Board of Studies, M.Sc. (Data Science)
Curriculum Structure for M.Sc. Program 2024-2025
Savitribai Phule Pune University
(Formerly University of Pune)

Department of Technology

Note:

A) M.Sc. Data Science (Intake 60)

Eligibility: B.Sc./BCA (Data Science/Computer Science/IT/Statics/Mathematics/Physics/Chemistry/Electronics or allied Branches or any graduate with 12th Maths

B) M.Sc. Data Science (Direct 2nd Year Admissions)

Eligibility: Advance Course in Data Science and AI/PGD in Data Science & AI/ Cyber Security.

Sr. No.	Subject Category	Subject Code	Subject Name	Credits
Semester (I)				
1	Major (Mandatory)	MSC1	Research Methodology	4
2	Major (Mandatory)	MSC2	Python Programming	4
3	Major (Mandatory)	MSC3	R Programming for Data Science	4
4	Major (Mandatory)	MSC4	Introduction to Database Management System	4
5	Major (Mandatory)	MSC5	Data Visualization	2
6	Major (Mandatory)	MSC6	Lab Practice -1	4
			TOTAL CREDITS	**22**
Semester (II)				
7	Major (Mandatory)	MSC7	Statistics Essential for Data Science	4
8	Major (Mandatory)	MSC8	Machine Learning & Artificial Intelligence	4
9	Major (Mandatory)	MSC9	Natural Language Processing	4
10	Major (Mandatory)	MSC10	Introduction to Deep Learning	4
11	OJT/ FP	MSC11	Massive Open Online Courses (MOOCs)	2
12	Major (Mandatory)	MSC12	Lab Practice – 2	4
			TOTAL CREDITS	**22**
Semester (III)				
13	Major (Mandatory)	MSC13	Big Data and Data Clouds	4
14	Major (Mandatory)	MSC14	Web Framework	4
15	Major (Mandatory)	MSC15	LLMs and Generative AI	4
16	Major (Mandatory)	MSC16	Seminar-I	2
17	RP	MSC17	Interim Project	4
18	Major (Mandatory)	MSC18	Lab Practice – 3	4
			TOTAL CREDITS	**22**
Semester (IV)				
19	Major (Mandatory)	MSC19	Data Mining and Business Intelligence	4
20	Major (Elective)	MSC20	Open Elective	2
21	Major (Mandatory)	MSC21	Seminar-II	2

22	RP	MSC22	Journal\Conference\Internship	4
23	RP	MSC23	Final Project (Dissertation Submission)	10
			TOTAL CREDITS	**22**
			GRAND TOTAL CREDITS	**88**

– First Year of M.SC Data Science is the same as the Advanced course in Data Science and AI and the remainder is as per UGC.
† - Open Elective can be taken from any Discipline / Department / MOOC with the approval of the Board Coordinator.

Eligibility: BSC/BCS and allied branches.

Course Intake: 60

Open Electives:

1. **Time Series Analysis and Forecasting Techniques**
2. **Marketing Analytics using AI**
3. **Data Analysis in Economics & Financial Decision Making**
4. **Data Analytics in IOT**

Notes:

1) Maximum 25% Open Electives are allowed.
2) Candidates are expected to perform minimum eight (8) assignments for every Lab Practice, and submit a report as a bona fide document to the course instructor. The assignment may be in the form of modeling/ simulation/ programming/ experimental investigation/ fieldwork.
3) MOOC Courses should be Board Specific.
4) **MOOC-**Open Online Course- Students are required to complete an online course through Coursera/NPTEL and other standard open Online Platforms. The course has to be pre- approved from the Department by the Course-Coordinator.
5) **On-Job-Training / Field Project / Internship / Research Based Project:** Students are encouraged to do OJT/FP/Internship/RBP which will enable them towards state-of-art technologies and best practices followed by Industries. A Completion Letter for the same has to be submitted to the course coordinator. Post Internship Presentation and report is to be submitted.

Guidelines for On-Job-Training / Internship / Research Based Project completion:

Credit Assigned: 4 Credits

A. Duration of Internship:

- On-Job-Training / Internship / Research Based Project have a specified duration or number of hours that need to be completed for 4 Credits (typically 6 to 8 weeks)
- The OJT / internship /RBP can be completed during any available slots in Ist, IInd Semesters and the Semester breaks in between.

B. Performance Evaluation:

- OJT / Internship / RBP will be evaluated on their performance during the training. This evaluation will include:

  a) Report submission as per Department Format (Latex file format.)

  b) Presentation and open discussion session

C. Project or Task Completion:

- Certificate of completion form respective company / organization. (As per company policy on its letterhead.) Or (Generalize draft for certificate provided by the Department)

D. Attendance and Punctuality:

- Regular attendance and punctuality are important factors in completing an internship successfully.

E. Evaluation will also be required from the OJT/FP/RBP supervisor for 50 marks.

6) **Journal\Conference-** Students are expected to present their research findings in standard Research conferences and encouraged to publish in reputed Journals approved by Course Coordinator and Research Guide.

7) Students can do their Projects either in Industry or in an academic Institution's\Research Lab. Students pursuing Projects in Industry cannot earn credits through Internship, they are encouraged to earn those credits through MOOCs\Journal\Conference.

8) Exit Norms: Students can exit the programme after one year of completion of earning 44 credits, and can opt for Post Graduate Diploma (PG Diploma) and exit the Degree Program as per NEP 2020.

Program Outcome (PO)

PO1:Data Analysis and Interpretation: Students will demonstrate proficiency in collecting, cleaning, analyzing, and interpreting complex data sets using statistical and machine learning techniques.

PO2: Programming Proficiency: Students will be proficient in programming languages commonly used in data science such as Python, R, and SQL, as well as in relevant libraries and frameworks (e.g., Pandas, NumPy, SciPy, TensorFlow, PyTorch).
Machine Learning and Artificial Intelligence: Students will have a deep understanding of various machine learning algorithms and techniques, including supervised and unsupervised learning, deep learning, reinforcement learning, and natural language processing.

PO3: Data Visualization: Students will be able to effectively visualize data and communicate insights through the use of appropriate visualization tools and techniques, such as Matplotlib, Seaborn, Plotly, and Tableau.

PO4: Big Data Technologies: Students will understand the principles and technologies underlying big data processing, storage, and analysis, including distributed computing frameworks (e.g., Hadoop, Spark), NoSQL databases, and cloud computing platforms.

PO5: Data Mining and Pattern Recognition: Students will be equipped with the knowledge and skills to discover meaningful patterns and insights from large and complex datasets using data mining techniques and pattern recognition algorithms.

PO6: Ethical and Legal Considerations: Students will understand the ethical and legal considerations surrounding data science, including issues related to privacy, bias, fairness, and data protection regulations (e.g., GDPR).

PO7: Problem-Solving and Critical Thinking: Students will develop strong problem-solving and critical thinking skills, enabling them to formulate data-driven solutions to real-world problems across various domains.

PO8: Collaboration and Communication: Students will be able to work effectively in interdisciplinary teams, communicate technical concepts and findings to diverse stakeholders, and contribute to the development of data-driven strategies and solutions.

PO9: Continuous Learning and Professional Development: Students will recognize the importance of lifelong learning and professional development in the rapidly evolving field of data science, and will be prepared to adapt to new technologies, tools, and methodologies throughout their careers.

Research Methodology

Course Code: MSC1 **Course Credit:** 4

Course Outcome(CO):

CO1:Understand Research Principles: Students will grasp the fundamental principles of research, including its purpose, types, and significance in the field of data science.

CO2: Research Design: Students will learn how to formulate research questions, hypotheses, and objectives, and design appropriate research methodologies for data science projects.

CO3: Literature Review: Students will be able to conduct comprehensive literature reviews to identify existing research gaps, theories, and methodologies relevant to their research topics.

CO4: Data Collection Methods: Students will acquire knowledge of various data collection methods, including surveys, interviews, experiments, and observational studies, and understand their applicability to different research contexts in data science.

CO5: Data Processing and Analysis: Students will learn techniques for data preprocessing, cleaning, and analysis using statistical and computational methods,

including descriptive statistics, hypothesis testing, regression analysis, and machine learning algorithms.

CO6: Ethical Considerations: Students will understand the ethical issues and considerations associated with data collection, analysis, and interpretation, and adhere to ethical standards and guidelines in their research practices.

CO7: Research Proposal Development: Students will develop research proposals outlining their research objectives, methodologies, timelines, and expected outcomes, demonstrating their ability to plan and execute research projects in data science.

CO8: Critical Thinking and Problem-Solving: Students will develop critical thinking skills to evaluate research findings, assess the validity and reliability of data, and propose innovative solutions to research problems in data science.

CO9: Communication Skills: Students will effectively communicate their research findings, methodologies, and implications to diverse audiences through written reports, presentations, and visualizations, demonstrating clear and concise communication skills.

CO10:Project Management: Students will manage research projects effectively by setting goals, allocating resources, tracking progress, and adapting research plans as needed to ensure timely completion and quality outcomes.

CO11: Collaboration and Teamwork: Students will collaborate with peers, advisors, and external stakeholders to exchange ideas, share resources, and contribute to interdisciplinary research projects in data science.

CO12: Continuous Learning and Professional Development: Students will cultivate a mindset of lifelong learning and engage in ongoing professional development activities to stay abreast of emerging trends, technologies, and best practices in research methodology and data science.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	3		2			2	2		2
CO2	1	1	2		3			2	3
CO3	1		2			3			
CO4		1		3	2			3	2
CO5	2		2			1	1		2
CO6		2		2	2	3	3	2	
CO7			1		3				2

CO8	3		2	2	3		2	3	3
CO9							3		
CO10		1		3		1	2	3	3
C011	3	3			1		3		
CO12	2	3		3	3		3	3	2

Introduction to Research Methodology:Overview of research methods in data science,Understanding the research process,Importance of research ethics

Research Design:Types of research designs (experimental, observational, quasi-experimental),Selection of research design based on research questions and objectives,Control groups and randomization

Data Collection Methods:Primary data collection methods (surveys, interviews, experiments),Secondary data collection methods (using existing datasets),Sampling techniques (random sampling, stratified sampling, cluster sampling)

Data Preprocessing and Cleaning:Data preprocessing techniques (missing value imputation, outlier detection and treatment, data transformation),Cleaning and preparing datasets for analysis

Statistical Analysis:Descriptive statistics (mean, median, mode, variance, standard deviation),Inferential statistics (hypothesis testing, confidence intervals, p-values),Correlation and regression analysis

Quantitative Research Methods:Experimental design and analysis,Analysis of variance (ANOVA),Factor analysis and principal component analysis (PCA)

Qualitative Research Methods:Overview of qualitative research methods,Content analysis,Thematic analysis

Machine Learning for Research:Overview of machine learning algorithms,Supervised learning (classification, regression),Unsupervised learning (clustering, dimensionality reduction),Model evaluation and validation techniques

Research Ethics and Data Privacy:Ethical considerations in data science research,Privacy and data protection regulations (GDPR, HIPAA),Ensuring anonymity and confidentiality of research participants

Research Project:Planning and executing a research project in data science,Formulating research questions and hypotheses,Designing experiments or studies,Analyzing data and drawing conclusions,Communicating research findings through reports and presentations

Python Programming

Course Code: MSC2 **Course Credit:** 4

Course Outcome(CO):

