AC –07/07/2023
Item No. –6.28(N)

As Per NEP 2020

University of Mumbai

Title of the program

A) P.G. Diploma in Data Science– **2023-24**
B) M.Sc. (Data Science) (Two Year) – **2023-24**
C) M.Sc. (Data Science) (One Year) – **2027-28**

Syllabus for

Semester - Sem I & II

Ref: GR dated 16th May, 2023 for Credit Structure of PG

(With effect from the academic year 2023-24)

University of Mumbai

(As per NEP 2020)

Sr. No.	Heading	Particulars		
1	**Title of program:** O.SP-39A	**A**	P.G. Diploma in Data Science	
	O. SP-39B	**B**	M.Sc. (Data Science)	(Two Year)
	O. SP-39C	**C**	M.Sc. (Data Science)	(One Year)
2	**Eligibility** O. SP- 40A	**A**	A candidate with minimum 50% score in Graduation can appear for entrance examination through which the admission merit list will be generated.	
	O. SP-40B	**B**	A candidate with minimum 50% score in Graduation can appear for entrance examination through which the admission merit list will be generated.	
	O. SP-40C	**C**	Graduate with 4 year UG Degree (Honours/Honours with Research) with specialization in concerned subject or equivalent academic level 6.0 OR Graduate with 4 years UG Degree program with maximum credits required for award of Minor degree is allowed to take up the Post Graduate program in Minor subject provided the student has acquired the required number of credits as prescribed by the concerned Board of Studies.	
3	**Duration of program** R.SP-96	**A**	1 Year	
		B	2 Years	
		C	1 Year	
4	**Intake Capacity** R. SP-97	120		
5	**Scheme of Examination** R. SP-98	NEP 50% Internal & 50% External, Semester End Examination Individual Passing in Internal and External Examination.		

6	**Standard of Passing:** R. SP-99	40%	
7	**Credit Structure:** R. SP-100A R. SP-100B	Attached herewith	
8	**Semesters**	**A**	Sem - I & II
		B	Sem - I, II, III & IV
		C	Sem - I & II
9	**Program Academic Level**	**A**	6.0
		B	6.5
		C	6.5
10	**Pattern**	Semester	
11	**Status**	New	
12	**To be implemented from Academic Year Progressively**	**A**	2023-24
		B	
		C	2027-28

DrJyotshnaDongardive
Head, Department of Computer Science
University of Mumbai

DrShivramGarje
Dean, Science and Technology
University of Mumbai

PREAMBLE

1. Introduction

The Master of Science (M.Sc.) in Data Science is a graduate-level program that focuses on the study and application of data analysis, data mining, machine learning, and statistical modeling techniques to extract valuable insights and knowledge from large and complex datasets. This program is designed to equip students with the necessary skills and knowledge to tackle the challenges and opportunities in the rapidly growing field of data science.

The curriculum of a M.Sc. in Data Science program usually includes a mix of core courses and elective courses. Core courses provide a solid foundation in fundamental concepts and techniques of data science, such as programming languages (e.g., Python, R), Essential technologies for data science,Data visualization, Statistical analysis, Natural Language processing,Soft computing, Time series analysis and Forecasting and Deep Neural Network. Elective courses allow students to specialize in specific areas of interest, such as, Health care analytics, Sports analytics, Legal analytics, Human Resource and Retail marketing analytics, Block chain Technologies, Image and video analytics and Social media analytics.

In addition, MSc in Data Science programs also include On Job Training (OJT) and hands-on projects. This intensive OJT training, equivalent to a full course, provides invaluable exposure to real-world scenarios within IT or IT-related organizations. By applying their theoretical knowledge in practical settings, students gain firsthand experience and develop the necessary skills to thrive in the professional world.The projects help students gain practical experience in applying their knowledge to solve complex data challenges and provide valuable insights to organizations.The inclusion of a Research Methodology Course helps students develop a strong research attitude, enabling them to contribute meaningfully to the advancement of Data Science.

Career prospects for graduates of MSc in Data Science programs are quite promising, as data science skills are in high demand across various industries. Graduates can pursue careers as data scientists, data analysts, machine learning engineers, data engineers, or data consultants in sectors such as finance, healthcare, e-commerce, marketing, and technology. They may also find opportunities in research institutions or pursue further studies at the doctoral level.

Overall, an MSc in Data Science program offers a comprehensive and rigorous education in the field of data science, equipping students with the skills and knowledge necessary to succeed in a data-driven world.

The curriculum's continuous refinement has been made possible through the valuable inputs, suggestions, and observations of colleagues at the University of Mumbai, experts from premier institutions, and industry professionals. We extend our gratitude to all those who have directly or indirectly contributed to the development of this program

2. Aims and Objectives

The Aim of M.Sc.(Data Science) program is to equip students with the knowledge, skills, and mindset required to excel in the field of data science, while also emphasizing ethical practices, teamwork, and effective communication.

Objectives:

- **Provide In-depth Knowledge**: The program aims to provide students with a comprehensive understanding of the key concepts, theories, and methodologies in data science. It covers a range of topics including statistics, machine learning, data mining, data visualization, and data management, enabling students to develop a deep knowledge base in the field.

- **Develop Analytical and Technical Skills**: The program aims to equip students with the analytical and technical skills necessary to analyze and interpret large and complex datasets. Students learn how to apply statistical methods, implement machine learning algorithms, and utilize data visualization techniques to extract valuable insights and make data-driven decisions.

- **Foster Problem-solving Abilities**: The program aims to enhance students' problem-solving abilities by training them to approach real-world data challenges critically and creatively. Students learn to identify problems, design appropriate data analysis strategies, and develop innovative solutions using data science techniques.

- **Encourage Collaboration and Teamwork**: The program aims to foster collaboration and teamwork skills among students. Data science projects often require interdisciplinary collaboration, where individuals with diverse skills work together to solve complex problems. Students learn to effectively communicate, collaborate, and contribute as part of a team.

- **Foster Industry Relevance:** The program aims to stay up-to-date with industry trends and technologies to ensure graduates are well-prepared for the demands of the data science job market. It may include guest lectures, industry collaborations, or internship opportunities to provide students with practical exposure and relevant industry experience.

- **Professional Development:** The program aims to prepare students for successful careers in the field of data science. This includes developing professional skills such as teamwork, project management, and leadership.

3. Learning Outcomes

- **Knowledge of Data Science Concepts:** Students will acquire a comprehensive understanding of the fundamental concepts, theories, and principles of data science. This includes knowledge of statistical analysis, machine learning algorithms, data visualization techniques, and data management strategies.

- **Proficiency in Data Analysis Techniques:** Students will develop proficiency in applying various data analysis techniques to extract insights from complex datasets. They will learn how to clean and preprocess data, perform exploratory data analysis, implement statistical modeling, and utilize machine learning algorithms for predictive and descriptive analytics.

- **Technical Skills in Data Science Tools and Programming Languages:** Students will gain proficiency in using data science tools and programming languages commonly used in the field, such as Python or R. They will learn how to leverage libraries and frameworks for data manipulation, visualization, and modeling.

- **Ability to Apply Analytical Methods:** Students will develop the ability to apply analytical methods to solve real-world data problems. They will learn how to choose appropriate statistical tests, evaluate model performance, and make data-driven decisions based on the results of their analysis.

- **Ethical Considerations in Data Science:** Students will gain an understanding of ethical considerations and responsibilities in data science. They will learn about privacy concerns, data security, and ethical guidelines for handling and analyzing data. They will be able to make ethical decisions in data collection, analysis, and reporting.

- **Collaboration and Teamwork:** Students will develop the ability to work collaboratively as part of a team on data science projects. They will learn to effectively communicate and share ideas, contribute their expertise to team efforts, and collaborate with diverse stakeholders to solve complex data challenges.

4. Credit Structure of the Program (Sem I, II, III & IV)

R. SP-100A

Year	Level	Sem	Major				RM	OJT/FP	RP	Cum.Cr.	Degree
			Mandatory			Electives					
1	6.0	Sem I	2*4+2*2+2			4	4	-	-	22	PG Diploma (after 3 Years Degree)
			Essential Technologies for Data Science (501)	TH	4	SPARK Technologies (506a) 2 TH+2PR **(OR)** Retail Marketing Analytics (506b) 2 TH + 2 PR **(OR)** Sports Data Analytics(506c) 2 TH +2PR	Research Methodology (510)				
			Essential Technologies for Data Science Practical(502)	PR	2						
			Data Analysis and Visualization(503)	TH	4						
			Data Analysis and Visualization Practical(504)	PR	2						
			Statistical Methods for Data Science(505)	TH	2						
		Sem II	2*4+2*2+2			4	-	OJT(516) 4	-	22	
			Soft Computing(511)	TH	4	Human Resource Analytics(516a) 2 TH +2PR **(OR)** Public Healthcare Analytics(516b) 2 TH +2PR **(OR)** Social Media Analytics(516c) 2 TH +2PR					
			Soft Computing Practical(512)	PR	2						
			Time Series Analysis and Forecasting(513)	TH	4						
			Time Series Analysis and Forecasting Practical (514)	PR	2						
			Ethical Issues in Data Science(515)	TH	2						

Cum.Cr.For PG Diploma	28	8	4	4		44
Exit Option: PG Diploma(44credits) after Three Year UG Degree						

R. SP-100B

Year	Level	Sem (2yr)	Major				RM	OJT/FP	RP	Cum.Cr.	Degree
			Mandatory			Electives					
2	6.5	Sem III	2*4+2*2+2			4	-	-	RP (607) 4	22	PG Degree after 3-yr UG
			Next Generation Databases (601)	TH	4	Image and Video Analytics (606a) 2TH +2PR **(OR)** Econometrics and Finance (606b) 2TH +2PR **(OR)** Data Science for Agriculture (606c) 2TH+2PR					
			Next Generation Databases Practical (602)	PR	2						
			Text Mining and Natural Langauge Processing(603)	TH	4						
			Text Mining and Natural Langauge Processing Practical (604)	PR	2						
			Data Compliance and Auditing(605)	TH	2						
		Sem IV	2*4+2*2			4	-	-	RP (616) 6	22	
			Deep Neural Networks (611)	TH	4	Blockchain Technologies for Data Science (615a) 2TH+2PR **(OR)** Financial Risk Analytics and Management					
			Deep Neural Networks Practical (612)	PR	2						
			Optimization Methods for Data Science (613)	TH	4						
			Optimization Methods	PR	2						

			for Data Science Practical (614)			(615b) 2TH+2PR **(OR)** Legal Analytics (615c)2TH +2PR					
Cum.Cr. For 1 Yr PG Degree			26			8			10	44	
Cum.Cr. For 2 Yr PG Degree			54			16	4	4	10	88	

DrJyotshnaDongardive
Head, Department of Computer Science
University of Mumbai

DrShivramGarje
Dean, Science and Technology
University of Mumbai

Sem-I

SEMESTER-I

Programme Name: M.Sc. Data Science (Semester-I)	Course Name: Essential technologies for Data science
Total Credits: 04	Total Marks: 100
University assessment: 50	College assessment: 50

Prerequisite:
Basic mathematics and problem-solving skills

Course outcomes:

- Students should be able to understand the core concepts of programming before starting to write new programs. Students should be able to develop logic for Problem Solving.
- To manage data by relying on data structures such as strings, arrays, files, lists, and dictionaries. Exemplify the various levels of decision making on a program and implement a mix of loops, functions, and control flow to extract information from a data structure.
- Student should able to learn different programming techniques and tools related to data science.

