. (Data Science)	**Semester – V**
Course Name: Advanced Web Technology	**Course Code: SIUDSMN311**
Credits	**1**

Course Objective:

- To design valid, well formed, scalable and meaningful pages using emerging technologies.
- To develop and implement data driven websites.
- Articulate what React is and why it is useful
- Creating dynamic web application
- Articulate what flask is and why it is useful

Course Outcomes:

- **CO1:** Students will be able to design single page application
- **CO2:** Students will be able to create design web page using bootstrap
- **CO3:** Students will be able to create database driven application
- **CO4:** Students will be able to create application using flask framework

Unit	Contents	No. of Lectures
I	**React**: Introduction,What is React ,What is single page application(SPA),React Installation,React ES6, React JSX, First React Code,Creating a React Project,Setting Up a Code Editor How React Works & Understanding Components,React Class,More About Components & Styling with CSS Classes ,Handling Events ,Introducing State,Event Props,Stateless and Stateful Components,Adding Routing,Adding Links & Navigation,CSS Modules, Dom Management with React,Adding a Form ,Getting User Input & Handling Form Submission,Preparing the App for HTTP,Sending a Post Request, Navigating Programmatically,Getting Started with Fetching Data ,React Hooks: useEffect ,Introducing React Context,Context Logic **MongoDB:** Introduction to mongoDB, create database, Create a Form in React to Add New Data to MongoDB, Display Data in React and Update It Using MongoDB. **PHP:** Why PHP and MySQL? Server-side scripting, PHP syntax and variables, comments, types, control structures, branching, looping, termination, functions, passing information with PHP, GET, POST,	15

	formatting form variables, superglobal arrays, strings and string functions, arrays, number handling, basic PHP errors/problems Advanced PHP **MySQL:** PHP/MySQL Functions, Integrating web forms and databases, Displaying queries in tables, Building Forms from queries, Regular Expressions, Sessions, Cookies and HTTP, E-Mail **Flask:** Installation, Basic application structure, routing, variables, redirect and errors, Templates, cookies, session, webforms, Databases	

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	React in Action	Mark Tielens Thomas	Manning Publications	5th	2018
2	PHP 5.1 For Beginners	Shroff/X-Team	Cengage Learning	1st	2010
3	PHP6 And MySQL Bible	Joyce Park Steve Suehring, Tim Converse	Wiley India Private Limited		2009
4	Flask Web Development	Miguel Grinberg	O'Reilly Media	1st	2014
5	Bootstrap: Responsive Web Development	Jake Spurlock	O'Reilly	1st	2013

SIUDSMNP311: Advanced Web Technology Practical

B.Sc. (Data Science)	**Semester – V**
Course Name: Advanced Web Technology Practical	**Course Code: SIUDSMNP311**
Credits	**1**

List of Practical:

1.	**React:**
a.	Creating an application using react. (Component,State and Props)
b.	Demonstrating React JSX, React Router
c.	Demonstrate Form handling using React-Login form, Registration form,Working with Event Listeners.
2.	**MongoDB:**
a.	Build a React form that allows the user to input a name and email address.
b.	When the form is submitted, send the data to an Express API to store it in a MongoDB collection.
c.	After the data is successfully added, display a success message in the React app
3	**Basic PHP**
a	Write a PHP Program to accept a number from the user and print it factorial.
b	Write a PHP program to accept a number from the user and print whether it is prime or not.
c	Write a PHP program to demonstrate different string functions.
d	Write a PHP program to create one dimensional array
4	**PHP and Database**
a.	Write a PHP code to create: Create a database College Create a table Department (Dname, Dno, Number_Of_faculty)
b.	Write a PHP program to create a database named “College”. Create a table named “Student” with following fields (sno, sname, percentage). Insert 3 records of your choice. Display the names of the students whose percentage is between 35 to 75 in a tabular format
c.	Design a PHP page for authenticating a user.
5	**implement the following using flask:**
a.	create Flask Application
b.	Show the use of cookies and sessions

c	Connect Flask to a Database with Flask-SQLAlchemy

SIUDSVS311: Data Analysis & Visualization with Microsoft Power BI

B.Sc. (Data Science)	**Semester – V**
Course Name: Data Analysis & Visualization with Microsoft Power BI	**Course Code: SIUDSVS311**
Credits	**1**

Course Objectives:

- To understand data and generate insights from it is by visualising it using a range of data visualization tools available.
- To understand large volume of data, discover trends, communicate effectively with all stakeholders and influence decisions.
- To develop Business Analytics skillset about how to create effective charts and interactive dashboards is extremely useful.

Course Outcomes:

By The End Of This Course, Student Will Be Able To:

- **CO1:** Students will be able to understand and describe the main concepts of data visualization.
- **CO2:** Students will be able to determine the right type of graph for different types of data available or provided through hands on experience with handling real data sets
- **CO3:** students will be able to read reports, charts, graphs, figures, maps and derive meaning from them.
- **CO4:** Students will be able to create reports, data visualizations, and dashboards using Power BI.
- **CO5:** Students will be able to understand how to automate tasks, perform ETL, create data models, perform computations, and present insights using data visualization and dashboards.