CO1: Python Fundamentals: Understand the fundamental concepts of Python programming language, including syntax, data types, control structures, functions, and modules.
CO2: Data Structures and Manipulation: Learn how to work with various data structures in Python such as lists, tuples, dictionaries, and sets, and how to manipulate and transform data using built-in functions and libraries like Pandas and NumPy.
CO3: File Handling and Input/Output Operations: Gain proficiency in reading from and writing to files, handling different file formats (e.g., CSV, JSON, XML), and performing input/output operations using standard Python libraries.
CO4: Object-Oriented Programming (OOP): Understand the principles of object-oriented programming and how to create and use classes, objects, inheritance, encapsulation, and polymorphism in Python.
CO5: Functional Programming: Explore functional programming concepts in Python, including higher-order functions, lambda functions, map, filter, and reduce functions, and their application in data processing and manipulation.
CO6: Exception Handling: Learn how to handle errors and exceptions gracefully in Python using try-except blocks, raise statements, and exception handling mechanisms to write robust and reliable code.
CO7: Regular Expressions: Understand the use of regular expressions in Python for pattern matching and text processing tasks, and how to apply them in data cleaning and extraction.
CO8: Python Libraries for Data Science: Gain familiarity with popular Python libraries and frameworks used in data science, such as Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, TensorFlow, and Keras, and learn how to use them for data analysis, visualization, and machine learning.
CO9: Integration with Data Sources and APIs: Learn how to connect Python applications with various data sources and APIs to retrieve, manipulate, and analyze data from different sources such as databases, web APIs, and streaming services.
CO10: Best Practices and Code Optimization: Understand best practices for writing efficient, maintainable, and scalable Python code, including code organization, documentation, testing, and optimization techniques.
CO11:Project Work and Application: Apply Python programming skills to real-world data science projects, including data analysis, visualization, machine learning model development, and deployment, and effectively communicate findings and insights.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	3		2			2	2		2
CO2		1			3				3
CO3			2			3			
CO4		1		3	2			3	
CO5	2		2						2
CO6		2		2	2	3	3		
CO7			1		3				2
CO8	3				3		2	3	
CO9							3		
CO10		1		3		1			3
C011	3	3			1		3		

General Problem-Solving Concepts- Problem solving in everyday life, types of problems, problem solving with computers, difficulties with problem-solving, problem-solving aspects, top-down design. Problem Solving Strategies, Program Design Tools: Algorithms, Flowcharts and Pseudo-codes, implementation of algorithms. Basics of Python Programming: Features of Python, History and Future of Python, Writing and executing Python program, Literal constants, variables and identifiers, Data Types, Input operation, Comments, Reserved words, Indentation, Operators and expressions, Expressions in Python.

Decision Control Statements: Decision control statements, Selection/conditional branching Statements: if, if-else, nested if, if-elif-else statements. Basic loop Structures/Iterative statements: while loop, for loop, selecting appropriate loop. Nested loops, the break, continue, pass, else statement used with loops. Other data types-Tuples, Lists and Dictionaries.

Need for functions, Function: definition, call, variable scope and lifetime, the return

statement. Defining functions, Lambda or anonymous function, documentation string, good programming practices. Introduction to modules, Introduction to packages in Python, Introduction to standard library modules.

Strings and Operations- concatenation, appending, multiplication and slicing. Strings are immutable, string formatting operators, built in string methods and functions. Slice operation, ord() and chr() functions, in and not in operators, comparing strings, Iterating strings, the string module.

Programming Paradigms-monolithic, procedural, structured and object oriented, Features of Object-oriented programming-classes, objects, methods, and message passing, inheritance, polymorphism, containership, reusability, delegation, data abstraction and encapsulation. Classes and Objects: classes and objects, class method and self-object, class variables and object variables, public and private members, class methods.

Introduction• Working with NumPy Arrays, Pandas in Python, Useful Ways to View Data Frame objects in Python.

Reference Books:

1. Python Data Science Handbook - Essential Tools for Working with Data (Jake VanderPlas)
2. DATA SCIENCE AND ANALYTICS WITH PYTHON (JESUS ROGEL - SALAZAR) 3. Mastering Python for Data Science (Madhavan Samir
4. R. G. Dromey, “How to Solve it by Computer”, Pearson Education India; 1st edition, ISBN10: 8131705625, ISBN-13: 978-8131705629 Maureen Spankle, “Problem Solving And Programming Concepts”, Pearson; 9th edition, ISBN-10: 9780132492645, ISBN- 13: 978- 0132492645
5. Romano Fabrizio, “Learning Python,” Packt Publishing Limited, ISBN: 9781783551712, 1783551712
6. Paul Barry, “Head First Python- A Brain Friendly Guide”, SPD O’Reilly, 2nd Edition, ISBN:978-93-5213-482-3
7. Martin C. Brown, “Python: The Complete Reference”, McGraw Hill Education,

ISBN-10: 9789387572942, ISBN-13: 978-9387572942, ASIN: 9387572943

8. Jeeva Jose, P. Sojan Lal, "Introduction to Computing & Problem Solving with Python", Khanna Computer Book Store; First edition, ISBN-10: 9789382609810, ISBN-13: 978- 9382609810

9. Reema Thareja, "Python Programming Using Problem Solving Approach", OxfordUniversity Press, ISBN 13: 978-0-19-948017-6

10. R. Nageswara Rao, "Core Python Programming", Dreamtech Press; Second editionISBN10: 938605230X, ISBN-13: 978-9386052308 ASIN: B07BFSR3LL

R-Programming for Data Science

Course Code: MSC3 **Course Credit:** 4

Course Outcome(CO):

CO1: Proficiency in R Programming: Students will gain proficiency in the R programming language, including data types, data structures, control structures, functions, and packages.
CO2: Data Manipulation: Students will be able to manipulate and transform data using R, including tasks such as subsetting, merging, reshaping, and aggregating datasets.
CO3: Data Visualization: Students will learn how to create informative and visually appealing data visualizations using R packages such as ggplot2, plotly, and lattice.
CO4: Statistical Analysis: Students will understand basic statistical concepts and techniques and will be able to perform statistical analysis using R, including hypothesis testing, regression analysis, and analysis of variance.
CO5: Exploratory Data Analysis (EDA): Students will learn how to conduct exploratory data analysis using R, including techniques for summarizing data, detecting outliers, and identifying patterns and trends.
CO6: Data Cleaning and Preprocessing: Students will learn how to clean and preprocess data using R, including tasks such as handling missing values, dealing with outliers, and standardizing or normalizing data.
CO7: Data Mining and Machine Learning: Students will gain an understanding of data mining and machine learning concepts and will be able to apply various machine learning algorithms using R packages such as caret, randomForest, and e1071.
CO8: Real-World Applications: Students will work on hands-on projects and case studies that apply R programming skills to real-world data science problems across various domains, such as finance, healthcare, marketing, and social media analysis.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	2			2		3			3
CO2			3		3			2	
CO3		3	3			3	2		
CO4			3						2
CO5	2			2	1	2	3		
CO6		2						1	3
CO7	2		2			1			3
CO8	3	3		2	2			3	3

Evolution of R, Features of R, Local Environment support, R Command prompt, R Script File, Comment, R Data types, R Variables, R Operators-function.

R-Decision Making: - R-If statement, R-If…. else statement, R- The if…. else if…else statement-Switch Statement, R- Loop: - Repeat loop, While loop, for loop, Loop, Control statement: - Break, Next.

Vectors: -Vector Creation, Accessing Vector Elements, Vector Manipulation, Lists: - Creating Lists, Naming List Elements, Accessing List Elements, Manipulating List Elements, Merging Lists, Covering Lists To Vectors. Matrices: - Accessing Elements of Matrix, Matrix Computation. Arrays: - Naming columns & rows, Accessing Array Elements, Manipulating Array Elements, and Calculations across Array Elements. Factors: - Factors in Data Frames, changing order of levels, Generating factor levels. Data frames: - Extract data from data frame, Expand data frame.

R-Reshaping: - Joining rows and columns, merging data frames, melting and casting. R- CSV Files: - Getting and starting with directory, Input as a CSV file, Reading CSV file, Analyzing CSV file, writing to CSV file. R- EXCEL File: - Install xlsx Packages, Verify & Load "xlsx" packages, Input as a xlsx file, Reading excel file. R- Binary File: - writing binary file, reading binary file. R- XML File: - Input data, Reading XML file, details of the first node, xml to data node.

R- Statistics Introduction: Mean, median and mode, Minimum and maximum value, Percentiles, Variance and Standard Deviation, Covariance and Correlation, Probability distributions. R Data Set: - Print Variable Values, Sort Variable Values, Analyzing the Data.

R Max and Min, Max and Min, Outliers. R Mean: - Mean, Median, and Mode, R Percentiles.

R- Pie charts: - Pie chart title and color, 3-D Pie Chart. R- Bar Chart: - Bar Chart Labels, Title and color, Group Bar chart and stacked bar chart. R- Box Plot: - Creating a box plot, Box plot with notch. R- Histogram: - Range of x and y values.

Reference Books: -

1. Mark Gardener, Beginning R: The Statistical Programming Language, Wiley India, ISBN:9788126541201 16.

2. Mark Gardener, The Essential R Reference, Wiley India, ISBN: 9788126546015 17. 3. Judith Hurwitz, Alan Nugent, Big Data for Dummies, Wiley India, ISBN: 9788126543281.

4. "Practical Data Science with R", Second Edition, Nina Zumel and John Mount *Foreword by Jeremy Howard and Rachel Thomas,* November 2019, ISBN 9781617295874.

5. "Hands-On Programming with R: Write Your Own Functions and Simulations," by Garrett Grolemund.

Introduction to Database Management System

Course Code: MSC4 **Course Credit**:4

Course Outcome(CO):

CO1: Understanding Database Fundamentals: Students will gain a solid understanding of fundamental concepts in database management systems (DBMS), including data models, database architectures, relational algebra, and database normalization.

CO2: Proficiency in SQL: Students will be proficient in writing SQL queries to perform data manipulation, retrieval, and management tasks on relational databases. They will understand the syntax and semantics of SQL statements, including SELECT, INSERT, UPDATE, DELETE, JOIN, and GROUP BY.

CO3: Database Design and Implementation: Students will learn the principles of database design and implementation, including entity-relationship modeling, schema design, indexing, and optimization techniques. They will be able to design and implement efficient and scalable databases for various applications.

CO4: Web Scraping Techniques: Students will learn the fundamentals of web scraping, including HTML parsing, web crawling, and data extraction from websites using libraries such as BeautifulSoup and Scrapy in Python. They will be able to extract structured data from web pages for analysis and integration into databases.

CO5: Data Integration and ETL Processes: Students will understand the process of data integration, transformation, and loading (ETL) and how it relates to database management. They will learn how to automate ETL processes to extract data from various sources, transform it into a usable format, and load it into a database for analysis.

CO6: Database Security and Privacy: Students will learn about database security principles and techniques to protect sensitive data from unauthorized access, manipulation, and disclosure. They will understand common security threats and vulnerabilities in databases and how to mitigate them.

CO7: Data Mining and Analysis: Students will leverage the data stored in databases to perform data mining and analysis tasks, including exploratory data analysis, pattern discovery, and predictive modeling. They will use SQL queries and analytical tools to extract insights and knowledge from large datasets.

CO8: Application Development with Databases: Students will learn how to develop database-driven web applications and services using frameworks such as Django, Flask, or Ruby on Rails. They will integrate databases with web applications to enable data storage, retrieval, and manipulation.

CO9: Real-World Projects: Students will apply their knowledge and skills in database management and web scraping to real-world projects and case studies. They will work on practical assignments and projects that involve designing databases, scraping data from websites, and analyzing the collected data for insights.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1		3	2	2	2		2		1
CO2				2				2	2
CO3	3	2			1	1		2	
CO4			1				2		2
CO5			1		3				
CO6	1		2	3		3		2	3
CO7				3					
CO8				3		3		3	3
CO9	1	2			2		2		3

SQL Basics – Fundamentals of Structured Query Language, SQL Tables, Joins, Variables Advanced SQL – SQL Functions, Subqueries, Rules, Views, Nested Queries, string functions, pattern matching, Mathematical functions, Date-time functions, etc.