Course Code	Course Title	Total Credits
PSDS501	Essential technologies for Data science	04
MODULE - I **Unit 1: Introduction to Data science and Python** a) Introduction to Data Science, data science life cycle, Applications, and advantages of Python over other programming languages b) What is Python? Why Should I learn Python? Installing Python How to execute Python program Writing your first program. c) Basic programming elements of Python-variables and constants, identifiers, Typecasting or Type Conversion in Python, indentation, comments, rules of writing identifiers, primitive data types, writing command line programs in python d) Operators in Python: Arithmetic operators, relational operators, Logical operators, Membership operators, Taking user input. **Unit 2: Data structures and control flow** a) Collection data structures in Python- List, tuples, dictionary, sets and strings b) Control flow- Sequential, Branching or Conditional, Iteration or Repetition, Modular or Subroutines Conditional and iteration statements: if elif else statements, loops, for loop and while loops		02

c) c) User defined functions in Python- No Value Pass and No Return, Value Pass and No Return, Value Pass and Return, Function with default arguments, Function with variable arguments, Higher order functions, list comprehension	
MODULE - II **Unit 3:Statistics for Data Analysts** a) Permutations and combinations, probability, Descriptive statistics (mean, median, mode), point estimation, quartiles and boxplot, methods of dispersion, random variables and probability distribution b) Measures of shape- skewness, kurtosis, outlier detection, transformation (log, square root) c) Inferential statistics- Sampling techniques, Hypothesis testing, Z-score normalization, correlation, ANOVA d) Introduction to NumPy, creating NumPy arrays, indexing and slicing, vectorization, Boolean indexing, transformation, inferential statistics using NumPy **Unit 4: Data wrangling using Pandas.** a) Introduction to data: NOIR (nominal, Ordinal, Interval and Ratio), continuous and discrete numeric data. Types of data analysis (descriptive, diagnostic, predictive and prescriptive analysis) b) Data wrangling using Pandas - Creating Series, Creating Data frame from dictionary, attributes, and method description of a data frame. Drop columns, add columns, add rows, iloc , loc, indexing and slicing data frames, selection with condition, group by summary operation, sorting operations c) Introduction to R IDE- components of R IDE, Basic data types in R, Data structures in R, data coercion, importing files, visualisation using ggplot2. d) Basic visualisation using matplotlib- Components of a chart, line chart, scatter chart, pie chart, sub plots.	**02**

Reference Books:

1. Data Analysis with Pandas and Python by Boris Paskhaver, Manning Publications. Available at: https://www.perlego.com/book/2881120/pandas-in-action-pd
2. Practical Statistics for Data Scientists: 50 Essential Concepts by Peter Bruce, Andrew Bruce, Peter Gedeck, O'Reilly Media, 2017 ISBN-10: 1491952962 ISBN-13: 978-1491952962
3. Foundations of Statistics for Data Scientists With R and Python By Alan Agresti, Maria Kateri, CRC Press Taylor and Francis group, 2022

Programme Name: M.Sc. Data Science (Semester-I)	Course Name: Essential technologies for Data Science Practical
Total Credits: 02	Total Marks: 50
University assessment: 50	

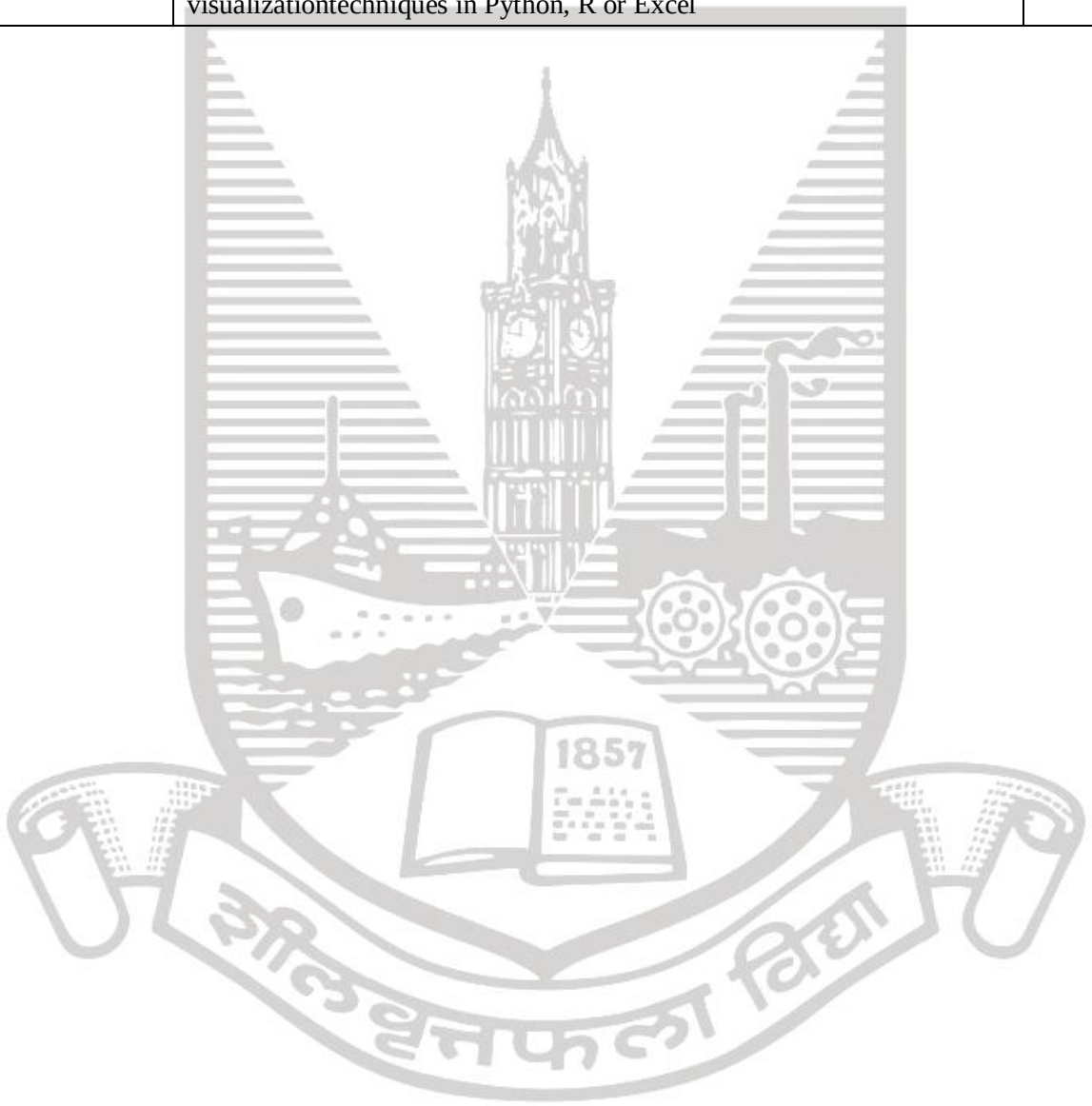
Prerequisite:
Basic computer skills and Basic understanding of elementary Math.

Course outcomes:

- Should be able to write basic programming in Python.
- Should be able to use Python data structures and able to use conditional and iterativecontrol flow.
- Should be able to demonstrate descriptive, diagnostic, and inferential statistics using, Python, R or Excel (use Data analysis tool pack in Excel or Data analyzer tool in Microsoft office 365)
- Perform basic data wrangling using R or Pandas
- Perform data visualization using R or Pandas

Course Code	Course Title	Total Credits
PSDSP502	Essential technologies for Data Science Practical	02
1	Write a Python program to accept inputs from users and perform arithmetic operations.	
2	Write a program to demonstrate relational and logical operators in Python.	
3	Write a Python program to demonstrate usage of loops. Use both for and while loops to distinguish between them. [e.g., Reversing the digits of a number without converting to String]	
4	Demonstrate the use of data structures list, sets, dictionary.	
5	Import a dataset and perform univariate analysis on the numeric columns to analyze the shape of the data. Write inference of the output. [Python or R or Excel]	
6	Demonstrate Hypothesis testing, and ANOVA using a dataset [use Python, R or Excel]	
7	Demonstrate correlation analysis. Use heatmap for visualization. Writeinferences.	

8	Import a csv or Excel dataset and demonstrate data wrangling, viewshape, dimension, column names of the dataset, ways to select data usingcolumn number, column names, simple and compound conditionalselection, update and modify dataset.	
9	Demonstrate group by summary operations and sorting techniques.	
10	Perform univariate, bivariate and multivariate analysis using visualizationtechniques in Python, R or Excel	



Programme Name: M.Sc. Data Science (Semester-I)	Course Name: Data Analysis and Visualization
Total Credits: 04	Total Marks: 100
University assessment: 50	College assessment: 50

Prerequisite:
Basic understanding of data, types of data and visualization

Course outcomes:

- Understand various data formats, sources and storage mechanisms.
- Handle missing data and manage data wrangling and manipulation
- Create data visualization and report making using various software tools
- Demonstrate the visualizations and make interpretations
- Create a data story using various software tools.