Unit	Contents	No. of Lectures
I	**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.	15

	Data Preparation: Connecting to different data sets, Basic data prep and model on Power Query, Drill down and Tooltip, AI visuals (Q&A, Analyze, Decomposition) **Data visualization and dashboards:** Inbuilt visuals, Custom visuals, Learn from existing reports, Visualization as a Tooltip, Final dashboard- putting it together Filter, slicer, bookmarks, buttons, **Performing computations**: Combine multiple files and folders, Merge and append, Custom calculations Conditional columns, Column from examples, Advanced Editor. **DAX**: Introduction to Measures, Calculated Columns vs Quick Measures, Creating a Date Table, Time Intelligence Functions **Power bi service**: Publishing to Power BI Service, Power BI dataflows, Dashboards and Cross-Reporting **Visualizations:** Introduction to Dimensions and Measures, Bar Chart, Line Chart, Table, Heat Map, Treemap, Packed Bubble, Tooltip, **Calculations:** Calculated Fields, Parameters, Introduction to Level of Detail (LOD) **Final dashboard:** Animations, Tooltips, Dashboard and Stories	

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	Fundamentals of Data Visualization	Claus O. Wilke	O'Reilly Media	1st	2019
2.	Microsoft Power BI Data Analyst Certification Guide	Orrin Edenfield, Edward Corcoran	O' Reilly		June 2022
3.	Mastering Power BI	Chandraish Sinha	BPB Publications		30 September 2021
4.	Interactive Data Visualization: Foundations, Techniques, and Applications.	Ward, Grinstein Keim	A K Peters/CRC Press	Second	2015
5.	The Visual Display of Quantitative Information	E. Tufte	Graphics Press	Second	2001

SIUDSVS311: Data Analysis & Visualization with Microsoft Power BI

B.Sc. (Data Science)	**Semester – V**
Course Name: Data Analysis & Visualization with Microsoft Power BI	**Course Code: SIUDSVS311**
Credits	**1**

List of Practical:

1.	Installing and configuring Power BI Desktop.
2.	1. Import data: 2. Data Cleaning i. Check for missing or incorrect data. ii. Correct any anomalies (e.g., negative values in sales revenue or quantity sold).
3.	1. Create a Time-Based Analysis: i. Create a line chart showing **total sales revenue** over the last 12 months. ii. Include month-over-month growth in sales revenue in the chart. **2. Product Performance:** i. Create a bar chart that shows **total sales revenue** by **product name**. ii. Highlight the top 5 performing products.
4.	**1. Regional Sales Insights:** i. Create a map visualization that shows **total sales revenue by region**. ii. Identify the region with the highest sales and provide an analysis. **2. Quantity vs Revenue:** i. Create a scatter plot to show the relationship between **quantity sold** and **sales revenue**. ii. Identify any outliers where a product sold a large quantity but generated low revenue, or vice versa
5.	Create a dashboard that includes the following visualizations: • Total Sales Revenue (Card) • Sales Revenue over Time (Line Chart)

Course Code	Course Type	Course Title	Credits
SIUDSMJ321	Core Subject	Cloud Computing	**3**
SIUDSMJ322	Core Subject	Natural Language Processing	**3**
SIUDSMJ323	Core Subject	Data Warehouse and Mining	**3**
SIUDSEL321	Core Subject (Elective)	Applied Hypothesis Testing in Data Science	**3**
SIUDSMN321	Minor	Optimization Technique	**1**
	OJT	On Job Training	**4**
SIUDSMJP321	Core Subject Practical	Cloud Computing Practical	**1**
SIUDSMJP322	Core Subject Practical	Natural Language Processing Practical	**1**
SIUDSMJP323	Core Subject Practical	Data Warehouse and Mining Practical	**1**
SIUDSELP321	Core Subject Practical	Applied Hypothesis Testing in Data Science Practical	**1**
SIUDSMNP321	Minor	Optimization Technique	**1**
		TOTAL CREDITS	**22**

	• Product Sales (Bar Chart) • Regional Sales (Map) Ensure the dashboard is easy to read and highlights key insights for decision-makers.
6.	Marketing Campaign Dashboard.
7.	Product Availability Dashboard.
8.	Flight Price Analysis Dashboard.
9.	Demonstrate the process of publishing Power BI reports.

T.Y. B.Sc. (DS) – Semester VI

SIUDSMJ321: Cloud Computing

B.Sc. (Data Science)	**Semester – VI**
Course Name: Cloud Computing	**Course Code: SIUDSMJ321**
Credits	**3**

Course Objective:

- This course gives students an insight into the basics of cloud computing along with virtualization; cloud computing has been one of the fastest-growing domains for a while now.
- It will provide the students with a basic understanding of clouds and virtualization along with how one can migrate over it.
- Cloud computing software security objectives, design principles and development practices.
- Cloud computing risks, challenges and threats to infrastructure, data and access control.
- Cloud computing security architectural issues, Identity management and Autonomic security.