Deep Dive into User Defined Functions -Types of UDFs, Inline table value, multi-statement table, Stored procedures, rank function, triggers, etc. SQL Optimization and Performance - Record grouping, searching, sorting, etc., Clustered indexes, common table expressions. Record grouping, searching, sorting, etc.

Apache spark framework, RDDs, Stopgaps in existing computing methodologies RDDs – RDD persistence, caching, General operations: Transformation, Actions, and Functions, Concept of Key-Value pair in RDDs, Other pair, two pair RDDs, RDD Lineage, RDD Persistence, Word Count Program Using RDD Concepts, RDD Partitioning & How it Helps Achieve Parallelization.

Passing Functions to Spark, Spark SQL Architecture, SQLContext in Spark SQL, User-Defined Functions, Data Frames, Interoperating with RDDs, Loading Data through Different Sources, Performance Tuning, Spark-Hive Integration.

Web Scraping, Interacting with APIs**,** Data Handling with NumPy - NumPy Arrays, CRUD Operations, etc., Linear Algebra – Matrix multiplication, CRUD operations, Inverse, Transpose, Rank, Determinant of a matrix, Scalars, Vectors, Matrices.

Loading the data, data frames, series, CRUD operations, splitting the data, etc. Data Pre-processing -Exploratory Data Analysis, Feature engineering, Feature scaling, Normalization, standardization, etc. Null Value Imputations, Outliers Analysis And Handling, VIF, Bias-variance trade-off, cross validation techniques, train test split, etc.

Data Visualization-Bar charts, scatter plots, count plots, line plots, pie charts, donut charts, etc, with Python Matplotlib, Regression plots, categorical plots, area plots, etc, with Python seaborn.

Introduction to MLOps-MLOps lifecycle, MLOps pipeline, MLOps Components, Processes, etc. Deploying Machine Learning Models- Introduction to Azure Machine Learning, Deploying Machine Learning Models using Azure.

Reference Books:

1. Introduction to Database Systems, CJ Date, Pearson

2. Fundamentals of Database Systems, Elmasri Navrate Pearson Education 3. The Database Systems – The Complete Book, HG Molina, J D Ullman, J Widom Pearson 4. Database Management Systems, Raghuram Krishnan, Johannes Gehrke, TATA McGraw Hill 3rd Edition

5. Database Systems design, Implementation, and Management, Peter Rob & Carlos Coronel 7th Edition.

Data Visualization

Course Code: MSC5 **Course Credit:** 2

Course Outcome(CO):

CO1:Understanding Visualization Principles: Students will demonstrate an understanding of fundamental principles of data visualization, including perceptual principles, graphical integrity, and effective use of visual encodings.

CO2: Proficiency in Visualization Tools: Students will gain proficiency in using a variety of data visualization tools and libraries, such as Matplotlib, Seaborn, Plotly, ggplot2, D3.js, and Tableau, to create static and interactive visualizations.

CO3: Data Preprocessing for Visualization: Students will learn techniques for preprocessing and cleaning data to prepare it for visualization, including handling missing values, transforming data formats, and aggregating data for visualization purposes.

CO4:Exploratory Data Analysis (EDA): Students will be able to conduct exploratory data analysis using visualizations to gain insights into the underlying patterns, trends, and relationships in the data.

CO5:Visualization Techniques for Different Data Types: Students will learn appropriate

visualization techniques for different types of data, including numerical data, categorical data, time series data, and geospatial data.

CO6: Interactive Visualization Design: Students will learn how to design interactive visualizations that allow users to explore and interact with the data dynamically, enabling deeper insights and understanding.

CO7: Application of Visualization in Data Science Projects: Students will apply data visualization techniques and tools to real-world data science projects, demonstrating their ability to use visualization as a key component of the data analysis process.

CO8: Integration with Data Analysis and Machine Learning: Students will learn how to integrate data visualization with data analysis and machine learning workflows, using visualization to explore data, validate models, and communicate results.

CO9: Critical Evaluation of Visualizations: Students will critically evaluate existing visualizations in terms of design choices, effectiveness in conveying information, and adherence to best practices in visualization.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	3	3	2	2	2		2		1
CO2				2		2		2	2
CO3	3	2	3		1	1		2	
CO4	2		1	2			2		2
CO5			1		3		2		
CO6	1	3	2	3		3		2	3
CO7				3	3		2		
CO8				3		3		3	3
CO9	1	2			2		2		3

Introduction to Data Visualization, BI Lifecycle, What is Analysis, importance of data

visualization to the industry, Why Data Visualization became so popular and where we used it. what is Data and what is Visualization Techniques, Data Wrangling, Getting Started Importing Data: Excel As a source, SQL Server as a Source, Web as a Source Direct Query, Limitations, Live Connection, Limitation, which should I choose.

Data Transformation Strategies The power query editor, transform basics Use first row as header, Remove Column, Change type Add Column from examples, Advanced Data Transformation options, Conditional Columns, Fill Up, Down, Pivot, Unpivot, merging queries, appending queries. Leveraging R, Installation and configuration, The R Script transform, M Formula language, shared

Building the Data Model building relationships, Editing relationships, creating a new relationship#xA0, Working with complex relationship, many to many relationships, cross filtering directions, Enabling filtering from the many side of a relationship role playing tables, Importing the date table, Usability enhancements, Hiding tables and columns, Renaming tables and columns, Default summarization, How to display one column but sort by another, Data Categorization, Creating hierarchies summary.

Leveraging DAX, Building calculated columns, String functions – Month, Year Format function – Month, Year, Age Calculation, Switch () – the basics calculated measures – basic aggregation, Total Sales, Total Cost, Profit, Profit Margin, Optional Parameters, Filter Context, Percentage of total calculation, Time Intelligence, year to date sales, YTD sales (Fiscal Calendar), Prior Year Sales#xA0

Data Visualization Basics, Visuals for Filtering, Interactive Filtering, The Slicer Visual Visualizing tabular data: the table visual, the matrix visual, Visualizing categorical data: Bar and Column charts, Pie and Donut Charts, The Tree map Visual, The scatter Chart, Visualizing Trend Data: Line and Area Charts, The Waterfall Charts, The Funnel Chart, Visualizing KPI Data: The Gauge Visual, The KPI Visual, Visualizing Geographical Data: The map visual, The Filled Map Visual, The ArcGIS map visual

Configuring drill through filters, Storytelling with the selection pane and bookmarks Bookmarks pane, Selection pane, Summary

Text Books:

1. Microsoft Power BI Quick Start Guide, By Devin Knight, Brain Knight, Mitchell Pearson and Manual Quintana, Published by Packt Publishing Ltd, ISBN 978-1-78913-822-1
2. Microsoft Power BI Data Analyst Certification Guide by Orrin Edenfield-

Lab Practice-I

Course Code: MSC6 **Course Credit**: 4

Lab Overview:

The Python, SQL, and R Programs Lab is designed to provide hands-on experience in data manipulation, analysis, and visualization using three powerful programming languages commonly used in data science and analytics.

Objectives:

1. Python Programming Skills: Develop proficiency in Python programming for data handling, analysis, and visualization.
2. SQL Database Operations: Learn SQL for managing databases, querying data, and performing data manipulation tasks.
3. R Programming for Analytics: Gain skills in R programming for statistical analysis, machine learning, and advanced visualization.

Lab Structure:

1. Python Programming:

- Introduction to Python: Basic Python syntax, data types, and control structures.
- Data Handling with Pandas: Importing, cleaning, and transforming data using the Pandas library.
- Data Visualization with Matplotlib and Seaborn: Creating visualizations to explore data patterns and relationships.
- Introduction to NumPy: Introduction to numerical computing with NumPy for array operations and mathematical functions.
- Introduction to Scikit-Learn: Basic machine learning concepts and implementation of models for classification and regression tasks.

2. SQL Database Operations:

- Introduction to Relational Databases: Overview of relational database concepts and SQL language.
- Creating and Managing Databases: Creating databases, tables, and managing database objects.
- Data Querying with SQL: Writing SQL queries to retrieve, filter, and aggregate data from databases.
- Data Manipulation with SQL: Performing data manipulation tasks such as inserting, updating, and deleting records.
- Advanced SQL Concepts: Working with joins, subqueries, and indexing for optimizing database operations.

3. R Programming:

- Introduction to R Programming: Basics of R syntax, data types, and control structures.
- Data Manipulation with dplyr: Data manipulation tasks using the dplyr package for

filtering, summarizing, and arranging data.

- Statistical Analysis with R: Perform statistical analysis and hypothesis testing using built-in functions and packages like stats.
- Introduction to ggplot2: Creating advanced visualizations using the ggplot2 package for exploratory data analysis.
- Machine Learning with caret: Introduction to machine learning algorithms and implementation using the caret package.

Lab Requirements:

- Software: Install Python (with Anaconda distribution), MySQL or PostgreSQL database, R programming language, and necessary libraries and packages (Pandas, Matplotlib, Seaborn, NumPy, Scikit-Learn).
- Database Access: Access to a local or remote MySQL or PostgreSQL database for SQL exercises.
- Integrated Development Environment (IDE): Use Jupyter Notebook or any preferred IDE for writing and executing Python and R code.
- Learning Resources: Provide textbooks, online tutorials, and reference materials for additional learning and practice.

Statistics Essential for Data Science

Subject Code: MSC7 **Course Credit:4**

Course Outcome(CO):

CO1. Understanding of Fundamental Concepts: Mastery of foundational statistical concepts including probability, descriptive statistics, inferential statistics, and hypothesis testing.
CO2. Statistical Methods and Techniques: Proficiency in applying various statistical methods and techniques commonly used in data science such as regression analysis, analysis of variance (ANOVA), and non-parametric methods.
CO3. Probability Theory: Solid understanding of probability theory including random variables, probability distributions, and probability density functions, essential for modeling uncertainty in data.
CO4. Sampling Techniques: Knowledge of sampling techniques and their application in collecting data from populations, including simple random sampling, stratified sampling, and cluster sampling.
CO5. Exploratory Data Analysis (EDA): Ability to perform exploratory data analysis techniques such as data visualization, summary statistics, and outlier detection to gain insights into the underlying patterns and structure of the data.
CO6. Statistical Modeling: Proficiency in building statistical models to describe and analyze relationships between variables in datasets, including linear models, logistic regression, and time series analysis.

CO7. Inferential Statistics: Understanding of inferential statistics methods such as confidence intervals and hypothesis testing, and their application in making inferences about population parameters based on sample data.
CO8. Statistical Software Proficiency: Familiarity with statistical software packages such as R, Python (with libraries like NumPy, Pandas, and SciPy), or specialized statistical software like SPSS or SAS for data analysis and visualization.
CO9. Practical Application in Data Science: Ability to apply statistical concepts and techniques to real-world data science problems, including data preprocessing, feature engineering, model evaluation, and interpretation of results.
CO10. Critical Thinking and Problem-Solving: Development of critical thinking skills to assess the appropriateness of statistical methods for different types of data and problem domains, and to formulate effective solutions to data analysis challenges.
CO11. Communication of Statistical Findings: Proficiency in communicating statistical findings and insights effectively to both technical and non-technical stakeholders through written reports, visualizations, and presentations.

Basic Statistics for Data Science:Introduction**:** Understanding the Data, various data types, Various variable types, Population and Sample, Sampling techniques, Measures of Central Tendency, Measures of Variability, Asymmetry, Measures of Dispersion

The Fundamentals of Descriptive Statistics:Random variable and probability, Z-Scores, Standard Normal Distribution, Probability Distributions, Discrete Probability Distribution: Binomial Distribution, Normal Distribution, Practical Example: Descriptive Statistics.