Course Code	Course Title	Total Credits
PSDS503	Data Analysis and Visualization	04
MODULE - I **Unit 1: Introduction to Data Analysis** Data Analysis - Exploratory Data Analysis and Data Science Process - Responsibilities of a Data Analyst - Data Analytics vs. Data Analysis - Types of Data - Understanding Different Types of File Formats - Sources of Data - Languages for Data Professionals - Overview of Data Repositories - Data Marts, Data Lakes, ETL, and Data Pipelines - Foundations of Big Data - Identifying Data for Analysis **Unit2: Data Wrangling** Data Sources - How to gather and Import Data - Data Loading, Storage and File Formats - Reading and Writing Data in Text Format, Web Scraping, Binary Data Formats, interacting with Web APIs, Interacting with Databases – Data Wrangling - Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting - Tools for Data Wrangling - Data Cleaning and Preparation - Handling Missing Data, Data Transformation, String Manipulation		02
MODULE - II **Unit 3: Data Visualization** Intro to data visualization - Introduction to Visualization and Dashboarding Software - Visualization Tools - Getting started with Tableau Desktop – Connecting to the dataset - Creating charts – Creating common visualizations (bar charts, line charts etc.) - Filtering		02

and sorting data - Adding Titles, Labels, and descriptions - Publish your work to Tableau Cloud - Interactivity with text and visual tooltips - Interactivity with actions (filter, highlight, URL) – Assembling dashboards from multiple charts **Unit 4: Story Telling** Introduction to Power BI - Understanding Desktop - Understanding Power BI Report Designer - Report Canvas, Report Pages: Creation, Renames - Report Visuals, Fields and UI Options - Experimenting Visual Interactions, Advantages - Reports with Multiple Pages and Advantages - Pages with Multiple Visualizations - PUBLISH Options and Report Verification in Cloud - Adding Report Titles. Report Format Options - Introduction to data storytelling - Creating a data story	

Reference Books:

1. Python for Data Analysis: Data Wrangling with Pandas, NumPy and IPython by McKinney, W., 2nd edition. O'Reilly Media, 2017
2. Doing Data Science: Straight Talk from the Frontline by O'Neil, C., & Schutt, R, O'Reilly Media, 2013
3. The Big Book of Dashboards by Steve Wexler, Jeffrey Shaffer, Andy Cotgreave, John Wiley & Sons, 2017
4. Practical Tableau by Ryan Sleeper, O'Reilly Media, 2018
5. Power BI. Book-1, Business Intelligence Clinic: Create and Learn by Roger F Silva, 2018
6. Introducing Microsoft Power BI by Alberto Ferrari and Marco Russo, Microsoft Press, Washington, 2016

Programme Name: M.Sc. Data Science (Semester-I)	Course Name: Data Analysis and Visualization Practical
Total Credits: 02	Total Marks: 50
Universityassessment:50	

Prerequisite:
Basic understanding of data, types of data, visualization mechanisms and basic computer skills.

Course outcomes:

- Handle missing data and manage data wrangling and manipulation
- Create data visualization and report making using various software tools
- Demonstrate the visualizations and make interpretations
- Create a data story using various software tools.

Course Code	Course Title	Total Credits
PSDSP504	Data Analysis and VisualizationPractical	02
Note: - The following set of practicals could be implemented in Python/ R/ Power BI/ Tableau or any other suitable software.		
1	Implement Data Loading, Storage and File Formats. Read data and store them in text format.	
2	Implement the code to interact with Web APIs and to perform web scrapping.	
3	Demonstrate Data Cleaning and Preparation.	
4	Implement Data wrangling on a data set.	
5	Demonstrate the handling of missing data and string manipulation.	
6	Create common charts with title, labels and descriptions using Tableau.	
7	Perform sorting and filtering using tableau, create visualizations and publish it on Tableau Cloud.	
8	Perform data visualization using Power BI.	
9	Create reports using Power BI.	
10	Create a data story in Tableau or power BI.	

Programme Name: M.Sc. Data Science (Semester-I)	Course Name: Statistical Methods for Data Science
Total Credits: 02	Total Marks: 50
University assessment: 25	College assessment: 25

Prerequisite:
Knowledge of statistics and mathematical concepts

Course outcomes:

- Students will gain a solid understanding of foundational statistical concepts, including probability, sampling distributions, hypothesis testing, and confidence intervals. They will learn the principles and techniques used in statistical analysis.
- Students will learn how to apply statistical methods to analyze data in the context of data science. They will become proficient in using statistical techniques such as regression analysis, analysis of variance (ANOVA), chi-square tests, and non-parametric tests.
- Students will gain proficiency in using statistical software and programming languages such as R or Python to implement statistical analyses. They will learn how to write code to perform statistical calculations, visualize data, and automate data analysis processes.
- Students will enhance their critical thinking skills and ability to solve problems using statistical methods.

Course Code	Course Title	Total Credits
PSDS505	Statistical Methods for Data Science	02
MODULE - I:		02
Unit 1: Introduction to Applied Statistics The Nature of Statistics and Inference, What is "Big Data"?,Statistical Modelling, Statistical Significance Testing and Error Rates, Simple Example of InferenceUsing a Coin, Statistics is for Messy Situations, Type I versus Type II Errors, Point Estimates andConfidence Intervals, Variable Types, Sample Size, Statistical Power, and Statistical Significance, TheVerdict on Significance Testing, Training versus Test Data. Means, Correlations, Counts: Drawing Inferences: Computing z and Related Scores, Statistical Tests, Plotting Normal Distributions, Correlation Coefficients, Evaluating Pearson's r for StatisticalSignificance, Spearman's Rho: A Nonparametric Alternative to Pearson. Tests of Mean Differences: t-Tests for One Sample, Two Sample t-Test, Paired-Samples t-Test Categorical Data: Binomial Test, Categorical Data Having More Than Two Possibilities.		

Power Analysis and Sample Size Estimation: Power for t-Tests, Power for One-Way ANOVA, Power for Correlations. Analysis of Variance: Fixed Effects, Random Effects, Mixed Models, Introducing the Analysis of Variance (ANOVA), Performing the ANOVA, Random Effects ANOVA and Mixed Models, One-Way Random Effects ANOVA **Unit 2: Multivariate Techniques** Simple and Multiple Linear Regression, Hierarchical Regression, How Forward Regression Works Logistic Regression and the Generalized Linear Model, Predicting Probabilities, Multiple LogisticRegression, Training Error Rate Versus Test Error Rate. Multivariate Analysis of Variance (MANOVA)and Discriminant Analysis: Multivariate Tests of Significance, Example of MANOVA, Outliers,Homogeneity of Covariance Matrices, Linear Discriminant Function Analysis, Theory of DiscriminantAnalysis, Predicting Group Membership, Visualizing Separation. Principal Component Analysis: Principal Component Analysis Versus Factor Analysis, Properties ofPrincipal Components, Component Scores, How Many Components to Keep? Exploratory FactorAnalysis, Common Factor Analysis Model, Factor Analysis Versus Principal Component Analysis onthe Same, Initial Eigenvalues in Factor Analysis, Rotation in Exploratory Factor Analysis, Estimation inFactor Analysis. Cluster Analysis: k-Means Cluster Analysis, Minimizing Criteria, Example of k-Means Clustering Hierarchical Cluster Analysis, Why Clustering Is Inherently Subjective. Nonparametric Tests: Mann– Whitney U Test, Kruskal–Wallis Test, Nonparametric Test for PairedComparisons and Repeated	

Reference Books:

1. Gupta S. C., Kapoor V. K.: Fundamentals of Mathematical Statistics; Tenth Edition.Sultan Chand &Sons. (2000)
2. Johnson, R.A., Wichern, D.W.: Applied Multivariate Statistical Analysis, Prentice-Hall, NewJersey, 2002.
3. Draper, N. R. and Smith, H. (1998), Applied Regression Analysis (John Wiley), Third Edition.
4. Purohit, S. G. Gore, S.D. and Deshmukh, S.R. (2015). Statistics using R, second edition. Narosa Publishing House, New Delhi.
5. Daniel W. W.: Applied Non-Parametric Statistics, First edition Boston-Houghton Mifflin Company.

ELECTIVES

Programme Name: M.Sc. Data Science (Semester-I)	Course Name: SPARK Technologies
Total Credits: 02	Total Marks: 50
University assessment: 25	College assessment: 25

Prerequisite:
Data mining Techniques. Knowledge of Python for implementation.

Course outcomes:

- Understand the concept of SPARK Technologies and its implementation
- Understand the concept of RDD
- Understand the implementation of SPARK SQL, GraphX, Performance Tuning.
- Use Sparks Resilient Distributed Datasets to process and analyze large data sets across many CPUs
- Understand how the GraphX library helps with network analysis problems
- Understand how Spark SQL lets you work with structured data

Course Code	Course Title	Total Credits
PSDS506a	SPARK Technologies	02

MODULE - I:Basics of SPARK Technologies **Unit 1: Introduction to SPARK Technologies** Components of the Spark unified stack, Features of Spark, Spark Web UI, an introduction to RDDs - Resilient Distributed Datasets, Launching and using Spark's Scala and Python shell, Spark Context, Spark Ecosystem, In-Memory data – Spark, Creating,Loading and Saving RDD, Transformations in RDD, Actions in RDD, Key-Value Pair RDD, Map Reduce and Pair RDD operations RDD Partitions **Unit 2: Implementation of SPARK Technologies** Spark Applications vs. Spark Shell, Creating Spark Context, Building a Spark Application, Spark and Hadoop Integration-HDFS, Handling Sequence File, Spark RDD-RDD Lineage, RDD Persistence Overview, Distributed Persistence. Spark Streaming, ML library for Spark, Working with Statistics, SPARK SQL, GraphX, Performance Tuning.	**02**

Reference Books:

1) Learning Spark: Lightning-Fast Data Analytics 2nd Edition, by Jules S. Damji,Brooke Wenig , Tathagata Das, Denny Lee, O'Reilly , 2020
2) Apache Spark Machine Learning Blueprints 1st Edition, Kindle Edition by Alex Liu, Packt Publishing, 2016
3) Apache Spark 2.x Cookbook: Cloud-ready recipes for analytics and data science 2nd Edition, by Rishi Yadav, Packt Publishing, 2017

Programme Name: M.Sc. Data Science (Semester-I)	Course Name:SPARK Technologies Practical
Total Credits: 02	Total Marks: 50
University assessment: 50	

Prerequisite:
Basic understanding of programming language (python) and Data mining techniques

Course outcomes:

- Understand the concept of SPARK Technologies and its implementation
- Understand the concept of RDD
- Understand the implementation of SPARK SQL, GraphX, Performance Tuning

Course Code	Course Title	Total Credits
PSDSP506a	SPARK Technologies Practical	02
1	Installation of Apache Spark	
2	Spark Basics and RDD interface	
3	Filtering RDDs, and the Minimum Temperature by Location Example	
4	Counting Word Occurrences using flatmap()	
5	Executing SQL commands and SQL-style functions on aDataFrame	
6	Implement Total Spent by Customer with DataFrames	
7	Use Broadcast Variables to Display Movie Names Instead of ID Numbers	
8	Create Similar Movies from One Million Rating	
9	Using Spark ML to Produce Movie Recommendations	
10	Use Windows with Structured Streaming to Track Most-Viewed URLs (SparkStreaming)	

Programme Name: M.Sc. Data Science (Semester-I)	Course Name: Retail Marketing Analytics
Total Credits: 02	Total Marks: 50
University assessment: 25	College assessment: 25

Prerequisite:
Knowledge of statistics and mathematical concepts

Course outcomes:

- Understand the importance of marketing analytics for forward looking and systematic allocation of marketing resources
- Know how to use marketing analytics to develop predictive marketing dashboard for organization
- Analyze data and develop insights from it to address strategic marketing challenges