Course Outcomes (COs):

By the completion of the course, the students are able to:

CO1: Articulate the main concepts, key technologies, strengths, limitations of cloud computing and the possible applications for state-of-the-art cloud computing.

CO2: Identify the architecture and infrastructure of cloud computing, including cloud delivery and deployment models.

CO3: Analyze the core issues of cloud computing such as security, privacy, and interoperability.
CO4: Identify problems, analyze, and evaluate various cloud computing solutions.
CO5: Analyze appropriate cloud computing solutions and recommendations according to the applications used.

Unit	Contents	No. of Lectures
I	**IKS: Indian Cloud engineers**, **Homegrown Cloud Startups & Products**(Global SaaS players (Zoho, Freshworks, Postman), **Government & Policy-Driven Innovation** (Digital India, MeghRaj, UPI, Aadhaar cloud infra**), Cloud Infrastructure Investments (**Data Centers in India**), Talent Pool & Skill Developmen**t, **Research** (Cloud-native computing, edge, AI, security**), Academia & Open Source**. **Overview of Cloud Computing:** Introduction to cloud computing, Characteristics of cloud computing, Advantages of cloud computing, Disadvantages of cloud computing, Cloud service models, Cloud computing deployment models, Cloud computing deployment models. Comparing cloud providers with traditional IT service providers, Roots of cloud computing. **Cloud Architecture and Applications:** Cloud architecture, Components of cloud computing architecture, Working of cloud computing, Applications of cloud computing. **Cloud Services:** Cloud services, Benefits of cloud services, Types of cloud service models-Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS), features of PaaS and benefits, Infrastructure as a Service (IaaS), features of IaaS and benefits, Service providers, challenges and risks in cloud adoption.	15
II	**Cloud Simulators**- Cloud Sim and Green Cloud Introduction to Simulator, understanding Cloud Sim simulator, Cloud Sim Architecture (User code, Cloud Sim, Grid Sim, Sim Java) Understanding Working platform for Cloud Sim. **Introduction to VMWare Simulator:** Basics of VMWare, advantages of VMware virtualization, using VMware workstation, creating virtual machines-understanding virtual machines, create a new virtual machine on local host, cloning virtual machines, virtualize a physical machine, starting and stopping a virtual machine.	15
III	**Cloud Computing on AWS:** Delve into AWS' score services Compute, Load Balancing, Autoscaling, Storage, and more. **Managed Services on AWS:** Explore AWS's managed services like Databases (RDS, DynamoDB), SNS, SQS, Cloud watch, and more.	15

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	Cloud Security A Comprehensive Guide to secure Cloud Computing	Ronald L. Krutz, Russell Dean Vines	Wiley	1st	
2.	Cloud computing a practical approach	Anthony T.Velte , Toby J. Velte Robert	Elsenpeter, TATA McGraw-Hill		2010
3.	Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online	Michael Miller	Que		2008
4.	Cloud Computing Implementation, Management and Security	John W. itinghouse james F.Ransome	CRC Press		

SIUDSMJP321: Cloud Computing Practical

B.Sc. (Data Science)	**Semester – VI**
Course Name: Cloud Computing	**Course Code: SIUDSMJP321**
Credits	**1**

List of Practical:

1.	To Study and implementation of Infrastructure as a Service. Scenario: Tech Nova, an AI-driven analytics startup, is struggling with its on-premises infrastructure. Their current servers cannot handle increasing customer demand, leading to performance issues, high maintenance costs, and limited scalability. To overcome these challenges, Tech Nova decides to migrate to Infrastructure as a Service (IaaS).
2.	Installation and Configuration of virtualization using KVM: To study and implement Kernel-based Virtual Machine (KVM) for virtualization, enabling multiple operating systems to run simultaneously on a Linux-based host system. A. install and configure KVM on a Linux host. B. Create and manage virtual machines using CLI and GUI. C. Set up networking, storage, snapshots, and backups. D. Optimize VM performance, security, and monitoring.