Statistical Inference :Sampling variation, Statistic, Central Limit Theorem, Confidence Interval, Point Estimate, Practical Example: Inferential Statistics

Testing the Data:Hypothesis Testing, Procedure of Hypothesis Testing, Type I and Type II error, One sample t- test, Two sample t- test, Anova and Chi-Square Test, Parametric Test, Non Parametric Test, Practical Example: Hypothesis Testing

Exploratory Data Analysis:Outliers, Boxplot, Scatter Plot, Correlation, Pearson Correlation, Covariance, Practical Example

Regression Modeling**:**Logistic and Regression Techniques, Sensitivity,Information Gain, Entropy, Heteroscedasticity, Homoscedasticity.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	3			2				2	
CO2		3		2		2	3		
CO3	2								3
CO4			1		3		3	3	
CO5		2							
CO6	3			3		2	1	1	1
CO7		3	3	3	3				
CO8	3			2	3	2	2	2	1
CO9									
CO10	2		2	2			3		
C011		3			3			3	2

Reference Books: -

1. Statistics for Data Science (Miller James D.)
2. Practical Statistics for Data Scientists (Bruce Peter)
3. Statistical Data Book 2nd Edition (R. S. Naagarazan, C. Muralidharan)

Machine Learning & Artificial Intelligence

Course Code: MSC8 **Course Credit:** 4

Course Outcome(CO):

CO1: Understanding of Machine Learning Principles: Students will demonstrate a comprehensive understanding of fundamental machine learning concepts, including supervised learning, unsupervised learning, and reinforcement learning.

CO2: Proficiency in Machine Learning Algorithms: Students will be proficient in implementing and applying a variety of machine learning algorithms, such as linear regression, logistic regression, decision trees, random forests, support vector machines, k-nearest neighbors, clustering algorithms, and neural networks.

CO3: Advanced Techniques in Deep Learning: Students will acquire knowledge and skills in deep learning techniques, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), generative adversarial networks (GANs), and transfer learning.

CO4: Hands-on Experience with ML & AI Tools: Students will gain practical experience using popular libraries and frameworks for machine learning and deep learning, such as TensorFlow, Keras, PyTorch, scikit-learn, and NLTK.

CO5: Model Evaluation and Validation: Students will learn techniques for evaluating and validating machine learning models, including cross-validation, hyperparameter tuning, model selection, and performance metrics such as accuracy, precision, recall, and F1-score.

CO6: Applications of ML & AI in Data Science: Students will explore a wide range of applications of machine learning and artificial intelligence in data science, including predictive modeling, classification, regression, clustering, anomaly detection, natural language processing, computer vision, and recommender systems.

CO7: Ethical and Responsible AI Practices: Students will understand the ethical implications of deploying machine learning and artificial intelligence systems, including issues related to fairness, transparency, accountability, bias, and privacy, and will be able to incorporate responsible AI principles into their work.

CO8: Research and Innovation in ML & AI: Students will be prepared to engage in research and innovation in the field of machine learning and artificial intelligence, including the development of novel algorithms, models, and techniques to address emerging challenges and opportunities.

CO9: Real-world Applications and Case Studies: Students will analyze real-world datasets and case studies to identify appropriate machine learning and artificial intelligence techniques, apply them to solve practical problems, and interpret the results in the context of domain-specific challenges and requirements.

.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9

CO1	2	3	2	2	2		2		1
CO2	2	2		2				2	2
CO3	3	2			1	1		2	
CO4		2	1	2			2		2
CO5	3		1		3		2		3
CO6	1		2	3	3	3	2	2	3
CO7		3		3	3		2		
CO8	3	2		3		3		3	3
CO9	1	2			2		2		3

Decision trees, a type of data mining algorithm that can select from among a large number of variables those and their interactions that are most important in predicting the target or response variable to be explained. Decision trees create segmentations or subgroups in the data, by applying a series of simple rules or criteria over and over again, which choose variable constellations that best predict the target variable.

Random forests, a type of data mining algorithm that can select from among a large number of variables those that are most important in determining the target or response variable to be explained. Unlike decision trees, the results of random forests generalize well to new data.

Lasso regression analysis is a shrinkage and variable selection method for linear regression models. The goal of lasso regression is to obtain the subset of predictors that minimizes prediction error for a quantitative response variable. The lasso does this by imposing a constraint on the model parameters that causes regression coefficients for some variables to shrink toward zero. Variables with a regression coefficient equal to zero after the shrinkage process are excluded from the model. Variables with non-zero regression coefficient variables are most strongly associated with the response variable. Explanatory variables can be either quantitative, categorical or both. In this session, you will apply and interpret a lasso regression analysis. You will also develop experience using k-fold cross validation to select the best fitting model and

obtain a more accurate estimate of your model's test error rate.

Cluster analysis is an unsupervised machine learning method that partitions the observations in a data set into a smaller set of clusters where each observation belongs to only one cluster. The goal of cluster analysis is to group, or cluster, observations into subsets based on their similarity of responses on multiple variables. Clustering variables should be primarily quantitative variables, but binary variables may also be included. In this session, we will show you how to use k-means cluster analysis to identify clusters of observations in your data set. You will gain experience in interpreting cluster analysis results by using graphing methods to help you determine the number of clusters to interpret, and examining clustering variable means to evaluate the cluster profiles. Finally, you will get the opportunity to validate your cluster solution by examining differences between clusters on a variable not included in your cluster analysis.

Version Control -What is version control, types, SVN, Git Lifecycle, Common Git commands, working with branches in Git, GitHub collaboration (pull request), GitHub Authentication (SSH and Http), Merging branches, Resolving merge conflicts, Git workflow.
Artificial Intelligence Basics-Introduction to Keras API and TensorFlow, Neural Networks Neural networks, Multi-layered Neural Networks, Artificial Neural Networks, Deep Learning Deep neural networks, Convolutional Neural Networks, Recurrent Neural Networks, GPU in deep learning, Autoencoders, restricted Boltzmann machines.

Reference Books:

1. Peter Flach, Machine Learning: The Art and Science of Algorithms that make sense of data, Cambridge University Press, 1st Edition, 2012, ISBN No.: 978-1-316-50611-0 2. Ethem Alpaydin, Introduction to Machine Learning, PHI, 2nd edition, 2013, 978-0-262- 01243-0
3. Kevin Murphy, Machine Learning: A Probabilistic Approach, MIT Press, 1st Edition, 2012,

ISBN No.: 978-0262-30616-4

4. C.M. Bishop, Pattern Recognition and Machine learning, Springer, 1st Edition, 2013, ISBN

No.: 978-81-322-0906-5

5. Hastie, Tibshirani, Friedman, Introduction to statistical machine learning with applications in R,

Springer, 2nd Edition, 2013, ISBN No.: 978-1-4614-7138-7

6. Tom Mitchell, Machine Learning, McGraw Hill, 1997, 0-07-042807-7

7. Parag Kulkarni, Reinforcement and Systemic Machine learning for Decision Making, Wiley

IEEE Press, 2012, 978-0-470-91999-6

8. M. F. Der, L. K. Saul, S. Savage, and G. M. Voelker (2014). Knock it off: profiling the online

Natural Language Processing

Course Code: MSC9 **Course Credit:** 4

Course Outcome(CO):

CO1:Understanding NLP Fundamentals: Gain a solid understanding of the fundamental concepts, theories, and techniques in natural language processing, including tokenization, stemming, lemmatization, part-of-speech tagging, syntactic parsing, semantic analysis, and named entity recognition.

CO2: Text Representation and Feature Engineering: Learn various methods for representing text data, including bag-of-words, TF-IDF (Term Frequency-Inverse Document Frequency), word embeddings (e.g., Word2Vec, GloVe), and contextual embeddings (e.g., BERT, GPT).

CO3: Text Preprocessing Techniques: Acquire skills in preprocessing text data to remove noise, handle capitalization and punctuation, handle special characters, remove stop words, and perform text normalization.

CO4: Text Classification and Sentiment Analysis: Understand the principles and techniques of text classification and sentiment analysis, including binary and multiclass classification, sentiment lexicons, and machine learning algorithms for sentiment analysis (e.g., Naive Bayes, Support Vector Machines, Recurrent Neural Networks).

CO5: Topic Modeling and Document Clustering: Gain knowledge of topic modeling techniques (e.g., Latent Dirichlet Allocation) and document clustering algorithms (e.g., K-means clustering) for unsupervised analysis of text data.

CO6: Sequence Labeling and Sequence-to-Sequence Models: Understand sequence labeling tasks such as part-of-speech tagging and named entity recognition, and learn

about sequence-to-sequence models for tasks like machine translation and text summarization.

CO7: Text Generation and Language Modeling: Learn about language modeling techniques and models for text generation tasks, including n-gram models, recurrent neural networks (RNNs), and transformer-based models (e.g., GPT, BERT).

C08: Applications of NLP: Explore real-world applications of natural language processing across various domains, including information retrieval, question answering, chatbots, sentiment analysis in social media, document summarization, and machine translation.

CO9: Hands-on Projects and Practical Applications: Apply NLP techniques and algorithms to real-world datasets through hands-on projects, case studies, and practical exercises, and develop skills in implementing and evaluating NLP models using Python libraries such as NLTK, spaCy, scikit-learn, and TensorFlow/PyTorch.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	3	3	2	2	2		2		1
CO2			3	2	3	3		2	2
CO3	3	2			1	1		2	
CO4		3	1		1		2		2
CO5	2		1		3		1	3	
CO6	1		2	3		3		2	3
CO7	2	2		3	1				
CO8		2	2	3		3		3	3
CO9	1	2		2	2		2		3

What is NLP? Why is NLP Difficult? History of NLP, Advantages of NLP, Disadvantages of NLP,

Components of NLP, Applications of NLP, How to build an NLP pipeline? Phases of NLP, NLP APIs, NLP Libraries,Various Tokenizers, Tokenization, Frequency Distribution, Stemming, Types of Stemming, Stop Words, Normalization, POS Tagging, Lemmatization, Named Entity Recognition (NER), NLPModel- Unigram, Bigram, Trigram N-gram Model

Overview of Machine Learning, Bag-of-Words, Term Frequency, Count vectorizer, Inverse Document Frequency, Text conversion, Confusion Matrix, Classification Metrics, Naive Bayes Classifier, Smoothing technique, Support vector Machine

Language Modeling, Sequence Tagging, Sequence Tasks, Predicting Sequence of Tags, Syntax Trees, Context Free Grammars, Chunking, Automatic Paraphrasing of Texts, Chinking.

Need of MT, Problems of Machine Translation, MT Approaches, Direct Machine Translations, Rule-Based Machine Translation, Knowledge Based MT System, Statistical Machine Translation (SMT), Parameter learning in SMT (IBM models) using EM), Encoder-decoder architecture, Neural Machine Translation

Using the NLP concepts, build a recommendation engine and an AI chatbot assistant using AI.

Reference Books:

1. Natural Language Processing and Information Retrieval First Edition (TIWARY, U.S, SIDDIQUI ,TANVEER)

2. Speech and Language Processing - Written by Daniel Jurafsky and James Martin. 3. Natural Language Processing with Python. Steven Bird, Ewan Klein, and Edward Lope,

Introduction to Deep Learning

Course Code: MSC10 **Course Credit:** 4

Course Outcome(CO):

CO1: Understanding Deep Learning Fundamentals: Students will gain a comprehensive understanding of deep learning concepts, including neural networks, activation functions, loss functions, optimization algorithms, and backpropagation.

CO2: Implementing Deep Learning Models: Students will be able to implement deep learning models using popular frameworks such as TensorFlow and PyTorch. They will understand how to build, train, and evaluate deep neural networks for various tasks, including classification, regression, and clustering.