Course Code	Course Title	Total Credits
PSDS506b	Retail Marketing Analytics	02

MODULE I — **02**

Unit 1: Introduction to Marketing Analytics and Exploratory Data analytics using R

a) Course Introduction
- why marketing analytics?
- course description and learning objectives

b) Marketing Analytics Overview
- how analytics can assist marketing decision-making
- the framework of marketing optimization

c) Tabulate and Summarize data
- what cleaned data looks like
- simple histogram plot
- use histogram and boxplot to inform data distribution

d) Visualize data
- elements of data visualization
- histogram, scatter plot, line plot, bar charts, line fits with the ggplot() function

Unit 2:Marketing Campaigns - Experiment Design, Customer Lifetime Value (CLV) and Cohort Analysis

a) Design and Conduct Experiments

• design experiments, examples • randomization/sample splitting • conduct experiments b) Assess Experiment Outcome Using Hypothesis Testing • why hypothesis testing for experiment outcomes • terminologies for hypothesis testing • how does hypothesis testing work • power calculation • conduct hypothesis testing in R c) Calculate and Predict CLV • calculate CLV • typical frameworks in predicting CLV • using linear regression and logistic regression to predict CLV d) CLV Analysis and Cohort AnalysisIntroduction to Experiment	

Reference Books:

1. Hands-on Data Science for Marketing by Yoon Hyup Hwang, Packt Publishing, 2019
2. Retail Analytics: The Secret Weapon by Emmett Cox,1st edition , Weily , 2011
3. Cutting Edge Marketing Analytics: Real World Cases and Data Sets for Hands on Learning by Venkatesan Rajkumar, Farris Paul and Ronald Wilcox, Pearson FT Press, 2014
4. Marketing Analytics: A Practical Guide to Real Marketing Science by Grigsby Mike,Kogan Page, 2015

Programme Name: M.Sc. Data Science (Semester-I)	Course Name: Retail Marketing Analytics Practical
Total Credits: 02	Total Marks: 50
University assessment: 50	

Prerequisite:
Knowledge of statistics and mathematical concepts

Course outcomes:

- To Learn working and analyzing with marketing data
- To develop predictive marketing dashboard for organization
- Understand the concept of hypothesis testing and its role in assessing experiment outcomes
-

Course Code	Course Title	Total Credits
PSDSP506b	Retail Marketing Analytics Practical	02
Note: Being able to approach data using statistical software is one of the essential goals of this class. You are required to use R for all assignments and projects throughout this course. Completing homework and quizzes using Excel or other program languages is not accepted. Programming knowledge prior to the class is preferred, but not required. As we spend time in-class to familiarize you with the RStudio interface and basic functions in the first few weeks of the class, take this time ask questions and adapt to R as soon as possible. • Download R from http://cran.r-project.org/ • Download R Studio from http://www.rstudio.com/products/rstudio/download/		
1	**Learn how to tabulate and summarize marketing data using R.** • Clean and preprocess the marketing data. • Generate a simple histogram plot to visualize data distribution. • Use tabulation and summary functions to gain insights from the data. • Interpret the findings and discuss the implications for marketing analysis.	

2	**Gain proficiency in visualizing marketing data using R.** • Understand the key elements of data visualization. • Create various visualizations such as histograms, scatter plots, line plots, and bar charts using the ggplot() function in R. • Apply appropriate visualization techniques to effectively communicate marketing insights.	
3	**Design and conduct experiments for marketing campaigns.** • Learn about experimental design and its application in marketing. • Design experiments using examples from marketing scenarios. • Implement randomization and sample splitting techniques. • Conduct the experiments and collect relevant data for analysis.	
4	**Understand the concept of hypothesis testing and its role in assessing experiment outcomes.** • Explore the purpose of hypothesis testing in analyzing experiment results. • Familiarize with key terminologies related to hypothesis testing. • Learn the process of hypothesis testing and power calculation. • Conduct hypothesis testing using R to evaluate experiment outcomes.	
5	**Calculate and predict Customer Lifetime Value (CLV).** • Calculate CLV using different approaches and frameworks. • Explore predictive modeling techniques such as linear regression and logistic regression for CLV prediction. • Assess the accuracy and reliability of CLV predictions.	
6	**Apply CLV analysis and cohort analysis in marketing analytics.** • Analyze CLV data and identify patterns and trends. • Perform cohort analysis to segment customers based on their behavior or characteristics. • Interpret the results of CLV analysis and cohort analysis to derive actionable insights for marketing strategies.	
7	**Extract data from social media platforms and perform analysis to gain insights into customer behavior and preferences.** • Utilize Python libraries like BeautifulSoup and requests to scrape data from social media platforms. • Clean and preprocess the scraped data. • Analyze the data to identify trends, sentiment analysis, or customer engagement metrics. • Visualize the findings using appropriate charts or graphs.	
8	**Analyze customer purchasing patterns and build a recommender system based on market basket analysis.**	

	• Use transactional data to identify frequently occurring item sets using association rule mining algorithms. • Calculate support, confidence, and lift for the identified item sets. • Build a recommendation engine using collaborative filtering techniques. • Evaluate the performance of the recommender system and make recommendations based on customer preferences.	
9	**Segment customers based on their recency, frequency, and monetary value (RFM) to better target marketing efforts.** • Analyze customer transaction data to calculate RFM scores. • Segment customers into different groups using clustering algorithms such as k-means or hierarchical clustering. • Perform descriptive analysis on each customer segment to understand their characteristics. • Develop targeted marketing strategies for each segment based on their RFM profiles.	
10	**Conduct A/B testing to evaluate the impact of different marketing strategies and make data-driven decisions.** • Design and implement A/B tests for marketing campaigns using randomized assignment. • Collect relevant data and perform statistical analysis to compare the performance of different strategies. • Calculate key metrics such as conversion rates, click-through rates, or revenue. • Interpret the results and provide recommendations for optimizing marketing campaigns based on the findings.	

Programme Name: M.Sc. Data Science (Semester-I)	Course Name: Sports Data Analytics
Total Credits: 02	Total Marks: 50
University assessment: 25	College assessment: 25

Prerequisite:
Basic knowledge of statistics and probability, Familiarity with programming concepts and a programming language (e.g., Python), Background in computer science or related field

Course outcomes:

- Understand the role and importance of data analytics in sports
- Develop skills in collecting, cleaning, and managing sports data
- Gain proficiency in using statistical analysis techniques to analyze sports data
- Apply data visualization methods to present sports data effectively
- Learn how to apply predictive modeling techniques to sports data
- Explore the use of machine learning algorithms in sports analytics
- Understand ethical considerations and challenges in sports data analytics

Course Code	Course Title	Total Credits
PSDS506c	Sports Data Analytics	02
MODULE I **Unit 1: Fundamentals of Sports Data Analytics** A. Introduction to Sports Data Analytics • Overview of sports data analytics • Evolution of analytics in sports • Importance and applications of sports data analytics B. Data Collection and Preprocessing • Sources of sports data • Data collection methods • Data cleaning and preprocessing techniques • Database management for sports analytics C. Data Visualization for Sports Analytics • Principles of data visualization • Visualization tools and libraries		02

- Creating effective visualizations for sports data
- Interactive dashboards for sports analytics

D. Statistical Analysis in Sports
- Descriptive statistics for sports data
- Hypothesis testing in sports analytics
- Regression analysis in sports
- Analysis of variance (ANOVA) in sports

Unit 2:Advanced Techniques in Sports Data Analytics

E. Predictive Modeling in Sports Analytics
- Introduction to predictive modeling
- Feature selection and engineering for sports data
- Linear regression models for sports prediction
- Classification models for sports outcomes

F. Machine Learning in Sports Analytics
- Overview of machine learning algorithms
- Decision trees and random forests in sports analytics
- Support vector machines for sports prediction
- Neural networks and deep learning in sports analytics

G. Advanced Topics in Sports Data Analytics
- Sports performance analysis
- Player tracking and motion analytics
- Sports marketing and fan engagement analytics
- Sports injury prediction and prevention

H. Sports Business Analytics
- Revenue generation and marketing in sports
- Fan engagement and customer analytics

Reference Books:

1. Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers by Benjamin C. Alamar, Columbia university press, 2013
2. Sports Analytics and Data Science: Winning the Game with Methods and Models by Thomas Miller, 1st edition, Pearson FT Press, 2015
3. Sports Analytics: Analysis, Visualisation and Decision Making in Sports Performanceby Daniel Memmert, Tim McGarry, and Tony Reilly, 2018
4. Cricket Analytics: Analytics and Data Science in Cricket by Tapan Bagchi and S. Raghunathan
5. Machine Learning using Python by Manaranjan Pradhan and U. Dinesh Kumar, Weily , 2020

URL for Online Study Material –

1. www.coursera.org
2. www.kaggle.com
3. www.datacamp.com
4. www.sportsanalyticsinstitute.com

Programme Name: M.Sc. Data Science (Semester-I)	Course Name:Sports Data Analytics Practical
Total Credits: 02	Total Marks: 50
University assessment: 50	

Prerequisite:
Sports knowledge, Basic statistics, Programming skills**,** Data cleaning and preprocessing methods

Course outcomes:

- Understanding of Sports Data Analytics
- Proficiency in Data Manipulation and Exploratory Data Analysis
- Player Performance/ Team Performance Analysis
- Predictive Modeling
- Data Visualization and Reporting

Course Code	Course Title	Total Credits
PSDSP506c	Sports Data Analytics Practical	02
Note: -The following set of practical could be implemented in Python/ R/ Power BI/ Tableau or any other suitable software		
1	**Exploratory Data Analysis** • Perform exploratory data analysis on a cricket dataset, analyzing variables such as number of matches, runs, not outs, wickets, etc. • Visualize the distribution of player performance metrics using histograms, box plots, or scatter plots. Investigate the relationship between player age and performance metrics using correlation analysis.	
2	**Batting Performance Analysis** • Analyze batting performance in a Cricket dataset, calculating metrics such as batting average, strike rate, and runs scored by players. • Identify top-performing batsmen based on performance metrics and compare their performance against different opponents or in specific conditions.	