3.	To Study and implementation of Storage as a Service: To explore how cloud storage services (Google Drive, OneDrive, or Dropbox) enable users to store, access, and share Word documents, Excel sheets, PowerPoint presentations, and Forms efficiently. A. Google Drive/OneDrive Setup: B. Create a cloud storage account (Google Drive, OneDrive, Dropbox). C. Organize folders for different file types (Word, Excel, PPT, Forms).
4.	To Study and implementation of identity management: To understand and implement Identity Management (IDM), which involves securely managing user identities, access control, authentication, and authorization in an IT environment. A. created an AWS Free Tier account. B. Configured IAM users, groups, and roles. C. Implemented RBAC and Multi-Factor Authentication (MFA). D. Monitored IAM activity with CloudTrail.
5.	To Study Cloud Security Management Using AWS Free Trial: To understand and implement Cloud Security Management using AWS, focusing on Identity and Access Management (IAM), Multi-Factor Authentication (MFA), and Security Best Practices. A. secured AWS root account with MFA. B. Created IAM users with proper permissions. C. Implemented cloud security best practices for AWS management.
6.	Write a program for web feed: Set up an RSS Feed Reader in Chrome, allowing you to stay informed about your favorite topics without visiting multiple websites manually. Steps to Use an RSS Feed Reader in Chrome: Scenario: Imagine you are a digital marketing specialist who wants to stay updated with the latest industry trends, news, and blog updates without visiting multiple websites. To streamline this process, you decide to use an RSS Feed Reader in Chrome to collect and display the latest updates in one place. A. Install the RSS Feed Reader Extension B. Enable the RSS Feed Reader Icon C. Add a Website to Your RSS Feed Reader D. Follow the RSS Feed E. Access Your RSS Feeds
7.	To Study and implementation of Single-Sing-On: To understand and implement Single Sign-On (SSO) using Auth0, allowing users to log in once and access multiple applications seamlessly. Scenario: You are a developer working on a web application, and your team wants to integrate Single Sign-On (SSO) to allow users to log in with their Google account instead of creating multiple accounts for different services. You decide to use Auth0, a popular identity provider, to implement this feature. A. Search for an SSO Implementation Guide. B. Sign Up for an Auth0 Account C. Create a Social Connection for Google Sign-In D. Test the Google SSO Connection E. Integrate SSO with Your Application
8.	To Study and Implement User Management in Cloud (AWS IAM):

	To understand and implement User Management in Cloud by adding and removing users in AWS Identity and Access Management (IAM). A. Enable Multi-Factor Authentication (MFA) for added security. B. Follow the Principle of Least Privilege (grant only necessary permissions). C. Regularly audit and remove inactive users to improve security. D. Use IAM Roles instead of individual users for temporary access. Scenario: You are a Cloud Administrator managing user access for your organization in AWS Cloud. You need to create new IAM users and remove inactive users to maintain security and access control. A. Log in to AWS as a Root User B. Add a New IAM User C. Remove a User from AWS IAM
9.	Case study on Amazon EC2/Microsoft Azure/Google Cloud Platform: Scenario: You are a Cloud Engineer working for a company that wants to deploy its web application on AWS EC2. Your task is to launch an on-demand EC2 instance in AWS Cloud by selecting the appropriate configurations for the business needs. A. Accessing AWS Services B. Choosing an AWS Region C. Selecting an Amazon Machine Image (AMI) D. Choosing an EC2 Instance Type E. Configuring the EC2 Instance F. Network and Subnet Configuration G. Assigning a Public IP H. Cost Optimization for On-Demand Instances

SIUDSMJ322: Natural Language Processing

B.Sc. (Data Science)	**Semester – VI**
Course Name: Natural Language Processing	**Course Code: SIUDSMJ322**
Credits	**3**

Course Objective:

- Learn the importance of natural language modelling
- Understand the Applications of natural language processing
- Study spelling, error detection and correction methods and parsing techniques in NLP
- Illustrate the information retrieval models in natural language processing

Course Outcomes:

After completion of this course, student will be able to:

- **CO1:** Apply the fundamental concept of NLP, grammar-based language model and statistical-based language model.
- **CO2:** Model morphological analysis using Finite State Transducers and parsing using context-free grammar and different parsing approaches.
- **CO3:** Develop the Naïve Bayes classifier and sentiment analysis for Natural language problems and text classifications.

- **CO4:** Apply the concepts of information retrieval, lexical semantics, lexical dictionaries such as WordNet, lexical computational semantics, distributional word similarity.
- **CO5:** Identify the Machine Translation applications of NLP using Encode and Decoder.

Unit	Contents	No. of Lectures
I	**Introduction:** What is Natural Language Processing? Origins of NLP, Language and Knowledge, The Challenges of NLP, Language and Grammar, Processing Indian Languages, NLP Applications. **Language Modeling:** Statistical Language Model - N-gram model (unigram, bigram), Paninion Framework, Karaka theory. **Word Level Analysis:** Regular Expressions, Finite-State Automata, Morphological Parsing, Spelling Error Detection and Correction, Words and Word Classes, Part-of Speech Tagging. **Syntactic Analysis:** Context-Free Grammar, Constituency, Top-down and Bottom-up Parsing, CYK Parsing.	15
II	**Naive Bayes, Text Classification and Sentiment:** Naive Bayes Classifiers, Training the Naive Bayes Classifier, Worked Example, Optimizing for Sentiment Analysis, Naive Bayes for Other Text Classification Tasks, Naive Bayes as a Language Model **Information Retrieval:** Design Features of Information Retrieval Systems, Information Retrieval Models - Classical, Non-classical, Alternative Models of Information Retrieval - Custer model, Fuzzy model, LSTM model, Major Issues in Information Retrieval.	15
III	**Lexical Resources:** WordNet, FrameNet, Stemmers, Parts-of-Speech Tagger, Research Corpora. **Machine Translation:** Language Divergences and Typology, Machine Translation using Encoder Decoder, Details of the Encoder-Decoder Model, Translating in Low-Resource Situations, MT Evaluation, Bias and Ethical Issues.	15