CO3: Exploring Convolutional Neural Networks (CNNs): Students will learn about CNN architectures and their applications in computer vision tasks such as image classification, object detection, and image segmentation. They will gain hands-on experience in building and training CNNs for real-world problems.

CO4: Understanding Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM): Students will explore RNN architectures, including LSTM networks, and their applications in sequential data analysis, such as natural language processing (NLP), time series forecasting, and speech recognition.

CO5: Applying Deep Learning Techniques to Natural Language Processing (NLP): Students will learn how to apply deep learning techniques to process and analyze textual data. They will explore architectures such as word embeddings, recurrent neural networks (RNNs), and transformers for tasks such as sentiment analysis, text classification, and machine translation.

CO6: Handling Overfitting and Regularization Techniques: Students will understand common challenges in deep learning, such as overfitting, and learn techniques to mitigate them, including dropout regularization, batch normalization, and data augmentation.

CO7: Hyperparameter Tuning and Model Evaluation: Students will gain skills in hyperparameter tuning and model evaluation techniques to optimize deep learning models' performance. They will learn about techniques such as grid search, random search, cross-validation, and evaluation metrics for classification and regression tasks.

CO8: Deploying Deep Learning Models: Students will learn about different deployment strategies for deep learning models, including deploying models to production environments, containerization using Docker, and deploying models to cloud platforms such as AWS, Azure, or Google Cloud.

CO9: Ethical Considerations in Deep Learning: Students will explore ethical considerations and challenges associated with deep learning, including bias and fairness issues, privacy concerns, and the responsible use of AI technologies.

CO10:Hands-on Projects and Case Studies: Students will engage in hands-on projects and case studies to apply deep learning techniques to real-world datasets and problems. They will gain practical experience in designing, implementing, and evaluating deep learning solutions across various domains.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	2	3	2	2	2		2		1
CO2	2	2		2				2	2
CO3	3	2			1	1		2	
CO4		2	1	2			2		2
CO5	3		1		3		2		3
CO6	1		2	3	3	3	2	2	3
CO7		3		3	3		2		
CO8	3	2		3		3		3	3
CO9	1	2			2		2		3
CO10	2	3		3		3	2	3	

Membership Functions, Fuzzification and Methods, Defuzzification and Methods, Fuzzy Logic, Fuzzy Rules and Fuzzy Reasoning, Fuzzy Inference Systems, Fuzzy Expert Systems, Fuzzy Decision Making. Fuzzy Control Systems, Fuzzy Classification. Genetic Algorithms: Introduction to Genetic Algorithms (GA), Search space, Working Principle, Simple GA, Operators, Fitness function, Multi-level Optimization.

Perceptron, Perceptron Learning Algorithm, Sigmoid Neuron, Shallow neural networks,Deep neural networks, Feedforward Neural networks, Gradient descent and the backpropagation algorithm

Learning Parameters of a feedforward neural network, the vanishing gradient problem, and ways to mitigate it, RelU Heuristics for avoiding bad local minima, Heuristics for faster training, Nesterov Accelerated GD, Stochastic GD, AdaGrad, RMSProp, Adam, Momentum.Adagrad, Principal Component Analysis and its interpretations, Singular Value Decomposition. Convolutional Neural Networks, LeNet, AlexNet, ZF-Net, VGGNet, GoogLeNet, ResNet, Visualizing Convolutional Neural Networks, Guided Backpropagation, Deep Dream, Deep Art,

Fooling Convolutional Neural Networks

Introduction to soft computing: , Paradigms soft computing, Features, Components, Techniques,

Applications, Neural Networks, Fuzzy logic, Genetic Algorithms, Hybrid systems, Introduction to

Fuzzy logic: Classical and Fuzzy sets, operations, properties, FuzzyRelations.

RNN, LSTM, GRU models, Application to NLP, language models, machine translation, image

captioning, video processing, visual question answering, video processing, learning from descriptions,

Attention Mechanism, Attention over images

Reference Books:

1. S. N. Sivanandam & S.N.Deepa ``Principles of Soft computing", John Wiley & Sons, 2.

S. Rajasekaran, G. A. Vijayalakshmi, Neural Networks, Fuzzy Logic and Genetic

Algorithms: Synthesis & Applications, PHI.

3. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.

4. David E. Goldberg., Genetic Algorithms: in Search and Optimization,PHI

5. Jyh: Shing Roger Jang, Chuen:Tsai Sun, EijiMizutani, Neuro:Fuzzy and Soft

Computing, Prentice:Hall of India, 2003

6. Timothy J. Ross, Fuzzy Logic with Engineering Applications (Wiley)

7. George J. Klir and Bo Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications,

Prentice Hall, 3. An Introduction to Genetic Algorithm Melanic Mitchell (MIT Press)

8. Evolutionary Algorithm for Solving Multi-objective, Optimization Problems (2nd Edition),

Collelo, Lament, Veldhnizer (Springer)

9. Neural Networks and Learning Machines Simon Haykin (PHI).

10. Neural Networks, Fuzzy logic, and Genetic Algorithms, S. Rajasekaran& G. A. V.

Pai, PHI. Case Studies:

1. Image classifier for identifying cat vs dogs using CNN
2. Image classifier for identifying cat vs dogs using CNN

Massive Open Online Courses (MOOCs)

Course Code: MSC11 **Course Credit:** 2

Course Contents:

1. Introduction to Machine Learning - NPTEL (https://nptel.ac.in/courses/106/106/106106139/)
2. Machine Learning - NPTEL (https://nptel.ac.in/courses/106/106/106106202/)
3. Machine Learning for Science and Engineering Applications - NPTEL (https://nptel.ac.in/courses/106/106/106106198/)
4. Introduction to Machine Learning - NPTEL (https://nptel.ac.in/courses/106/105/106105152/)
5. Deep Learning (Part-I) - NPTEL (https://nptel.ac.in/courses/106/106/106106184/)
6. Deep Learning - NPTEL Online Courses (https://onlinecourses.nptel.ac.in/noc19_cs54/preview)
7. Naive Bayes from Scratch - Analytics Vidhya (https://courses.analyticsvidhya.com/courses/naive-bayes)
8. Getting Started with Neural Networks - Analytics Vidhya (https://courses.analyticsvidhya.com/courses/getting-started-with-neural-networks)
9. Machine Learning - Stanford Online (https://www.coursera.org/learn/machine-learning)
10. Microsoft Exam DA-100: Analyzing Data with Microsoft Power BI
11. Microsoft Exam PL-300: Microsoft Power BI Data Analyst.
12. Microsoft Exam: Microsoft Certified: Azure Data Scientist Associate

Lab Practice-II

Course Code: MSC12 **Course Credit:** 4

Lab Overview:

The Machine Learning and Artificial Intelligence, Natural Language Processing & Introduction to Deep Learning Programs Lab is designed to provide hands-on experience and theoretical knowledge in advanced topics in data science and artificial intelligence.

Objectives:

1. Foundations of Machine Learning and AI: Understand the core concepts and algorithms of machine learning and artificial intelligence.
2. Natural Language Processing: Learn techniques for processing and analyzing natural language data.
3. Introduction to Deep Learning: Gain insights into deep learning architectures and applications.

Lab Structure:

1. Machine Learning and Artificial Intelligence :

- Introduction to Machine Learning: Overview of machine learning concepts, types of learning, and supervised vs. unsupervised learning.
- Regression and Classification Algorithms: Implement regression and classification algorithms such as linear regression, logistic regression, decision trees, and k-nearest neighbors.
- Clustering and Dimensionality Reduction: Perform clustering using algorithms like K-means and hierarchical clustering, and understand dimensionality reduction techniques like PCA.
- Introduction to Artificial Intelligence: Understand the fundamentals of artificial intelligence, including expert systems, knowledge representation, and reasoning.

2. Natural Language Processing :

- Text Preprocessing: Learn techniques for text cleaning, tokenization, stemming, and lemmatization.
- Text Representation: Convert text data into numerical representations using methods like Bag-of-Words, TF-IDF, and word embeddings (Word2Vec, GloVe).
- Text Classification and Sentiment Analysis: Implement text classification models and sentiment analysis using techniques like Naive Bayes, Support Vector Machines, and Recurrent Neural Networks (RNNs).
- Named Entity Recognition and Part-of-Speech Tagging: Explore advanced NLP tasks like named entity recognition (NER) and part-of-speech (POS) tagging.

3. Introduction to Deep Learning :

- Neural Networks Basics: Understand the architecture of neural networks, including perceptrons, activation functions, and feedforward networks.
- Deep Learning Frameworks: Introduce popular deep learning frameworks such as TensorFlow and PyTorch.
- Convolutional Neural Networks (CNNs): Learn about CNN architecture and its applications in image classification and object detection.
- Recurrent Neural Networks (RNNs): Explore RNNs for sequential data processing, including text generation and time series prediction.
- Introduction to Generative Adversarial Networks (GANs): Understand the concept of GANs and their applications in generating synthetic data.

Lab Requirements:

- Software and Libraries: Install Python with necessary libraries (scikit-learn, TensorFlow, PyTorch, NLTK, spaCy) and deep learning frameworks.
- Datasets: Provide datasets for hands-on exercises covering various domains such as text data, image data, and tabular data.
- Computational Resources: Ensure access to computational resources like GPUs for training deep learning models, if possible.
- Learning Resources: Offer textbooks, online tutorials, and research papers for further reading and exploration.

Big Data and Data Clouds

Course Code: MSC13 **Course Credit:** 4

Course Outcome(CO):

CO1: Understanding Big Data Concepts: Students will gain a comprehensive understanding of the fundamental concepts of big data, including volume, velocity,

variety, veracity, and value, as well as the challenges and opportunities associated with processing large-scale datasets.

CO2: Exploration of Distributed Computing: Students will learn about distributed computing frameworks such as Hadoop and Apache Spark, including their architecture, components, and functionalities, and how they enable the processing and analysis of big data across distributed computing environments.

CO3: Data Storage and Management: Students will explore various storage and management technologies for big data, including distributed file systems (e.g., Hadoop Distributed File System), NoSQL databases (e.g., MongoDB, Cassandra), and NewSQL databases, and understand their advantages and limitations.

CO4: Data Processing and Analytics: Students will acquire hands-on experience in processing and analyzing big data using distributed computing frameworks and tools, including MapReduce, Spark RDDs, Spark SQL, and Spark MLlib, to perform tasks such as data transformation, aggregation, and machine learning.

CO5: Scalability and Performance Optimization: Students will learn techniques for optimizing the scalability, performance, and efficiency of big data processing and analytics workflows, including parallelization, data partitioning, caching, and resource management strategies.

CO6: Data Integration and ETL Pipelines: Students will gain proficiency in designing and implementing extract, transform, load (ETL) pipelines for integrating and preprocessing heterogeneous data sources into unified datasets suitable for analysis and visualization.

CO7: Cloud Computing Fundamentals: Students will understand the principles and architectures of cloud computing, including Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS), and how cloud platforms such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) support big data processing and analytics.

CO8: Security and Privacy Considerations: Students will explore security and privacy issues related to big data and data clouds, including data encryption, access control, authentication, and compliance with data protection regulations such as GDPR.

CO9: Real-world Applications and Use Cases: Students will examine real-world applications and use cases of big data and data clouds across various domains, including finance, healthcare, e-commerce, social media, and IoT, and analyze the impact of big data technologies on business innovation and decision-making.