3	**Bowling Performance Analysis** • Analyze bowling performance in the Cricket dataset, calculating metrics such as bowling average, economy rate, and wickets taken by players. • Identify top-performing bowlers based on performance metrics and analyze their performance against different teams or in various match situations	
4	**Performance Comparison** • Compare the scoring averages of top-performing batsman in different seasons. • Analyze the runs scoring (strike rate of Batting) of players from various teams in a specific league	
5	**Player Position Analysis** • Calculate position-specific performance metrics and compare players within each position.	
6	**Injury Analysis** • Investigate the relationship between player injuries and their subsequentperformance using historical injury and performance data. • Identify patterns and trends in the data to determine the impact of injuries on playerperformance and team success.	
7	**Team Analysis** • Analyze the impact of toss on a team's overall scoring and winning percentage. • Study the relationship between batting averages of players and their team's win-loss record	
8	**Sports Revenue Analysis** • Analyze revenue generation in sports organizations by examining factors such as ticket sales, merchandise sales, and sponsorship deals. • Identify key drivers of revenue and provide recommendations for maximizing financial performance.	
9	**Predictive Modeling** • Build a regression model to predict the number of runs scored by players based ontheir historical performance data. • Develop a classification model to predict the outcome of match based on team'sstatistics.	
10	**Visualization and Reporting: (Mini-Project)**	

	• Prepare a comprehensive report summarizing the findings of the analysis and providing actionable insights for sports teams or organizations.	

Programme Name: M.Sc. Data Science (Semester-I)	Course Name: Research Methodology
Total Credits: 04	Total Marks: 100
University assessment: 50	College assessment: 50

Prerequisite:

Basic Knowledge of statistical methods, analytical and logical thinking.

Course outcomes:

- Formulate clear and relevant research questions and objectives in the context of Data Sciences.
- Choose appropriate research designs, sampling methods, and data collection techniques for data-driven research projects.
- Gather, preprocess, and manage datasets, ensuring data quality and ethical handling.
- Apply statistical analysis, machine learning, and deep learning techniques to analyze and extract insights from data in Data Sciences.
- Effectively communicate research findings through research papers and presentations tailored for the Data Sciences community.
- Understand and adhere to ethical principles and data privacy regulations when conducting research involving sensitive data.
- Prepare and submit research manuscripts to conferences and journals, understanding the peer-review process in Data Sciences.
- Analyze and extract insights from large-scale datasets using big data analytics tools and techniques.
- Recognize and address ethical and security concerns related to data privacy and data protection in research.

Course Code	Course Title	Total Credits
PSDS510	Research Methodology	04
MODULE – I **Unit 1: Introduction to Research Methodology for Data Sciences** **Understanding Research in Data Sciences:** Definition and significance of research in Data Sciences, Types of research: Basic research, applied research, and practical research, Research approaches: Quantitative vs. Qualitative, The role of research in Data Sciences. **Research Process and Ethics:** Steps in the research process in Data Sciences, Research ethics and responsible conduct, Formulating research questions and objectives, Identifying research problems in Data Sciences, Ethical considerations in research, including data privacy and security. **Research Design and Sampling for Data Sciences:** Principles of research design in Data Sciences, Types of research design, Sampling techniques in Data Sciences, Choosing the appropriate research design for data-driven research, Ethical considerations in sampling and data collection, including informed consent.		02

Unit 2: Data Collection and Management for Data Sciences **Data Sources and Collection Methods:** Data sources: Primary vs. secondary data, Data collection methods: Surveys, questionnaires, web scraping, and APIs, Data quality and data cleaning, Ethical considerations in data collection, including data privacy. **Data preprocessing techniques:** Cleaning, transformation, and normalization, Data integration and data fusion, Handling missing data and outliers **Exploratory Data Analysis (EDA):** Techniques for exploring and visualizing data, Descriptive statistics and data visualization, Identifying patterns and trends.	
MODULE - II **Unit 3: Data Analysis and Modeling for Data Sciences** **Statistical Analysis for Data Sciences:** Inferential statistics for hypothesis testing, Regression analysis and correlation, Machine learning concepts for predictive modeling. **Machine Learning and Deep Learning:** Introduction to machine learning and deep learning algorithms, Model selection and evaluation, **Big Data Analytics:** Understanding big data and its challenges, Tools and frameworks for big data processing (e.g., Hadoop, Spark), Analyzing and extracting insights from large datasets. **Unit 4: Research Communication, Publishing, and Data Privacy** **Research Paper Writing and Presentation:** Structure of a research paper, Writing the abstract, introduction, literature review, methodology, and results sections, Effective research paper presentations. **Publishing Research in Data Sciences:** Choosing the right conferences and journals, The peer-review process in Data Sciences, Preparing manuscripts for submission, Ethical considerations in publishing, including data privacy and security in publications. **Ethical Considerations in Data Privacy and Security:** Data privacy regulations and compliance, Ethical considerations in data anonymization and de-identification, Securing research data and protecting sensitive information.	**02**

Textbook:

1. Research Methodology: Methods and Techniques, C.R. Kothari, New Age International

Sem-II

SEMESTER -II

Programme Name: M.Sc. Data Science (Semester-II)	Course Name: Soft Computing
Total Credits: 04	Total Marks: 100
University assessment: 50	College assessment: 50

Prerequisite:
Foundation of mathematics, statistics and basic principles of AI

Course outcomes:

- Understanding Soft Computing Paradigms.
- Comprehend the principles and mathematical foundations of Fuzzy logic.
- Acquire knowledge of the theory and concepts underlying neural networks, including artificial neurons, activation functions.
- Explore the theoretical understanding of Genetic algorithms, including genetic operators (selection, crossover, mutation), encoding schemes, fitness evaluation

Course Code	Course Title	Total Credits
PSDS511	Soft Computing	04
MODULE - I **Unit 1: Artificial Neural Network** Fundamentalconcepts,Evolution of neural network, basic model of Artificial Neural Network, Important terminologies, McCulloch Pits neuron, linear separability, Hebb network Supervised Learning Network: Perceptron networks, Adaline, MAdaline, Backpropogation network, Radial Basis Function, Time Delay Network, Functional Link Networks, Tree Neural Network. **UnSupervised Learning Networks**: Fixed weight competitive nets, Kohonen self-organizing feature maps, learning vectors quantization, counter propogation networks, adaptive resonance theory networks. Associative Memory Networks: Training algorithm for pattern Association, Autoassociative memory network, hetroassociative memory network, bi-directional associative memory, Hopfield networks, iterative autoassociative memory networks, temporal associative memory networks **Unit2: Special Networks** Simulated annealing, Boltzman machine, Gaussian Machine, Cauchy Machine, Probabilistic neural net, cascade correlation network, cognition network, neo-cognition network, cellular neural network, optical neural network Third Generation Neural Networks: Spiking Neural networks, convolutional neural networks, deep learning neural networks, extreme learning machine model.		02

MODULE - II **Unit 3: Fuzzy Logic** Introduction to Fuzzy Logic, Classical sets, Fuzzy sets, Classical Relations and Fuzzy Relations: Cartesian Product of relation, classical relation, fuzzy relations, tolerance and equivalence relations, non-iterative fuzzy sets. Membership Function: features of the membership functions, fuzzificationand methods of membership value assignments. Defuzzification: Lambda-cuts for fuzzy sets, Lambda-cuts for fuzzy relations, Defuzzification methods. Fuzzy Arithmetic and Fuzzy measures: fuzzy arithmetic, fuzzy measures, measures of fuzziness, fuzzy integrals **Unit 4: Genetic Algorithm** Biological Background, Traditional optimization and search techniques, genetic algorithm and search space, genetic algorithm vs. traditional algorithms, basic terminologies, simple genetic algorithm, general genetic algorithm, operators in genetic algorithm, stopping condition for genetic algorithm flow, constraints in genetic algorithm, problem solving using genetic algorithm, the schema theorem, classification of genetic algorithm, Holland classifier systems, genetic programming, advantages and limitations and applications of genetic algorithm	**02**

Reference Books:

1. Artificial Intelligence and Soft Computing by Anandita Das Bhattacharya, SPD, 3rd edition 2018
2. Principles of Soft computing by S.N.SivanandamS.N.Deepa , Wiley, 3rd edition, 2019
3. Neuro-Fuzzy and Soft Computing by J.S.R.Jang, C.T.Sun and E.Mizutani, Prentice Hall of India, 1st edition, 2004
4. Neural Networks, Fuzzy Logic and Genetic Algorithms: Synthesis & Applications by S.Rajasekaran, G. A. Vijayalakshami , Prentice Hall of India, 1st edition, 2004
5. Fuzzy Logic with Engineering Applications, Timothy J.Ross, McGrawHill 1st edition, 1997
6. Genetic Algorithms: Search, Optimization and Machine Learning by Davis E.Goldberg, Addison Wesley, 1st edition, 1989
7. Introduction to AI and Expert System by Dan W. Patterson, Prentice Hall of India, 2nd edition, 2009

Programme Name: M.Sc. Data Science (Semester-II)	Course Name:Soft Computing Practical
Total Credits: 02	Total Marks: 50
University assessment: 50	

Pre requisite:
Understanding of Mathematics, Programming skills

Course outcomes:

- Understanding of the basic principles, concepts, and techniques of Soft Computing.
- Acquire the knowledge and skills necessary to apply Soft Computing techniques to solve real-world problems.
- Implement and program Soft Computing algorithms.
- Provide foundation in Soft Computing.

Course Code	Course Title	Total Credits
PSDSP512	Soft Computing Practical	02
1	Write a program to implement logical gates AND, OR and NOT with McCulloch-Pitts.	
2	Write a program to implement Hebb's rule.	
3	Implement Kohonen Self organizing map.	
4	Solve the Hamming network given the exemplar vectors.	
5	Write a program for implementing BAM network.	
6	Implement a program to find the winning neuron using MaxNet.	
7	Implement De-Morgan's Law.	
8	Implement Union, Intersection, Complement and Difference operations on fuzzy sets.	
9	Create fuzzy relation by Cartesian product of any two fuzzy sets	
10	Perform max-min composition on any two fuzzy relations.	

Programme Name: M.Sc. Data Science (Semester-II)	Course Name: Time Series Analysis and Forecasting
Total Credits: 04	Total Marks: 100
University assessment: 50	College assessment: 50

Prerequisite:
It is assumed the learner is familiar with regression model estimation and related hypothesis testing techniques.