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	"Natural Language Processing and Information Retrieval", Oxford University Press.	Tanveer Siddiqui, U.S. Tiwary			
2	"Speech and Language Processing, An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition", Pearson Education, 2023.	Daniel Jurafsky, James H. Martin			
3	"Natural Language ProcessingRecipes - Unlocking Text Data with Machine Learning and Deep	Akshay Kulkarni, Adarsha Shivananda,			2019

	Learning using Python", Apress,				

SIUDSMJP322: Natural Language Processing Practical

B.Sc. (Data Science)	**Semester – VI**
Course Name: Natural Language Processing Practical	**Course Code: SIUDSMJP322**
Credits	**1**

List of Practical: (using python)

1.	Write a Python program for the following preprocessing of text in NLP: ● Tokenization ● Filtration ● Script Validation ● Stop Word Removal ● Stemming
2.	Demonstrate the N-gram modeling to analyze and establish the probability distribution across sentences and explore the utilization of unigrams, bigrams, and trigrams in

	diverse English sentences to illustrate the impact of varying n-gram orders on the calculated probabilities.
3.	Investigate the Minimum Edit Distance (MED) algorithm and its application in string comparison and the goal is to understand how the algorithm efficiently computes the minimum number of edit operations required to transform one string into another. ● Test the algorithm on strings with different type of variations (e.g., typos, substitutions, insertions, deletions) ● Evaluate its adaptability to different types of input variations
4.	Write a program to implement top-down and bottom-up parser using appropriate context free grammar.
5.	Given the following short movie reviews, each labeled with a genre, either comedy or action: ● fun, couple, love, love comedy ● fast, furious, shoot action ●couple, fly, fast, fun, fun comedy ● furious, shoot, shoot, fun action ● fly, fast, shoot, love action and A new document D: fast, couple, shoot, fly Compute the most likely class for D. Assume a Naive Bayes classifier and use add-1 smoothing for the likelihoods.
6.	Demonstrate the following using appropriate programming tool which illustrates the use of information retrieval in NLP: ● Study the various Corpus – Brown, Inaugural, Reuters, udhr with various methods like filelds, raw, words, sents, categories • Create and use your own corpora (plaintext, categorical) • Study Conditional frequency distributions • Study of tagged corpora with methods like tagged_sents, tagged_words • Write a program to find the most frequent noun tags • Map Words to Properties Using Python Dictionaries • Study Rule based tagger, Unigram Tagger Find different words from a given plain text without any space by comparing this text with a given corpus of words. Also find the score of words.
7.	Write a Python program to find synonyms and antonyms of the word "active" using WordNet.
8.	Implement the machine translation application of NLP where it needs to train a machine translation model for a language with limited parallel corpora. Investigate and incorporate techniques to improve performance in low-resource scenarios.

SIUDSMJ323: Data Warehouse and Mining

B.Sc. (Data Science)	Semester – VI
Course Name: Data Warehouse and Mining	**Course Code: SIUDSMJ323**
Credits	**3**

Course Objective:

- Understand business intelligence for an enterprise and review data warehouse with architectural types and architectural building blocks
- To discuss and understand changing dimensions and learn about aggregate tables and determine their usage.
- To learn basics of data mining, understand the need and the process of data mining in contrast with machine learning.
- To study the use of classification and clustering techniques for Data Mining.
- To appreciate the use of various data mining algorithms and learn about their specific applications.

Course Outcomes:

Upon completion of this course, student will be able to:

CO1: Learner is able to demonstrate knowledge of business intelligence, data warehouse with clear understanding of architectural types and will be able to establish the relationship between architectural building blocks.
CO2: Learner is able to elaborate changing dimensions with respect to current trends & using aggregate tables.
CO3: Learner is able to handle the processes of data preprocessing, data transformation and data reduction.
CO4: Learner has knowledge of using various Data Mining techniques for classification and clustering.
CO5: Learner is able to align the Data Mining techniques for analyzing the datasets using tools like Weka, R or Python.