CO10: Project Work and Case Studies: Students will apply their knowledge and skills to real-world projects and case studies involving the design, implementation, and evaluation of big data solutions using cloud-based technologies, and present their findings and recommendations to peers and stakeholders.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9

CO1		3	3	2	3		2	3	1
CO2			3	2		2		2	2
CO3	3	2			1	1		2	
CO4	2		1		2		2		2
CO5		2	1		3	2		2	
CO6	1		2	3	3	3		2	3
CO7		2		3	2				
CO8				3		3		3	3
CO9	1	2			2		2		3
CO10	3	3	3	3		2		2	2

Introduction Evolution of Cloud Computing –Essential Characteristics of cloud computing –

Operational models such as private, dedicated, virtual private, community, hybrid and public cloud – Service models such as IaaS, PaaS and SaaS – Governance and Change Management – Business drivers, metrics and ypical use cases. Example cloud vendors – Google cloud platform, Amazon AWS, Microsoft Azure,Pivotal cloud foundry and OpenStack.

Security – Virtual Machine Security. Design a machine learning solution:Determine the appropriate compute specifications for a training workload,Describe model deployment requirements,Select which development approach to use to build or train a model

Manage an Azure Machine Learning workspace:Create an Azure Machine Learning workspace,Manage a workspace by using developer tools for workspace interaction,Set up Git integration for source control,Create and manage registries,Manage data in an Azure Machine Learning workspace,Select Azure Storage resources,Register and maintain datastores,Create and manage data assets

Explore data by using data assets and data stores:Access and wrangle data during interactive developmentWrangle interactive data with Apache Spark,Create models by using the Azure Machine Learning designer,Create a training pipeline,Consume data assets from the designer,Use custom code components in designer,Evaluate the model, including responsible AI guidelines

Use automated machine learning to explore optimal models,Use automated machine learning for tabular data,Use automated machine learning for computer vision,Use automated machine learning for natural language processing,Select and understand training options, including preprocessing and algorithms,Evaluate an automated machine learning run, including responsible AI guidelines

Develop code by using a compute instance,Track model training by using MLflow,Evaluate a model,Train a model by using Python SDK v2,Use the terminal to configure a compute instance,Tune hyperparameters with Azure Machine Learning,Select a sampling method, Define the search space, Define the primary metric,Define early termination options,Implement training pipelines

Create a pipeline, Pass data between steps in a pipeline,Run and schedule a pipeline,Monitor pipeline runs, Create custom components, Use component-based pipelines,Manage models in Azure Machine Learning,Describe MLflow model output:Identify an appropriate framework to package a model,Assess a model by using responsible AI principles,Deploy a model

Apply machine learning operations (MLOps) practices:Trigger an Azure Machine Learning job, including from Azure DevOps or GitHub,Automate model retraining based on new data additions or data changes,Define event-based retraining triggers

Reference Books:

1. Cloud Computing: Concepts, Technology & Architecture" by Thomas Erl, Zaigham Mahmood, and Ricardo Puttini
2. Hands-On Machine Learning with Azure: Build powerful models with cognitive AI tools" by Jeff Prosise and Richard Conway
3. Azure Machine Learning for Beginners: Perform complex data analysis, create interactive visualizations, and deploy your model to the cloud using Azure Machine Learning" by Jeff Heaton
4. Mastering Azure Machine Learning: A comprehensive guide to implementing, deploying, and troubleshooting machine learning models using Azure Machine Learning" by Zoran Cvijanovic and Abhishek Kumar

Web Framework

Course Code: MSC14 **Course Credit:** 4

Course Outcome(CO):

CO1: Understanding of Web Development Concepts: Students will gain a comprehensive understanding of web development concepts, including client-server architecture, HTTP protocol, request-response cycle, and MVC (Model-View-Controller) architecture.

CO2: Introduction to Django: Students will be introduced to the Django web framework, its features, advantages, and components, including ORM (Object-Relational Mapping), URL routing, template system, and form handling.

CO3: Setting up Development Environment: Students will learn to set up a development environment for Django, including installing Django, configuring settings, managing virtual environments, and integrating version control systems like Git.

CO4: Creating Django Projects and Apps: Students will learn to create Django projects and apps, understand the structure of a Django project, and develop modular and reusable web applications using Django's app architecture.

CO5: Django Models and Database Integration: Students will learn to define data models using Django's ORM, understand database migrations and schema evolution, perform CRUD (Create, Read, Update, Delete) operations, and integrate Django with various database systems.

CO6: Views and Templates: Students will learn to create views to handle HTTP requests, render dynamic content using templates, understand URL patterns and routing mechanisms, and pass data from views to templates.

CO7: Django Forms: Students will learn to create HTML forms using Django's form classes, handle form validation and submissions, customize form behavior, and integrate forms with Django models for data validation and manipulation.

CO8: Admin Panel and Content Management: Students will explore Django's built-in admin interface for content management, learn to customize the admin panel for specific models, and add custom actions and filters to streamline content management tasks.

CO9: Authentication and Authorization: Students will learn to implement user authentication and authorization in Django, manage user sessions and passwords, configure permissions and access control rules, and secure web applications against common security threats.

C010:Integration with Frontend Frameworks: Students will learn to integrate Django with frontend frameworks like Bootstrap, utilize static files and media in Django projects, and implement asynchronous behavior using AJAX in Django applications.

CO11: Testing and Debugging: Students will learn best practices for testing Django applications, write unit tests for Django views and models, debug common errors using Django's debugging tools, and ensure the reliability and robustness of web applications.

CO12: Deployment and Scaling: Students will learn to prepare Django applications for deployment, choose appropriate hosting platforms (e.g., Heroku, AWS), configure production settings, and scale Django applications to handle increased traffic and workload.

CO13: Advanced Topics: Students may explore advanced topics such as signals and event handling, custom middleware development, caching strategies, internationalization and localization, and optimization techniques for improving Django application performance.

CO14: Project Work: Students will apply their knowledge and skills to develop a real-world Django project, from conception to deployment, demonstrating their ability to design, develop, and deploy scalable and secure web applications using Django.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1		3	2	2	2		2		1
CO2				2				2	2
CO3	3	2			1	1		2	
CO4			1				2		2
CO5			1		3		1		
CO6	1		2	3		3		2	3
CO7	3	2		3			2		
CO8		2		3		3		3	3
CO9	1	2			2		2		3
CO10		3	3	3		3			3
CO11	3		2	2			3	2	
CO12		2			2				
CO13			1			2	3	2	
CO14	1	2	3	2		2		2	2

Introduction to Django:Understanding web development frameworks,Introduction to Django and its features,Installing Django and setting up a development environment,Creating a simple Django project and app

Django Models and Database Integration:Creating models and defining database tables,Working with Django's Object-Relational Mapping (ORM),Performing database queries using Django's QuerySet API,Qw2003 Migrations and database schema evolution

Views and Templates:Building views to handle HTTP requests,Creating templates for dynamic HTML generation,Routing and URL patterns in Django.Passing data from views to templates

Django Forms:Creating HTML forms in Django,Form validation and handling form submissions,Customizing form behavior with Django form classes,Integrating forms with models

Django Admin Panel:Utilizing the Django admin interface for content management,Customizing the admin panel for specific models,Adding custom actions and filters

Authentication and Authorization:Implementing user authentication in Django,Managing user sessions and passwordsConfiguring permissions and authorization

Django REST Framework:Introduction to RESTful APIs,Building APIs with Django REST Framework,Serializers, views, and authentication for APIs,Consuming APIs in Django applications

Frontend Integration with Django:Integrating frontend frameworks (e.g., Bootstrap) with Django,Using static files and media in Django projects,AJAX and asynchronous behavior in Django applications

Testing and Debugging in Django:Writing unit tests for Django applications,Debugging techniques and tools,Best practices for testing in Django

Deployment and Scaling:Preparing a Django application for deployment,Choosing a hosting platform (e.g., Heroku, AWS),Configuring production settings,Scaling Django applications

Advanced Topics:Signals and event handling in Django,Building custom middleware,Caching strategies in Django,Internationalization and localization

Project Work:Applying knowledge to a real-world project,Working on a comprehensive Django project from start to finish,Code reviews and best practices

Reference Books:

1. Django for Beginners" by William S. Vincent
2. Django for APIs: Build web APIs with Python & Django" by William S. Vincent
3. Two Scoops of Django 3.x: Best Practices for the Django Web Framework" by Audrey Roy Greenfeld and Daniel Roy Greenfeld
4. Django 3 By Example" by Antonio Melé
5. High Performance Django" by Peter Baumgartner, Yann Malet, and Mark Lavin
6. Django Design Patterns and Best Practices" by Arun Ravindran

LLMs and Generative AI

Course Code: MSC15 **Course Credit:** 4

Course Outcome(CO):

CO1: Understanding LLMs: Students will gain an in-depth understanding of large language models, including their architecture, training methodologies, and applications in natural language processing (NLP).

CO2: Exploration of Generative AI: Students will explore generative AI techniques, including generative adversarial networks (GANs), variational autoencoders (VAEs), and transformer-based architectures such as GPT (Generative Pre-trained Transformer) models.

CO3: Applications in Text Generation: Students will learn how to use LLMs and generative AI techniques for text generation tasks, such as language modeling, text completion, and dialogue generation.

CO4: Understanding Ethical Considerations: Students will examine the ethical implications of LLMs and generative AI, including issues related to bias, fairness, misinformation, and the responsible use of AI-generated content.

CO5: Hands-on Experience: Students will gain practical experience through hands-on projects and exercises involving the implementation and fine-tuning of LLMs and generative AI models using frameworks such as TensorFlow, PyTorch, or Hugging Face's Transformers library.

CO6: Evaluation and Interpretation: Students will learn techniques for evaluating the performance of LLMs and generative AI models, as well as methods for interpreting their outputs and identifying areas for improvement.

CO7: Advanced Applications: Students will explore advanced applications of LLMs and generative AI in areas such as text summarization, question answering, machine translation, image generation, and creative content generation.

CO8: Research and Innovation: Students will be encouraged to explore current research trends and contribute to the advancement of LLMs and generative AI through research projects, papers, or thesis work.

CO9: Collaboration and Communication: Students will develop skills in collaborating with peers on AI projects, as well as effectively communicating their findings and insights to both technical and non-technical audiences.

CO10: Continuous Learning and Adaptation: Students will recognize the dynamic nature of LLMs and generative AI technologies and be prepared to adapt to new advancements and methodologies throughout their careers.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9

CO1	2	2	2		2		3		1
CO2		2		2				2	2
CO3	3	2			1	1		2	
CO4			1				2		2
CO5		2	1		3	2			
CO6	1		2	3		3		2	3
CO7				3					
CO8			2	3		3		3	3
CO9	1	2			2		2		3
CO10	3	3		2	2		2		3

Introduction to Large Language Models:Overview of natural language processing (NLP) and its applications.Evolution of language models: from n-grams to transformers.Introduction to transformer architecture and self-attention mechanisms.

Pretraining and Fine-tuning Language Models:Pretraining strategies: BERT, GPT, XLNet, etc.Fine-tuning language models for specific tasks: classification, summarization, translation.Hands-on lab: Fine-tuning a language model using Hugging Face Transformers.

Text Generation and Dialogue Systems:Generation techniques: autoregressive models, sampling strategies.Building conversational AI: chatbots, virtual assistants, dialogue generation.Ethical considerations in AI-generated text: bias, misinformation, and responsible AI.

Creative AI and Language Generation:Creative applications of language models: storytelling, poetry generation, art.Understanding style transfer and text-to-image generation.Design and implement a creative AI application using language models.

Advanced Topics in Generative AI:Recent advancements: GPT-3, BERT variants, T5, etc.Multi-modal generation: combining text with images and other modalities.Research trends and future directions in generative AI.

References Books:

- "Attention is All You Need" by Vaswani et al.

- "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" by Devlin et al.
- "Language Models are Few-Shot Learners" by Brown et al.
- "The Illustrated Transformer" by Jay Alammar.
- Various research papers and articles from leading conferences and journals in NLP and AI.