Course outcomes:

- Forecast the trend pattern exhibited by the given data by using various methods
- Run and interpret time series models and regression models for time series
- Use the Box-Jenkins approach to model and forecast time series data empirically
- Analyze and estimate the cyclic components using special processes

Course Code	Course Title	Total Credits
PSDS513	Time Series Analysis and Forecasting	04
MODULE - I **Unit 1: Introduction to Trend** Introduction to times series data, application of time series from various fields - Components of a time series - Decomposition of time series. Trend: Estimation of trend by free hand curve method - method of semi averages - fitting a various mathematical curve and growth curves. **Unit2: Trend and Seasonal component** Method of moving averages – Detrending - Effect of elimination of trend on other components of the time series. Seasonal Component: Estimation of seasonal component by Method of simple averages, Ratio to Trend, Ratio to moving average and Link relatives		02
MODULE - II **Unit 3: Forecasting** Variate component method - Stationary Time series: Weak stationary, autocorrelation function and correlogram of moving average Forecasting: Exponential smoothing methods, short term forecasting methods: Brown's discounted regression, Box-Jenkins Method. **Unit 4: Cyclic Component** Deseasonalization - Cyclic Component: Harmonic Analysis. Some Special Processes: Moving-average (MA) process and Autoregressive (AR) process of orders one and two - Estimation of the parameters of AR (1) and AR (2) – Yule-Walker equations.		02

Reference Books:

1. Kendall, M. (1976) Time Series. 2nd Edition, Charles Griffin and Co Ltd., London and High Wycombe.
2. Chatfield C. (1980). The Analysis of Time Series –An Introduction, 6th Edition, Chapman & Hall.
3. Mukhopadhyay P. (2011). Applied Statistics, 2nd ed. Revised reprint, Books and Allied
4. Shumway, R. H., and Stoffer, D. S. (2006). Time Series Analysis and Its Applications With R Examples, 2 ed. Springer, New York, NY
5. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). Time Series Analysis: Forecasting and Control. Prentice - Hall, Inc., Upper Saddle River, NJ.
6. Yaffee, R. and McGee, M. (2000). Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS. Academic Press, Inc., San Diego, CA.

Programme Name: M.Sc. Data Science (Semester-II)	Course Name: Time Series Analysis and Forecasting Practical
Total Credits: 02	Total Marks: 50
University assessment: 50	

Prerequisite:
Basic understanding of data and statistical programming tools/platforms.

Course outcomes:

- Fit various growth curves, trend and to measure seasonal indices
- Understand forecasting by different methods
- Able to calculate variance of a random component

Course Code	Course Title	Total Credits
PSDSP514	Time Series Analysis and Forecasting Practical	02
Note: Software: Time Series Lab/any statistical software/any programming platform		
1	Fitting and plotting of modified exponential curve.	
2	Fitting and plotting of Gompertz curve.	
3	Fitting and plotting of logistic curve.	
4	Fitting of trend by Moving Average Method.	
5	Measurement of Seasonal indices Ratio-to-Trend method.	
6	Measurement of Seasonal indices Ratio-to-Moving Average method.	
7	Measurement of seasonal indices Link Relative method.	
8	Calculation of variance of random component by variate difference method.	
9	Forecasting by exponential smoothing.	
10	Forecasting by short term forecasting methods.	

Programme Name: M.Sc. Data Science (Semester-II)	Course Name: Ethical Issues in Data Science
Total Credits: 02	Total Marks: 50
University assessment: 25	College assessment: 25

Prerequisite:
Data science concepts, Techniques used in data science.

Course outcomes:

- Understand the fundamental ethical issues that arise in the field of data science
- Understand the ethical implications of data collection, and decision-making processes
- Understand concerns related to data collection, storage, and sharing
- Understand the principles to make decisions in data science projects
- Recognize and address privacy concerns related to data collection, storage, and sharing
- Reflect on the emerging ethical issues and future directions in data science

Course Code	Course Title	Total Credits
PSDS515	Statistical Methods for Data Science	02
MODULE I **Unit 1: Ethical foundation in Data Science** Introduction to ethical frameworks and principles, Ethics in data science: challenges and importance, Data collection methods, storage, sharing and its ethical considerations, Types of bias in data, Importance of transparency in data science, Ethical considerations in automated decision-making, Data governance frameworks and practices, Ensuring accountability in data science projects. **Unit 2: Emerging Ethical issues in Data Science** Ethical Issues in Data Visualization, Ethical Issues in Machine Learning, Ethical challenges in emerging technologies e.g., AI, IoT, biometrics, blockchain, Ethical challenges in data science research, Ethical considerations in collaborative data science environments, Ethical issues in using the internet, privacy and security, in the context of data science.		02

Reference Books:

1. Data Science Ethics, David Martens ISBN: 9780192847263 Oxford University Press 2023
2. Ethics of Big Data: Balancing Risk and Innovation" by Kord Davis and Doug Patterson O'Reilly 2012.
3. Data Science Ethics Resources - Concepts, Techniques, and Cautionary Talesby David Martens-Oxford University Press-2022.
4. Data Science Association, Data Science Code of Professional Conduct.- https://www.datascienceassn.org/code-of-conduct.html

ELECTIVES

Programme Name: M.Sc. Data Science (Semester-II)	Course Name: Human Resource Analytics
Total Credits: 02	Total Marks: 50
University assessment: 25	College assessment: 25

Prerequisite:
Tools, methods, approaches and techniques of HR

Course outcomes:

- Analyze problems and issues in HR and the relevance of HR analytics.
- Logically synthesize the tools, methods and techniques of HR analytics to understand real world corporate scenario.
- Identify the application and uses of HR analytics in various HR sub-systems

Course Code	Course Title	Total Credits
PSDS516a	Human Resource Analytics	02
MODULE - I **Unit 1: HR Measurement** Need for HR Measurement, Significance and concept of HR Analytics, HR Analytics and business linkages, Prerequisites of HR Analytics; Models and frameworks of HR Analytics; Measuring intellectual capital, need and rationale for HR Accounting & Audit, Approaches and methods of HR Accounting & Audit **HRIS for HR Analytics**: What is Human Resource Information System; Role of HRIS in analytics; HRIS development and Implementation, the development process- need analysis, systems design, structure and culture; HRIS Applications-making HRIS work. **Unit 2: Analytics for HR sub-systems** HR Analytics for Staffing, Training & Development, Performance Management Systems, Career Planning Systems, Rewards and Compensation Management, Employee Relations Systems. **Analytics for HR system:** HR performance frameworks and measurement systems; Measuring HR Climate and People Management Capabilities; Competency Management Frameworks & Competency Mapping, Integration of competency-based HR System. Measuring HR Effectiveness, The HR Scorecard **Trends and Future Challenges:**Technology and changes in HR Analytics, Role of social media, Big Data and Predictive Analytics in HR, Assessing the effectiveness of HR Analytics, Post analysis steps, Review and monitoring, Issues in HR valuation and measurement; Emerging challenges: Global and Indian Experience		02

Reference Books:

1. Ulrich, D. & Brockbank, W., The HR Value Proposition. Harvard Business School Press 2016
2. How to measure HRM by Jac Fitz-enz 2002
3. Predictive Analytics for Human Resources by Jac Fitz-enz, John Mattox II, Wiley 2014
4. Making Human Capital Analytics Work: Measuring the ROI of Human Capital Processes and Outcomes. By by Jack Phillips,Patricia Pulliam Phillips- 2014

Programme Name: M.Sc. Data Science (Semester-II)	Course Name: Human Resource Analytics Practical
Total Credits: 02	Total Marks: 50
University assessment: 50	

Prerequisite:
Understanding of R

Course outcomes:

- Analyze HR analytics.
- Use and understand tools require for HR analytics
- Identify the application and uses of HR analytics in various HR sub-systems

Course Code	Course Title	Total Credits
PSDSP516a	Human Resource AnalyticsPractical	02
Note: -Being able to approach data using statistical software is one of the essential goals of this class. You are required to use R for all assignments and projects throughout this course. Completing homework and quizzes using Excel or other program languages is not accepted. Programming knowledge prior to the class is preferred, but not required. As we spend time in-class to familiarize you with the RStudio interface and basic functions in the first few weeks of the class, take this time ask questions and adapt to R as soon as possible. Download R from http://cran.r-project.org/ Download R Studio fromhttp://www.rstudio.com/products/rstudio/download/		
1	**Analyze employee turnover rates and identify factors contributing to attrition** • Collect historical employee data, including tenure, performance ratings, salary, and job satisfaction. • Calculate employee turnover rates for different departments and job roles. • Conduct statistical analysis to identify correlations between turnover and variables such as salary, job satisfaction, and performance. • Generate visualizations (e.g., charts, graphs) to present the findings and propose recommendations to reduce turnover.	

2	**Develop a user-friendly HRIS dashboard for monitoring and analyzing HR metrics** • Identify key HR metrics to be displayed on the dashboard (e.g., headcount, recruitment pipeline, training hours). • Design the layout and interface of the HRIS dashboard using appropriate programming languages and tools. • Integrate data from various HR systems and databases to populate the dashboard in real-time. • Implement interactive features, such as drill-down capabilities and data filters, to facilitate data exploration and analysis	
3	**Analyze training effectiveness and identify skill gaps in the organization** • Collect training data, including participant demographics, training modules, pre/post-assessment scores, and performance metrics. • Perform statistical analysis to evaluate the impact of training on employee performance. • Identify areas of improvement and recommend targeted training programs based on identified skill gaps. • Develop a visualization or report summarizing the training needs analysis results.	
4	**Develop an HR scorecard to measure HR effectiveness and align HR strategies with organizational goals** • Identify key HR performance indicators aligned with the organization's strategic objectives. • Collect relevant data for each HR indicator, such as employee satisfaction surveys, training investment data, and performance metrics. • Calculate HR metrics and indicators, such as turnover rate, training ROI, and employee engagement index. • Design a dashboard or report to present the HR scorecard and analyze trends over time.	
5	**Use predictive analytics to forecast employee attrition and develop retention strategies** • Gather historical HR data, including employee demographics, performance metrics, compensation, and employee exit data. • Build a predictive model (e.g., logistic regression, decision tree) to predict employee attrition. • Validate the model's accuracy and evaluate its performance using appropriate evaluation metrics.	

	• Generate actionable insights and recommendations to proactively address potential attrition risks.	
6	**Use predictive analytics to forecast employee attrition and develop retention strategies** • Gather historical HR data, including employee demographics, performance metrics, compensation, and employee exit data. • Build a predictive model (e.g., logistic regression, decision tree) to predict employee attrition. • Validate the model's accuracy and evaluate its performance using appropriate evaluation metrics. • Generate actionable insights and recommendations to proactively address potential attrition risks.	
7	**Measure and analyze employee engagement levels within the organization** • Collect employee engagement survey data, including responses to survey questions related to job satisfaction, work environment, and organizational culture. • Calculate engagement scores and identify key drivers of engagement. • Conduct a sentiment analysis on employee feedback to understand areas of improvement. • Present the findings and propose strategies to enhance employee engagement based on the analysis.	
8	**Develop a program to automate repetitive HR processes, such as leave management or performance appraisal** • Identify the HR process to be automated and define the required functionalities. • Design and implement a web-based application or script to streamline the process using appropriate programming languages and frameworks. • Integrate the application with relevant HR systems and databases to ensure data accuracy and consistency. • Test and validate the automated process, considering different scenarios and user inputs.	
9	**Analyze the effectiveness of the organization's performance management system and provide insights for improvement.** • Collect performance evaluation data, including performance ratings, goal achievement metrics, and feedback.	