Unit	Contents	No. of Lectures
I	**THE COMPELLING NEED FOR DATA WAREHOUSING:** Escalating Need for Strategic Information, Failures of Past Decision-Support Systems, Operational Versus Decision-Support Systems, Data Warehousing—The Only Viable Solution, Data Warehouse Defined, The Data Warehousing Movement, Evolution of Business Intelligence **DATA WAREHOUSE:** The Building Blocks: Defining Features, Data Warehouses and Data Marts, Architectural Types, Overview of The Components, Metadata in The Data Warehouse **TRENDS IN DATA WAREHOUSING:** Continued Growth in Data Warehousing, Significant Trends, Emergence of Standards, Web-Enabled Data Warehouse **ARCHITECTURAL COMPONENTS:** Understanding Data Warehouse Architecture, Distinguishing Characteristics, Architectural Framework, Technical Architecture, Architectural Types **THE SIGNIFICANT ROLE OF METADATA:** Why Metadata Is Important, Metadata Types By Functional Areas, Business Metadata, Technical Metadata, How To Provide Metadata **PRINCIPLES OF DIMENSIONAL MODELING:** From Requirements to Data Design, The Star Schema, Star Schema Keys, Advantages of The Star Schema, Star Schema: Examples **DIMENSIONAL MODELING: ADVANCED TOPICS:** Updates to The Dimension Tables, Miscellaneous Dimensions, The Snowflake Schema, Aggregate Fact Tables, Families of Stars **DATA EXTRACTION, TRANSFORMATION, AND LOADING:** ETL Overview, ETL Requirements and Steps, Data Extraction, Data Transformation, Data Loading, ETL Summary, Other Integration Approaches	15
II	**INTRODUCTION TO DATA MINING:** Introduction to Data Mining, Need of Data Mining, What Can Data Mining Do and Not Do? Data Mining Applications, Data Mining Process, Data Mining Techniques, Difference between Data Mining and Machine Learning **CLASSIFICATION:** Introduction to Classification, Types of Classification, Input and Output Attributes, Guidelines for Size and Quality of the Training Dataset, Introduction to the Decision Tree Classifier, Naive Bayes Method, Understanding Metrics to Assess the Quality of Classifiers	15

	CLUSTER ANALYSIS: Introduction to Cluster Analysis, Applications of Cluster Analysis, Desired Features of Clustering, Distance Metrics, Major Clustering Methods/Algorithms, Partitioning Clustering,	
III	**HIERARCHICAL CLUSTERING ALGORITHMS:** Web Mining and Search Engines: Introduction, Web Content Mining, Web Usage Mining, Web Structure Mining, Hyperlink Induced Topic Search algorithm, Introduction to Modern Search Engines, Working of a Search Engine, PageRank Algorithm, Precision and Recall **INTRODUCTION TO ASSOCIATION RULE MINING:** Defining Association Rule Mining, Representations of Items for Association Mining, The Metrics to Evaluate the Strength of Association Rules, The Naive Algorithm for Finding Association Rules, Approaches for Transaction Database Storage **THE APRIORI ALGORITHM,** Closed and Maximal Item sets, The Apriori– TID Algorithm for Generating Association Mining Rules, Direct Hashing and Pruning (DHP), Dynamic Itemset Counting (DIC), Mining Frequent Patterns without Candidate Generation (FP Growth)	15

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	DATA WAREHOUSING FUNDAMENTALS FOR IT PROFESSIONALS	PAULRAJ PONNIAH	Wiley	Second	2010
2	Data Mining and Data Warehousing : Principles and Practical Techniques	Parteek Bhatia	Cambridge University Press	First	2019
3	Data Mining and Data Warehousing	S.K. Mourya Shalu Gupta	Alpha Science International Ltd	First	2013

SIUDSMJP323: Data Warehouse and Mining Practical

B.Sc. (Data Science)	**Semester – VI**
Course Name: Data Warehouse and Mining Practical	**Course Code: SIUDSMJP323**
Credits	**1**

Practical List

	1.	**Data warehouse design**
	a.	Design dimension tables.
	b.	Design fact tables.
	c.	Create an indexed view and rebuild column store indexes.
	2.	**OLAP operations**

	a.	Perform various OLAP operations such slice, dice, roll up, drill down and pivot
	2.	**Data Warehouse with Azure**
	a.	Create an Azure SQL Data Warehouse Project.
	b.	Develop tables in Azure SQL Data Warehouse.
	c.	Migrate Data Warehouse to Azure.
	d.	Pause and remove Azure data warehouse.
	3.	**Data Warehouse implementation and use**
	a.	Cleanse data with SQL Server Data Quality Services.
	b.	Create custom knowledge base.
	c.	Install Master Data Services and IIS
	d.	Configure MDS and deploy sample MDS model.
	e.	Install MDS excel add-in and Update master data in excel.
	f.	Consume the data from the warehouse.
	4	**Working with Data and Data Preprocessing**
	a.	Demonstrate the use of ARFF files taking input and display the output of the files.
	b.	Create your own excel file. Convert the excel file to .csv format and prepare it as ARFF files.
	c.	Preprocess and classify Customer dataset.http://archive.ics.uci.edu/ml/
	d.	Perform Preprocessing, Classification techniques onAgriculture dataset. (http://archive.ics.uci.edu/ml/)
	e.	Preprocess and classify Weather dataset.http://archive.ics.uci.edu/ml/
	f.	Perform data Cleansing of customer dataset. http://archive.ics.uci.edu/ml/ www.kdnuggets.com/datasets/
	5	**Performing classification on data sets**
	a.	Building a Decision Tree Classifier in Weka
	b.	Applying Naïve Bayes on Dataset for classification
	c.	Creating the Testing Dataset
	d.	Decision Tree Operation with R
	e.	Naïve Bayes Operation using R
	f.	Classify the dataset using decision tree. www.kdnuggets.com/datasets/
	6	**Simple Clustering**
	a.	Perform Clustering technique on Customer dataset. http://archive.ics.uci.edu/ml/
	b.	Perform Clustering technique on Agriculture dataset. http://archive.ics.uci.edu/ml/
	c.	Perform Clustering technique on Weather dataset. http://archive.ics.uci.edu/ml/
	7	**Implementing Clustering with Weka and R**
	a.	Clustering Fisher‘s Iris Dataset with the Simple k-Means Algorithm
	b.	Handling Missing Values
	c.	Results Analysis after Applying Clustering
	d.	Classification of Unlabeled Data
	e.	Clustering in R using Simple k-Means
	8	**Implementing Apriori Algorithm with Weka and R**
	a.	Applying Predictive Apriori in Weka