Seminar-I

Course Code: MSC16 **Course Credit:** 2

Course outcome(CO) :

CO1:Demonstrate a sound technical knowledge of their selected seminar topic.
CO2:Undertake problem identification, formulation and solution.
CO3:Design engineering solutions to complex problems utilising a systems approach.
CO4:Communicate with engineers and the community at large.
CO5:Demonstrate the knowledge, skills and attitudes of a professional engineer

In-depth Understanding: Acquire a comprehensive understanding of the chosen topic through

research, analysis, and critical evaluation of relevant literature, theories, and perspectives.

Effective Presentation Skills: Develop and refine skills in presenting complex ideas, findings, and

insights in a coherent, engaging, and structured manner suitable for a seminar audience.

Critical Thinking and Discussion: Foster an environment for critical thinking, intellectual discourse,

and constructive feedback among participants, promoting deeper insights and broader perspectives

on the topic.

Note :- Seminar related to any of their current subjects and latest technology in related subject (Case Study/Demo)

Interim Project

Course Code: MSC17 **Course Credit:** 4

Step 1: Data Processing

- Data Collection: Gather relevant datasets containing features and target variables for predictive modeling.
- Data Cleaning: Perform data cleaning tasks such as handling missing values, removing duplicates, and addressing outliers.
- Feature Engineering: Create new features or transform existing ones to improve model performance.
- Data Encoding: Convert categorical variables into numerical representations using techniques like one-hot encoding or label encoding.
- Data Splitting: Split the dataset into training, validation, and testing sets to facilitate model training and evaluation.

Step 2: Model Building

- Algorithm Selection: Choose appropriate machine learning algorithms based on the nature of the prediction task (e.g., regression, classification).
- Model Training: Train machine learning models on the training dataset using techniques like linear regression, decision trees, or ensemble methods.
- Cross-Validation: Implement k-fold cross-validation to assess model performance and ensure robustness.
- Hyperparameter Tuning: Fine-tune model hyperparameters using techniques such as grid search or random search to optimize performance.
- Model Evaluation: Evaluate model performance on the validation dataset using relevant metrics like mean squared error (MSE), accuracy, or F1-score.

Step 3: Model Fine-tuning

- Performance Analysis: Analyze model performance metrics and identify areas for improvement.
- Feature Importance: Assess the importance of different features in the model and consider feature selection or dimensionality reduction techniques.
- Ensemble Methods: Explore ensemble learning techniques like bagging, boosting, or stacking to combine multiple models for improved performance.
- Regularization: Apply regularization techniques like L1 or L2 regularization to prevent overfitting and improve generalization.

Step 4: Dashboarding and Representing Results

- Exporting Results: Export model predictions and performance metrics to a format compatible with visualization tools.

- Dashboard Creation: Design interactive dashboards using Tableau to visualize model outputs, insights, and performance metrics.
- Insights Generation: Extract meaningful insights from model results and communicate them effectively through visualization.

Step 5: Assessment and Monitoring

- Model Assessment: Assess the final model's performance on the test dataset to ensure it meets predefined accuracy or performance criteria.
- Monitoring: Implement monitoring mechanisms to track model performance over time and detect any degradation in predictive accuracy.
- Feedback Loop: Establish a feedback loop to collect user feedback and incorporate it into model updates and improvements.

Lab Practice-3

Course Code: MSC18 **Course Credit:** 4

1. Big Data and Data Clouds:
    - Setting up a Hadoop cluster for distributed computing.
    - Working with Hadoop Distributed File System (HDFS) for storage.
    - MapReduce programming paradigm for data processing.
    - Implementing big data analytics using Apache Spark.
    - Hands-on exercises on processing large datasets with Hive and Pig.
    - Introduction to cloud-based data storage solutions like Amazon S3 and Google Cloud Storage.
2. Web Framework:
    - Introduction to web frameworks such as Django, Flask, and Ruby on Rails.
    - Building web applications with Django: setting up routes, views, and templates.
    - Implementing RESTful APIs with Flask for web services.
    - Integrating databases with web applications using Object-Relational Mapping (ORM).
    - Front-end development with HTML, CSS, and JavaScript frameworks like React or Angular.
    - Deploying web applications on cloud platforms like Heroku or AWS Elastic Beanstalk.
3. Large Language Models (LLMs) and Generative AI:
    - Understanding the architecture and working principles of large language models like GPT (Generative Pre-trained Transformer).

- Hands-on exploration of pre-trained language models using Hugging Face's Transformers library.
- Fine-tuning language models for specific tasks such as text generation, sentiment analysis, or summarization.
- Implementing conversational AI agents using LLMs for chatbots or virtual assistants.
- Ethical considerations and biases in language models: discussion and case studies.
- Advanced topics in generative AI: conditional text generation, style transfer, and multimodal models.

Data Mining and Business Intelligence

Course Code: MSC19 **Course Credit:** 4

Course Outcome(CO):

CO1: Understanding Data Mining Concepts: Students will gain a solid understanding of fundamental concepts, techniques, and algorithms used in data mining, including association rule mining, clustering, classification, and anomaly detection.

CO2: Data Preprocessing and Cleaning: Students will learn techniques for preprocessing and cleaning data to prepare it for data mining tasks, including handling missing values, dealing with outliers, and transforming variables.

CO3:Exploratory Data Analysis: Students will be able to perform exploratory data analysis to gain insights into data distributions, patterns, and relationships, using visualization techniques and statistical methods.

CO4: Data Mining Algorithms and Models: Students will become familiar with a variety of data mining algorithms and models, such as decision trees, k-nearest neighbors, support vector machines, and neural networks, and understand their strengths, weaknesses, and applications.

CO5: Model Evaluation and Validation: Students will learn how to evaluate and validate data mining models using appropriate metrics and techniques, including cross-validation, confusion matrices, ROC curves, and precision-recall curves.

CO6: Feature Selection and Engineering: Students will acquire skills in feature selection and engineering techniques to identify and extract relevant features from data, reduce dimensionality, and improve model performance.

CO7: Business Intelligence Tools and Techniques: Students will be introduced to business intelligence tools and techniques for analyzing and visualizing data, generating reports and dashboards, and making data-driven decisions in business contexts.

CO8: Real-World Applications: Students will explore real-world applications of data mining and business intelligence across various domains, including marketing, finance, healthcare, and e-commerce, and understand how these techniques are used to solve practical problems.

CO9: Ethical and Legal Considerations: Students will understand the ethical and legal considerations related to data mining and business intelligence, including privacy, confidentiality, bias, and compliance with regulations such as GDPR.

CO10: Project Work: Students will have the opportunity to apply their knowledge and skills to a hands-on project, where they will design and implement a data mining solution to address a real-world business problem, from data collection and preprocessing to model deployment and evaluation.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	2	3	3		2	3	3		1
CO2		2		2				2	2
CO3	3	2			1	1	3	2	3
CO4			1				2		2
CO5		2	1		3	2			
CO6	1		2	3		3		2	3
CO7	2	2		3		3	3		
CO8			2	3		3		3	3
CO9	1	2			2		2		3
CO10	3	3		2	2	3	2	3	3

Introduction to Data Mining and Business Intelligence:Overview of data mining concepts, process, and methodologies.Role of business intelligence in decision-making and strategic planning.Introduction to data mining tools and software.

Data Preprocessing and Cleaning:Data cleaning techniques: handling missing values, outliers, and noise.Data transformation and normalization.Feature engineering and selection.

Supervised Learning Techniques:Classification algorithms: decision trees, logistic regression, support vector machines.Evaluation metrics for classification models: accuracy, precision, recall, F1-score.Hands-on lab: Building a classification model using scikit-learn.

Unsupervised Learning Techniques:Clustering algorithms: k-means, hierarchical clustering, DBSCAN.Association rule mining: Apriori algorithm.Anomaly detection methods.Case study: Customer segmentation and market basket analysis.

Text Mining and Sentiment Analysis:Text preprocessing techniques: tokenization, stemming, lemmatization.Sentiment analysis using machine learning and deep learning approaches.Applications of text mining in social media analytics and customer feedback analysis.

Advanced Topics in Data Mining:Ensemble learning techniques: random forests, gradient boosting.Feature importance and model interpretation.Time-series analysis and forecasting.Case study: Predictive maintenance in manufacturing.

Business Intelligence and Decision Support Systems:Introduction to business intelligence platforms and tools.Data visualization techniques for effective communication of insights.Dashboard design principles and best practices.Project: Designing a business intelligence dashboard.

Ethics, Privacy, and Security in Data Mining:Ethical considerations in data collection, analysis, and decision-making.Privacy-preserving data mining techniques.Security challenges and risk mitigation strategies.

References Books:

- "Data Mining: Concepts and Techniques" by Jiawei Han and Micheline Kamber.
- "Introduction to Business Intelligence" by Ronald Ramalho.
- "Python Data Science Handbook" by Jake VanderPlas.
- Various research papers and articles from leading conferences and journals in data mining and business intelligence.

Open Electives

Course Code: MSC20 **Course Credit:** 2

1. Time Series Analysis and Forecasting Techniques

Course Outcome(CO):

CO1: To understand the fundamentals of time series data and its characteristics.
CO2: To explore different time series decomposition methods and trend analysis techniques.
CO3: To learn how to model and forecast time series data using statistical methods.
CO4: To apply machine learning algorithms for time series forecasting and evaluate model performance.
CO5:To develop practical skills for time series analysis and forecasting using software tools and libraries.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	2	3	3		2	3	3		1
CO2		2		2				2	2
CO3	3	2			1	1	3	2	3
CO4			1				2		2
CO5		2	1		3	2			

Introduction to Time Series Data:Definition and characteristics of time series data,Importance of time series analysis in data science and business applications

Time Series Decomposition:Trend, seasonality, and noise components,Decomposition methods: additive vs. multiplicative decomposition

Exploratory Data Analysis for Time Series:Visualization techniques: line plots, scatter plots, histograms,Autocorrelation and partial autocorrelation analysis,Stationarity tests and transforming non-stationary data

Statistical Forecasting Methods:Moving average, exponential smoothing, and Holt-Winters methods,Autoregressive Integrated Moving Average (ARIMA) models,Seasonal ARIMA (SARIMA) models and parameter selection

Machine Learning for Time Series Forecasting:Supervised learning algorithms: Linear regression, Decision trees, Random forests,Time series regression and feature engineering,Ensemble methods and model selection for time series forecasting

Deep Learning for Time Series Forecasting:Introduction to recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) networks,Time series forecasting with RNNs and LSTMs,Hyperparameter tuning and optimization for deep learning models
Model Evaluation and Validation:Performance metrics for time series forecasting: MAE, RMSE, MAPE, etc.Cross-validation techniques: K-fold cross-validation, time series split,Hyperparameter tuning and model selection strategies
Practical Applications and Case Studies:Real-world applications of time series analysis and forecasting,Case studies from finance, economics, healthcare, and other domains,Hands-on projects with time series datasets using Python libraries (e.g., pandas, statsmodels, scikit-learn, TensorFlow)

Textbooks:

1. "Forecasting: Principles and Practice" by Rob J Hyndman and George Athanasopoulos
2. "Time Series Analysis and Its Applications: With R Examples" by Robert H. Shumway and David S. Stoffer
3. "Deep Learning for Time Series Forecasting" by Dr. N D Lewis

2. Marketing Analytics using AI

Course Outcome(CO):

CO1: To understand the role of data science and AI in marketing analytics.
CO2: To explore various techniques for analyzing customer behavior and preferences.
CO3: To learn how to build predictive models for marketing outcomes using AI and machine learning algorithms.
CO4: To apply AI-driven approaches to segment customers, personalize marketing campaigns, and optimize marketing ROI.
CO5: To develop practical skills for marketing analytics using AI tools and platforms.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	2	3	3		2	3	3		1
CO2		2		2				2	2
CO3	3	2			1	1	3	2	3
CO4			1				2		2
CO5		2	1		3	2			