	• Analyze the distribution of performance ratings across different departments or job roles. • Identify trends and patterns in performance data and assess the fairness and consistency of the evaluation process. • Propose recommendations for enhancing the performance management system based on the analysis results.	
10	**Analyze the organization's compensation structure and compare it to industry benchmarks.** • Gather salary data for different job roles and levels within the organization. • Perform a salary analysis, including measures like average salary, salary distribution, and salary competitiveness. • Conduct benchmarking by comparing the organization's salary data with industry standards or competitor data.	

Programme Name: M.Sc. Data Science (Semester-II)	Course Name: Public Health Analytics
Total Credits: 02	Total Marks: 50
University assessment: 25	College assessment: 25

Prerequisite:
Basic concepts of data analytics and machine learning

Course outcomes:

- Discuss the evolving landscape of healthcare services, including the growing importance of value-based healthcare systems and the role of data in enhancing outcomes. Explore the utilization and management of data in electronic health record (EHR) systems. Identify and cite relevant sources of public health data and information.
- Illustrate the various functions performed by data analysis in the field of public health. Apply commonly used graphical and descriptive techniques to summarize public health data effectively.
- Investigate the reliability, accuracy, and comparability of health and genomic data. Examine the integrity of these data sets and their suitability for meaningful comparisons.
- Acquire fundamental skills in using popular software tools for conducting data analyses.
- Develop data models that integrate patient information from multiple origins to create comprehensive and patient-centric perspectives.

Course Code	Course Title	Total Credits
PSDS516b	Public Health Analytics	02
MODULE - I **Unit 1: Healthcare data management** • What is Health Data Management? Benefits and challenges of health data management, how to store all that data • Electronic Health Records– Components of EHR- Coding Systems- Benefits of EHR- Barrier to Adopting EHR Challenges • Phenotyping Algorithms • Statistical analysis of healthcare data- Measures of Central Tendency and Dispersion, Confidence Limits and Hypothesis Testing, Statistical Tests for Categorical Data, T-Tests for Related and Unrelated Data, Analysis of Variance • Data Quality and Governance **Unit 2: Healthcare data Analysis:** • Biomedical Image and Signal Analysis, Genomic Data Analysis for Personalized		02

Medicine • Natural Language Processing and Data Mining for Clinical Text, Mining the Biomedical- Social Media Analytics for Healthcare. • Predictive Models for Integrating Clinical and Genomic Data, Privacy-Preserving Data, Publishing Methods in Healthcare • Mobile Imaging and Analytics for Biomedical Data, Data Analytics for Pharmaceutical Discoveries- Clinical Decision Support Systems	

Reference Books:

1. Healthcare Business Intelligence + Website - A Guide to Empowering Successful Data Reporting and Analytics Hardcover by LB Madsen (Author),2012
2. Practical Text Analytics: Interpreting Text and Unstructured Data for Business Intelligence (Marketing Science) 1st Edition by Dr. Steven Struhl,2016

Programme Name: M.Sc. Data Science	Course Name: Public Health Analytics Practical
Total Credits: 02	Total Marks: 50
University assessment: 50	

Prerequisite:
Basic computer skills and Basic understanding of elementary Math.

Course outcomes:

- Should be able to understand, visualize and infer healthcare data.
- Should be able to use linear regression and forecasting methods for predicting growth rates, expenditure or any other numeric variable related to the medical field.
- Should be able to demonstrate descriptive, diagnostic, and inferential statistics using Python, R or Excel
- Perform predictive analysis using machine learning algorithms and deep learning.
- Perform NLP and sentiment analysis.

Course Code	Course Title	Total Credits
PSDSP516b	**Public Health Analytics Practical**	**02**
1	EHR data modeling, data mining, exploratory data analysis using tableau or power BI.	
2	Medical expenditure prediction	
3	A Twitter Healthcare data extraction, processing and sentiment analysis use keywords related to health (e.g., COVID-19 vaccine).	
4	Using classification algorithms in prediction of diseases	
5	Clinical Text Data Categorization and Feature Extraction	
6	Medical Image classification (x-rays)	
7	Image object detection using deep learning.	
8	Health forecasting using time-series.	
9	Heart sound classification using signal processing.	
10	AI conversational chatbot for primary healthcare diagnosis.	

Programme Name: M.Sc. Data Science (Semester-II)	Course Name: Social Media Analytics
Total Credits: 02	Total Marks: 50
University assessment: 25	College assessment: 25

Prerequisite:
Graph Theory, Data mining Techniques. Knowledge of Python for implementation.

Course outcomes:

- To understand and familiarize the learners with the concept of social media.
- Social media analytics integrates with the learners to understand the significance.
- Enable the learners to develop skills required for analyzing the effectiveness of
- social media.
- Familiarize the learner with different visualization techniques for social media
- decisions.
- Examine the ethical and legal implications of leveraging social media data.

Course Code	Course Title	Total Credits
PSDS516c	Social Media Analytics	02
MODULE - I **Unit 1: Social Media Analytics: An Overview** Core Characteristics of social media, Types of social media, social media landscape, Need for Social Media Analytics (SMA), Seven Layers of Social Media Analytics, Types of Social Media Analytics, Social Media Analytics Cycle, Location Analytics - Sources of Location Data, Categories of Location Analytics, Social Information Filtering: Social Sharing and filtering, Automated Recommendation systems, Traditional Vs social Recommendation Systems, understanding social media and Business Alignment, social media KPI, formulating a Social Media Strategy, Managing Social Media Risks **CaseStudy**: Online Behavior on Twitter **Unit 2: Social Network Structure, Measures & Visualization** Basics of Social Network Structure - Nodes, Edges & Tie Describing the Networks Measures - Degree Distribution, Density, Connectivity, Centralization, Tie Strength & Trust. Network Visualization - Graph Layout, Visualizing Network features, Scale Issues. Capturing Correlations: Triangles, Clustering, and Assortativity. Social Media Network Analytics - Common Network Terms, Common Social Media Network Types, Types of Networks, Common Network Terminologies, Network Analytics Tools. **Case Study**: LinkedIn		02

Reference Books:

1. Seven Layers of Social Media Analytics_ Mining Business Insights from Social Media Text, Actions, Networks, Hyperlinks, Apps, Search Engine, and Location Data, Gohar F. Khan, 2015
2. Analyzing the Social Web 1st Edition by Jennifer Golbeck, 2013
3. Mining the Social Web_ Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites, Matthew A Russell, O'Reilly, 2019
4. Charu Aggarwal (ed.), Social Network Data Analytics, Springer, 2011

Useful Links

1. https://cse.iitkgp.ac.in/~pawang/courses/SC16.html
2. https://onlinecourses.nptel.ac.in/noc20_cs78/preview
3. https://nptel.ac.in/courses/106106146
4. https://7layersanalytics.com/

Programme Name: M.Sc. Data Science (Semester-II)	Course Name: Social Media Analytics Practical
Total Credits: 02	Total Marks: 50
University assessment: 50	

Prerequisite:
Python, Types of social media

Course outcomes:

- To understand the fundamental concepts of social media networks.
- To Collect, monitor, store and track social media data
- To analyze and visualize social media data
- To design and develop social media analytics models.

Course Code	Course Title	Total Credits
PSDSP516c	Social Media AnalyticsPractical	02
Note: - The following set of practicals should be implemented in Scrape, python: Link: -Python: https://www.python.org/downloads/ **Prepare Research Paper with Publication using any the derived analysis**		
1	**Study Various** • Social Media platforms (Facebook, twitter, YouTube etc) • Social Media analytics tools (Facebook insights, google analytics netlyticetc) • Social Media Analytics techniques and engagement metrics (page level, post level, member level) using Gephi Tool	
2	Scrape an online Social Media Site for Data. Use python to scrape information from twitter. Exploratory Data Analysis and visualization of Social Media Data	
3	Create sociograms for the persons-by-persons network and the community-by- community network for a given relevant problem. Create a one-mode network and two- node network for the same. Datasets: les-Misérables, Airlines, Internet Core Routers.	
4	Develop Content (text, emoticons, image, audio, video) based social media analytics model for business. (e.g., Content Based Analysis: Topic, Issue, Trend, sentiment/opinion analysis, audio, video, image analytics)	
5	Develop Structure based social media analytics model for any business. (e.g., Structure Based Models -community detection, influence analysis)	

6	Develop a dashboard and reporting tool based on real time social media data Using Power BI	
7	Use Google Visualization Charts to analyze social media data	
8	Analyze social media data Network Analysis with Orange Software	
9	Use Graph Neural Networks on the datasets (Planetoid Cora Dataset)/ Jazz Musicians Network.	
10	Analyze Twitter conversations to identify the most active and influential users using Machine Learning Algorithms with Gephi Tool.	

Programme Name: M.Sc. Data Science (Semester-II)	Course Name: On Job Training
Total Credits: 04	Total Marks: 100
University assessment: 50	College assessment: 50

A. Introduction

- On Job training (OJT) is an integral component of the M.Sc. Data Science program that provides students with a unique opportunity to bridge the gap between theoretical knowledge gained in the classroom and practical application in a real-world environment. This training aims to equip students with both technical and non-technical skills that are essential for success in the industry.
- By participating in OJT, students are able to apply the concepts and theories learned during their coursework to real-world scenarios. They gain hands-on experience, problem-solving skills, and a deeper understanding of how the industry operates. This practical exposure enhances their competence and confidence, preparing them to tackle the challenges they may encounter in their professional careers.
- From an organizational perspective, hosting OJT programs allows companies to gain insights into the curriculum and content of the M.Sc. Data Science program. They can provide valuable feedback on the relevance of the coursework and industry requirements, enabling academic institutions to continually improve the program's alignment with industry needs. This collaboration between academia and industry fosters a mutually beneficial relationship, ensuring that graduates are well-prepared for the job market.
- Moreover, OJT benefits the faculty members involved in the program. They have the opportunity to gain firsthand exposure to the industry and observe the type of work being performed. This experience enables them to enhance their teaching methodologies and delivery techniques, ensuring that they remain up-to-date with the latest industry practices. The insights gained from OJT also enable faculty members to provide relevant guidance and mentorship to students, preparing them for successful careers in the field of data science.

B. Enhancing Practical Skills through OJT

- The OnJob Training (OJT) program spans 4-6 weeks, requiring a minimum of 120 hours of physical presence at the organization.
- Students are expected to find their own OJT placements, although the institution provides support and guidance in securing positions with reputable organizations.