	b.	Applying the Apriori Algorithm in Weka on a Real World Dataset
	c.	Applying the Apriori Algorithm in Weka on a Real World Larger Dataset
	d.	Applying the Apriori Algorithm on a Numeric Dataset
	9	**Implementing Association Mining with R**
	a.	Applying Association Mining in R
	b.	Application of Association Mining on Numeric Data in R
	c.	Perform Association technique on Agriculture dataset. http://archive.ics.uci.edu/ml/,www.kdnuggets.com/datasets/
	d.	Perform Association technique on Agriculture dataset. http://archive.ics.uci.edu/ml/, www.kdnuggets.com/datasets/
	e.	Perform Association technique on Weather dataset.
	10	**Web Mining**
	a.	Implement Hyperlink Induced Topic Search (HITS) algorithm
	b.	Implement PageRank Algorithm

SIUDSEL321: Applied Hypothesis Testing in Data Science

B.Sc. (Data Science)	**Semester – VI**
Course Name: Applied Hypothesis Testing in Data Science	**Course Code: SIUDSEL321**
Credits	**3**

Course Objective:

- To impart statistical significance in solving complex problems.
- To critically test in developing robust, extensible and highly maintainable solutions to simple and complex problems.
- To implement various statistical functions using suitable programming languages and packages.
- To scientifically test the unknown and unlock possibilities in different dimensions of the system.
- To write the reports of analytical results generated by the system.

Course Outcomes:

Upon completion of this course, student will be able to:

CO1: Learner is developing null and alternative hypotheses to test for a given situation.
CO2: Learner is able to differentiate one- and two-tailed hypothesis tests.
CO3: Learner is able to do sampling a normal distribution and random sampling.
CO4: Learner is using statistical models and their associations in performing hypothesis testing.
CO5: Lerner is writing the reports and interpreting the data using the various programming languages and packages.

Unit	Contents	No. of Lectures
I	**Introduction to Hypothesis Testing:** Hypothesis Tests, stating a Hypothesis, Types of Errors and Level of Significance, Statistical Tests and P-Values, making a Decision and Interpreting the Decision, Strategies for Hypothesis Testing, Characteristics of a good hypothesis, Steps for hypothesis testing **Hypothesis Testing for the Mean (σ Known):** Using P-Values to Make Decisions, Using P-Values for a z-Test, Rejection Regions and Critical Values, Using Rejection Regions for a z-Test, Critical Values in at Distribution, The t-Test for a Mean μ, Using P-Values with t-Tests, Sums and case studies **Packages used for Hypothesis testing:** Introduction to statistical functions in R / Python / Excel, Packages used for finding P-value to make decision and hypothesis testing. **Goodness of fit tests:** Anderson-Darling, Chi-square test, Kolmogorov Smirnov, Ryan-Joiner, Shapiro-Wilk, Jarque-Bera, Lilliefors **Variance tests:** Chi-square test of a single variance, F-tests of two variances, Tests of homogeneity **Wilcoxon rank-sum/Mann-Whitney U test, Sign test Contingency tables:** Chi-square contingency table test, G contingency table test, Fisher's exact test, Measures of association, McNemar's test **Packages used for Hypothesis testing:** Packages used for finding goodness of fit test, variance test, Wilcoxon rank-sum / Mann-Whitney U test and sign test, Using Contingency table in R / Python / Excel.	15
II	**Analysis of variance and covariance:** ANOVA, Single factor or one way ANOVA, Two factor or two-way and higher-way ANOVA , MANOVA, ANCOVA **Non-Parametric ANOVA:** Kruskal-Wallis ANOVA, Friedman ANOVA test, Mood's Median **Packages used for Hypothesis testing:** Packages used for finding Anova, Manova, Anova and Non-Parametric Anova in R / Python / Excel. **Regression and smoothing:** Ridge regression, Generalized Linear Models (GLIM), Poisson regression for count data, Non-linear regression, Smoothing and Generalized Additive Models (GAM), Geographically weighted regression (GWR), Spatial series and spatial autoregression- SAR models, CAR models, Spatial filtering models	15
III	**Communicating and Documenting the Results of Analyses:** Introduction, The Difficulty of Good Communication, Communication Hurdles: Graphical Distortions, Communication Hurdles: Biased Samples & Sample Size, Preparing Data for Statistical Analysis, Guidelines for a Statistical Analysis and Report, Documentation and Storage of Results ,Supplementary Exercise **Data Storytelling:** What is a Data Story?, The Art and Science of Storytelling, Planning the Data Story, Elements of the Data Story, Parts of the Data Story, Framing and Formatting of the Data Story, False Narratives and Data Storytelling **Infographics:** What is an Infographic?, Why are Infographics Useful? Types of Infographics, Infographic Design Elements, Steps in Designing an Infographic, Best Practices in Designing an Infographic	15