Introduction to Marketing Analytics and AI:Overview of marketing analytics and its importance in business decision-making,Role of AI and machine learning in marketing analytics,Key concepts and terminology in marketing analytics using AI

Customer Segmentation and Profiling:Traditional segmentation methods vs. AI-driven segmentation techniques,Clustering algorithms: K-means, hierarchical clustering, DBSCAN,Customer profiling and persona development using machine learning

Predictive Modeling for Customer Behavior:Regression analysis for predicting customer purchase behavior,Classification algorithms: Logistic regression, decision trees, random forests,Churn prediction and customer lifetime value (CLV) modeling

Recommendation Systems and Personalization:Collaborative filtering and content-based recommendation algorithms,Building recommendation engines for personalized marketing,A/B testing and multivariate testing for evaluating recommendation systems

Natural Language Processing (NLP) for Marketing:Sentiment analysis of customer reviews and social media data:Topic modeling and text classification for marketing content analysis,Text generation and chatbot development for customer interaction

Optimizing Marketing Campaigns with AI:Marketing mix modeling and attribution analysis,Predictive analytics for campaign optimization and resource allocation,Real-time bidding and programmatic advertising using AI algorithms

Customer Journey Analytics:Mapping the customer journey using AI-driven analytics,Conversion funnel analysis and path analysis techniques,Predicting customer touchpoints and optimizing marketing channels

Ethical and Responsible AI in Marketing:Considerations for ethical AI use in marketing analytics,Bias and fairness in AI models for marketing,Privacy and data protection regulations in marketing analytics

Textbooks:

1. "Marketing Analytics: Data-Driven Techniques with Microsoft Excel" by Wayne L. Winston
2. "Predictive Analytics for Marketers: Using Data Mining for Business Advantage" by Barry Leventhal and James Fogarty
3. "AI for Marketing and Product Innovation: Powerful New Tools for Predicting Trends, Connecting with Customers, and Closing Sales" by A. K. Pradeep

3. Data Analysis in Economics & Financial Decision Making

Course Outcome(CO):

CO1: To understand the role of data analysis in economics and financial decision making.
CO2: To learn how to collect, preprocess, and visualize economic and financial data.
CO3: To explore statistical methods for analyzing economic indicators and financial metrics.
CO4: To apply machine learning algorithms for forecasting economic trends and financial outcomes.
CO5: To develop practical skills for analyzing and interpreting data in economic and financial contexts.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	2	3	3		2	3	3		1
CO2		2		2				2	2
CO3	3	2			1	1	3	2	3
CO4			1				2		2
CO5		2	1		3	2			

Introduction to Data Analysis in Economics and Finance:Role of data analysis in economics and financial decision making,Types of economic and financial data: time series, cross-sectional, panel data,Data sources and challenges in collecting economic and financial data

Data Collection and Preprocessing,Data collection methods: surveys, government reports, financial statements, web scraping,Data cleaning and preprocessing techniques: handling missing values, outliers, and duplicates,Exploratory data analysis (EDA) for economic and financial datasets

Descriptive Statistics and Visualization:Summary statistics: mean, median, variance, standard deviation,Data visualization techniques: histograms, box plots, scatter plots, time series plots,Correlation and covariance analysis for economic and financial variables

Statistical Analysis of Economic Indicators:Inflation rate, unemployment rate, GDP growth, interest rates,Time series analysis techniques: trend analysis, seasonal decomposition, forecasting methods,Cross-sectional analysis: regression analysis, hypothesis testing, economic modeling

Financial Metrics and Performance Analysis:Financial statement analysis: balance sheet, income statement, cash flow statement,Financial ratios and metrics: profitability ratios, liquidity ratios, solvency ratios,Performance evaluation of financial assets: stocks, bonds, derivatives

Machine Learning for Financial Forecasting:Introduction to machine learning algorithms: linear regression, decision trees, random forests,Predictive modeling for financial markets: stock price prediction, portfolio optimization,Sentiment analysis and natural language processing (NLP) for financial news and social media data

Risk Management and Portfolio Optimization:Risk assessment techniques: value-at-risk (VaR), conditional value-at-risk (CVaR),Portfolio theory and diversification strategies,Monte Carlo simulation for financial risk analysis

Ethical and Regulatory Considerations:Ethical issues in data analysis and decision making,Regulatory frameworks: GDPR, HIPAA, SEC regulations,Responsible use of data in economic and financial contexts

Textbooks:

1. "Data Analysis for Economics and Finance" by Daniel S. Wilks
2. "Statistics for Business and Economics" by Paul Newbold, William L. Carlson, and Betty Thorne
3. "Machine Learning for Financial Engineering" by Marcos López de Prado

4. **Data Analytics in IOT**

Course Outcome(CO):

CO1:To understand the concepts and architecture of the Internet of Things (IoT).
CO2:To learn about different types of IoT sensors and data collection methods.
CO3: To explore data preprocessing techniques for handling IoT data.
CO4: To apply statistical and machine learning algorithms for IoT data analytics.
CO5:To develop skills in visualizing and interpreting insights from IoT data.
CO6: To understand the challenges and ethical considerations in IoT data analytics.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	2	3	3		2	3	3		1
CO2		2		2				2	2
CO3	3	2			1	1	3	2	3
CO4			1				2		2
CO5		2	1		3	2			
CO6	3		2	2	1	2			2

Introduction to IoT and Data Analytics:Overview of IoT ecosystem and components,Role of data analytics in IoT applications,Challenges and opportunities in IoT data analytics

IoT Data Collection and Preprocessing:Types of IoT sensors and data sources,Data acquisition and transmission protocols (e.g., MQTT, CoAP),Data preprocessing techniques for handling noisy and missing data

Exploratory Data Analysis for IoT:Descriptive statistics and visualization methods for IoT data,Time series analysis and trend detection,Correlation and causality analysis in IoT datasets

Statistical Methods for IoT Data Analytics:Hypothesis testing and confidence intervals,Regression analysis for predicting IoT sensor readings,Anomaly detection and outlier detection techniques

Machine Learning for IoT Data Analytics:Supervised learning algorithms for classification and regression,Unsupervised learning techniques for clustering and anomaly detection,Deep learning approaches for IoT data analysis (e.g., CNNs, RNNs)

Real-time Analytics and Stream Processing:Introduction to stream processing frameworks (e.g., Apache Kafka, Apache Flink),Real-time analytics for IoT applications,Windowing and aggregation techniques for stream processing

Visualization and Interpretation of IoT Data:Dashboard design principles for IoT analytics,Interactive visualization tools for exploring IoT datasets,Interpretation of insights and actionable intelligence from IoT data

Ethical and Privacy Considerations in IoT Data Analytics:Data security and privacy challenges in IoT deployments,Regulatory compliance and data governance frameworks,Ethical implications of IoT data collection and analysis

Textbooks:

1. "Internet of Things: Principles and Paradigms" by Rajkumar Buyya et al.
2. "Data Science for IoT: Dummies" by John Paul Mueller and Luca Massaron
3. "Practical Internet of Things Security" by Brian Russell et al.

5. Intellectual Property Rights

Course Outcome(CO):

CO1:Understanding of Intellectual Property (IP) Concepts: Students will gain a comprehensive understanding of various forms of intellectual property, including patents, trademarks, copyrights, and trade secrets, and their relevance to data science and technology innovation.

CO2: Legal Framework and Regulations: Students will become familiar with the legal framework governing intellectual property rights at national and international levels, including relevant laws, regulations, treaties, and conventions.

CO3:Protection and Enforcement Mechanisms: Students will learn about the mechanisms and procedures for protecting intellectual property rights, such as registration, licensing, enforcement, and litigation, and their implications for data-driven innovations and technologies.

CO4: Data Privacy and Confidentiality: Students will explore the intersection of intellectual property rights with data privacy and confidentiality laws, regulations, and best practices, with a focus on safeguarding sensitive data and proprietary information.

CO5: Intellectual Property in Data Science Projects: Students will analyze case studies and real-world examples to understand how intellectual property rights apply to data science projects, including issues related to data ownership, access rights, data sharing, and data reuse.

CO6:Ethical and Social Implications: Students will critically examine the ethical and social implications of intellectual property rights in data science, including considerations of fairness, equity, transparency, and accountability in the use and dissemination of data and algorithms.

CO7:Intellectual Property Strategies: Students will develop strategies for effectively managing intellectual property assets in data science projects and organizations, including IP portfolio management, IP valuation, technology transfer, and commercialization.

CO8:Emerging Trends and Challenges: Students will explore emerging trends and challenges in intellectual property law and policy, such as the impact of artificial intelligence, machine learning, open source software, and big data analytics on intellectual property rights and innovation ecosystems.

CO9: Intellectual Property in Research and Innovation: Students will understand the role of intellectual property rights in fostering research and innovation in data science and related fields, including their influence on funding, collaboration, and technology commercialization initiatives.

CO10: Critical Analysis and Policy Recommendations: Students will engage in critical analysis of intellectual property laws, policies, and practices in the context of data science and technology, and develop informed policy recommendations to address current and future challenges.

CO-PO Mapping:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	2	2	2		2		3		1
CO2		2		2				2	2
CO3	3	2			1	1		2	

CO4			1				2		2
CO5		2	1		3	2			
CO6	1		2	3		3		2	3
CO7				3					
CO8			2	3		3		3	3
CO9	1	2			2		2		3
CO10	3	3		2	2		2		3

Introduction to Intellectual Property Rights:Definition and types of intellectual property,Historical evolution and significance of intellectual property rights.

Legal Framework for Intellectual Property:National and international laws, treaties, and conventions,Patent, trademark, copyright, and trade secret laws

Patents and Data Science:Patentability criteria and requirements,Patent protection for algorithms, software, and data-driven inventions
Copyrights and Data Analytics:Copyright protection for databases, software code, and digital content,Fair use, licensing, and open access in data science projects

Trademarks and Branding,Trademark registration and enforcement,Branding strategies and trademark infringement issues in data-driven marketing

Trade Secrets and Confidential Information:Protection of trade secrets and confidential information,Non-disclosure agreements and confidentiality clauses in data science contracts

Intellectual Property Management:Intellectual property audits and due diligence,Technology transfer, licensing, and commercialization strategies
Ethical and Social Implications of Intellectual Property:Equity, access, and diversity in intellectual property rights:Ethical considerations in data sharing, privacy, and data ownership

Emerging Issues in Intellectual Property and Data Science:Open source software and intellectual property implications,Artificial intelligence, machine learning, and IP challenges

Case Studies and Best Practices:Analysis of real-world cases involving intellectual property issues in data science,Best practices for managing intellectual property in data science projects and organizations

Textbooks:

1. "Intellectual Property Law for Engineers and Scientists" by Howard B. Rockman
2. "Understanding Intellectual Property Law" by Donald S. Chisum and Tyler T. Ochoa
3. "Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World" by Bruce Schneier

Seminar-II

Course Code: MSC21 **Course Credit:** 2

Course outcomes :

● Comprehensive Understanding: Gain an in-depth comprehension of the selected research paper subject, exploring its key concepts, methodologies, and findings. Develop critical thinking skills to assess the paper's significance within the broader academic context.

● Effective Communication: Enhance communication skills through seminar participation, fostering the ability to articulate complex ideas succinctly. Practice presenting and discussing research findings, promoting clarity and engagement with peers.

● Research Application: Apply acquired knowledge to formulate informed opinions and contribute to scholarly discussions. Develop the capacity to apply research methodologies in related areas, fostering a foundation for future academic pursuits.

Course Contents:

Seminar related to any Published Research Paper subjects