- OJT must be conducted outside the home institution to expose students to real-world work environments.
- OJT covers any subject within the syllabus, allowing students to align their experience with their academic interests.
- In recognition of changing dynamics, some OJT sessions can be conducted online to accommodate virtual work environments.
- OJT will offer students the opportunity to apply classroom learning in a real-world setting, fostering the development of technical and non-technical skills.
- Mutual Benefits: Organizations gain insights into the program's curriculum and industry requirements, enabling them to provide constructive feedback and enhance course relevance.
- OJT bridges the gap between theoretical knowledge and practical application, preparing students for successful careers in data science

C. **Interning organization:** Students have the flexibility to pursue their OJT in various types of organizations, including but not limited to:

- Software Development Firms: Gain practical experience in software development and programming.
- Hardware/Manufacturing Firms: Learn about hardware design, manufacturing processes, and quality assurance.
- Small-Scale Industries/Service Providers: Explore opportunities in diverse sectors such as banking, clinics, NGOs, and professional institutions like CA firms or law firms.
- Civic Departments: Engage with local civic departments such as ward offices, post offices, police stations, or panchayats to understand their functioning and contribute to their activities.
- Research Centre's/University Departments/Colleges: Contribute as research assistants or in similar roles for research projects or initiatives, fostering collaboration between academia and industry.

Note: The listed options provide a range of possible OJT placements, offering students valuable exposure to different sectors and professional settings.

D. OJT mentors:

To enhance the learning experience and ensure the quality of the MSc program, each student participating in the OJT will be assigned two mentors: a faculty mentor from the institution and an industry mentor from the organization where the student is interning.

- **Industry Mentor Role:** The industry mentor plays a crucial role in guiding the student during the internship. They ensure that the internee fulfills the requirements of the organization and successfully meets the demands of the assigned project. Through their

expertise and experience, industry mentors provide valuable insights into real-world practices and industry expectations.

- **Faculty Mentor Role:** The faculty mentor serves as the overall coordinator of the OJT program. They oversee the entire internship process and evaluate the quality of the OJT in a consistent manner across all students. The faculty mentor ensures that the OJT aligns with the program's objectives and provides valuable learning opportunities. They also facilitate communication between the institution, industry mentor, and student to ensure a fruitful OJT experience.

By having both an industry mentor and a faculty mentor, students benefit from a comprehensive guidance system that combines industry expertise and academic support. This dual mentoring approach ensures a well-rounded and rigorous OJT experience for every student in the program.

E. Submission of documentation for OJT

The student will make two documents as part of the OJT

- **Online diary:** This ensures that the student updates daily activity, which could be accessed by both the mentors. Daily entry can be of 3- 4 sentences giving a very brief account of the learning/activities/interaction taken place. The faculty mentor will be monitoring the entries in the diary regularly as shown in **Appendix-I**
- **OJT report:** A student is expected to make a report based on the OJT he or she has done in an organization. It should contain the following:
    - **Certificate**: A certificate in the prescribed Performa (given **in Appendix II** and **Appendix III**) from the organization where the OJT was done.
    - **Title**: A suitable title giving the idea about what work the student has performed during the OJT.
    - **Description of the organization**: A small description of the organization where the student has interned
    - **Description of the activities** done by the section where the intern has worked: A description of the section or cell of the organization where the intern worked. This should give an idea about the type of activity a new employee is expected to do in that section of the organization.
    - **Description of work allottedand done by the intern:** A detailed description of the work allotted, and actual work performed by the intern during the OJT period. It shall be the condensed and structured version of the daily report mentioned in the online diary.
    - **Self-assessment**: A self-assessment by the intern on what he or she has learned during the OJT period. It shall contain both technical as well as interpersonal skills learned in the process.

F. Interaction between mentors:

To ensure the smooth conduct of the OJT a meet-up involving the intern, industry mentor, and the faculty mentor will be scheduled as a mid-term review. The meeting can preferably be online to save time and resources. The meeting ensures the synergy between all stakeholders of the OJT. A typical meeting can be of around 15 minutes where at the initial stage the intern brief about the work and interaction goes for about 10 minutes. This can be followed by the interaction of the mentors in the absence of the intern. This ensures that issues between the intern and the organization, if any, are resolved amicably.

G. OJT workload for the faculty:

Every student is provided with a faculty member as a mentor. So, a faculty mentor will have a few students under him/her. A faculty mentor is the overall in charge of the OJT of the student. He/she constantly monitors the progress of the OJT by regularly overseeing the diary, interacting with the industry mentor, and guiding on the report writing etc. Considering the time and effort involved, a faculty mentor who is in-charge of 20 students shall be provided by a workload of 3 hours.

EVALUATION SCHEME

A. Evaluation for Mandatory Theory Courses (4 Credit Courses)

I. Internal Evaluation for Mandatory Theory Courses – 50 Marks

(i) Mid-Term Class Test – 30 Marks
(ii) Assignment/ Case study– 20 Marks

OR

(i) SWAYAM (Advanced Course) of minimum 20 hours and certification exam completed – 50 Marks

OR

(ii) NPTEL (Advanced Course) of minimum 20 hours and certification exam completed - 50 Marks

OR

(iii) Valid International Certifications (Prometric, Pearson, Certiport, Coursera, Udemy and the like) - 50 Marks

One certification marks shall be awarded one course only. For four courses, the students will have to complete four certifications.

II. External Examination for Mandatory Theory Courses – 50 Marks

- Duration: **2 Hours**
- Theory question paper pattern:

	All questions are compulsory.		
Question	**Based on**	**Options**	**Marks**
Q.1	Unit I	*Any 2 out of 4*	10
Q.2	Unit II	*Any 2 out of 4*	10
Q.3	Unit III	*Any 2 out of 4*	10
Q.4	Unit IV	*Any 2 out of 4*	10
Q.5	Unit I, II,III & IV	*Any 2 out of 4*	10

B. Evaluation for Elective Theory Courses (2 Credit Courses)

I. Internal Evaluation for Elective Theory Courses – 25 Marks

(i) Mid-Term Class Test – 15 Marks
(ii) Assignment/ Case study– 10 Marks

II. External Examination for Elective Theory Courses – 25 Marks

- Duration: **1 Hour**
- Theory question paper pattern:

	All questions are compulsory.		
Question	**Based on**	**Options**	**Marks**
Q.1	Unit I	*Any 2 out of 4*	10
Q.2	Unit II	*Any 2 out of 4*	10
Q.3	Unit I & II	*Any 1 out of 2*	5

C. Evaluation for Mandatory & Elective Practical Courses (2 Credit Courses)

- Each Practical Course carries 50 Marks
  - **40 marks + 05 marks (journal) + 05 marks (viva)**
- Duration: **2 Hours**for each practical course.
- Minimum **80% practical** from each core subjects are required to be completed.
- Certified Journal is compulsory for appearing at the time of Practical Exam

D. Evaluation of On Job Training Course (4 Credit Course)

Internal Evaluation	
Online diary	25
Mid-term interaction	25
Total	**50**
External Evaluation	
OJT Documentation	25
Quality & Relevance	10
OJT Viva	15
Total	**50**

APPENDIX-I

Maintain the weekly online diary for each week in the following format.

	Day	Date	Name of the Topic/Module Completed	Remarks
1st WEEK	MONDAY			
	TUESDAY			
	WEDNESDAY			
	THRUSDAY			
	FRIDAY			
	SATURDAY			

Signature of the Faculty mentor: _____

Seal of the University/College

APPENDIX-II

(Proforma for the certificate for internship in official letter head)

This is to certify that Mr. /Ms. ……. of .....College/Institution worked as an intern as part of his/her M.Sc. course in Data Science of University ofMumbai. The particulars of internship are given below:

Internship starting date: _____

Internship ending date: _____

Actual number of days worked: _____

Tentative number of hours worked: _____ Hours

Broad area of work: _____

A small description of work done by the intern during the period:

_____

_____

Signature: _____

Name:

Designation:

Contact details:

Email:

(Seal of the organization)

APPENDIX-III

(Proforma for the Evaluation of the intern by the industry mentor /to whom the intern was reporting in the organization)

Professional Evaluation of intern

Name of intern: _____

College/institution: _____

*[**Note**: Give a score in the 1 to 5 scale by putting √ in the respective cells]*

No	Particular	Excellent	Very Good	Good	Moderate	Satisfactory
1	Attendance & Punctuality					
2	Ability to work in a team					
3	Written and oral communication skills					
4	Problem solving skills					
5	Ability to grasp new concepts					
6	Technical skill in terms of technology, programming etc					
7	Ability to complete the task					
8	Quality of overall work done					

Comments:

_____

Signature: _____

Name :

Designation:

Contact details:

Email :

(Seal of the organization)

Letter Grades and Grade Points

Semester GPA/ Program CGPA Semester/Program	% of Marks	Alpha-Sign / Letter Grade Result
9.00-10.00	90.0-100	O (Outstanding)
8.00-<9.00	80.0-<90.0	A+ (Excellent)
7.00-<8.00	70.0-<80.0	A (Very Good)
6.00-<7.00	60.0-<70	B+ (Good)
5.50-<6.00	55.0-<60.0	B (Above Average)
5.00-<5.50	50.0-<55.0	C (Average)
4.00-<5.00	40.0-<50.0	P (Pass)
Below 4.00	Below 40	F (Fail)
Ab (Absent)	-	Absent

Appendix-B

Justification for M.Sc. (Data Science)

1.	Necessity for starting the program:	Data Science is at the forefront of technological advancements, and research in this field can lead to groundbreaking discoveries and innovations. It is an interdisciplinary field that combines elements of computer science, mathematics, statistics, and domain-specific knowledge. Starting an M.Sc. program can foster research collaborations and contribute to advancements in data analysis techniques and applications. Thus, data science program can provide a holistic education, preparing students to tackle real-world challenges with a diverse skill set.
2.	Whether the UGC has recommended the program:	**Yes**
3.	Whether all the programs have commenced from the academic year 2023-24	**Yes**
4.	The programs started by the University are self-financed, whether adequate number of eligible permanent faculties are available?	**Yes** Guest Faculty / Adjunct Professors/ IT Professionals will be invited
5.	To give details regarding the duration of the program and is it possible to compress the program?	2 years. Not possible Its Four semester Course, introduction, and learning research skill (Research Methodology) are taught in the first semester; Problem solving skills and On-Job-Training are the part of the second semester. The third semester and fourth semester comprise of Advanced concepts and completing a research project. Exit is available after first year. Corresponding credits should be earned bythe learners.

6.	The intake capacity of each program and no. of admissions given in the current academic year:	120 seats 2023-2024 admission starts from July
7.	Opportunities of Employability / Employment available after undertaking these courses:	The demand for data science professionals is rapidly increasing across various industries due to the growing importance of data-driven decision-making. Here are some common opportunities such as Data Scientist, Business Intelligence (BI) Analyst, Data Visualization Specialist, Artificial Intelligence (AI) Specialist, Healthcare Data Analyst, Market Research Analyst available after completing the course in data science. These are just a few examples of the diverse career opportunities available after undertaking courses in data science. The versatility of data science skills allows professionals to work in various industries, contributing to data-driven decision-making and innovation.

DrJyotshnaDongardive
Head, Department of Computer Science
University of Mumbai

DrShivramGarje
Dean, Science and Technology
University of Mumbai

****