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Hypothesis Testing	---	Pearson Higher Education	---	---
2	Statistical Analysis Handbook	Dr. Michael J de Smith	The Winchelsea Press, Drumlin Security Ltd, Edinburgh	---	2018
3	An Introduction to Statistical Methods and Data Analysis	R. Lyman Ott & Michael Longnecker	Thomson Learning	---	---

SIUDSELP321: Applied Hypothesis Testing in Data Science Practical

B.Sc. (Data Science)	**Semester – VI**
Course Name:	**Course Code: SIUDSELP321**
Credits	**1**

List of Practical : (Practical can be performed using R / Python / scilab / matlab / SPSS / MS Excel)

1	**Hypothesis Testing for the Mean**
a.	Perform testing of hypothesis using one sample t-test.
b.	Perform testing of hypothesis using two sample t-test.
c.	Perform testing of hypothesis using paired t-test.
d.	Perform testing of hypothesis using Z-test.
2	**Goodness-of-fit test**
a.	Perform goodness-of-fit test using chi-squared test.
b.	Perform goodness-of-fit test using KS-test.
c.	Perform testing of hypothesis using chi-squared Test of Independence
3	**Variance Testing**
a.	Using Chi-square test of a single variance
b.	Using F-tests of two variances
c.	Testing of homogeneity
4	**Analysis of variance and covariance**
a.	Perform testing of hypothesis using one-way ANOVA.
b.	Perform testing of hypothesis using two-way ANOVA.
c.	Perform testing of hypothesis using Multivariate ANOVA (MANOVA)
d.	Perform testing of hypothesis using one-way ANOVA.
5	**Regression**

a.	Perform simple linear regression
b.	Perform multiple linear regression
c.	Perform polynomial regression
6	Perform spatial series and spatial auto-regression
7	Perform time series analysis using Moving averages
8	Perform time series analysis using Trend Analysis
9	Perform Spectral analysis
10	Creating Infographics‖ using secondary data available on internet. (Use Canva / Adobe Spark / Prezi / Vennage

SIUDSMN321: Optimization Techniques

B.Sc. (Data Science)	**Semester – VI**
Course Name: Optimization Techniques	**Course Code: SIUDSMN321**
Credits	**1**

Course Objective:

- Introduce optimization methodology as a valuable decision support tool.
- Help develop skills in building and solving optimization models for variety of engineering and Data Science related decision problems.
- Expose key mathematical concepts underlying various optimization models and algorithms.
- Provide hands-on experience with optimization software for solving and analyzing optimization models.

Course Outcomes:

Upon completion of this course, student will be able to:

CO1: Formulate deterministic mathematical programs in various practical systems
CO2: Understand basic optimization techniques
CO3: Be able to interpret the results of a model and present the insights (sensitivity, duality)
CO4: Know the limitations of different solution methodology
CO5: Use software to solve problems

Unit	Contents	No. of Lectures
I	**Linear Programming Problem (LPP):** Mathematical Model, Standard Form, Canonical Form, Simplex Method, Big-M Simplex Method, Two Phase Simplex Method, Primal Vs Dual, Formulation of Dual from Primal, Comparison of Solutions of Primal & Dual of LPP **Assignment Problem (AP) :** Hungarian Method/ Reduced Matrix of Assignment problem. **Decision Making under Conditions of Certainty,Uncertainty & Risk :** Maximax, MaxiMin, MiniMax, Hurwicz, Laplace Criteria &	15

	Methods, EMV, EOL & EVPI Calculations, Incremental/ Marginal Analysis Method. **Simulation Techniques :** Monte Carlo Method, Problems of Simulation.	

Books and References

Sr. No.	Title	Author/s	Publisher	Edition	Year
1	Operations Research,	Prem Kumar Gupta & D S Hira	S Chand	---	---
2	Introduction to Mathematical Programming	Wayne L. Winston and Munirpallam Venkataramanan	---	4th Edition).	---

SIUDSMNP321: Optimization Techniques Practical

B.Sc. (Data Science)	**Semester – VI**
Course Name: Optimization Techniques	**Course Code: SIUDSMNP321**
Credits	**1**

List of Practical:

1	Simplex Method using R/ python/ scilab/ matlab.
2	North West Corner Method using R/ python/ scilab/ matlab.
3	Least Cost Method using R/ python/ scilab/ matlab.
4	Vogel's Approximation Method using R/ python/ scilab/ matlab.
5	Hungarian Method using R/ python/ scilab/ matlab.
	Decision Making under conditions of certainty / uncertainty: i. Maximax ii. Maximin iii. Minimax
7	Monte Carlo Model Simulation using R/ python/ scilab/ matlab.
8	Decision Making under Conditions of Conflict - Game Theory **[Only Demonstration]**