Anekant Education Society's

Tuljaram Chaturchand College

of Arts, Science and Commerce, Baramati

(Autonomous)

M.Sc. Degree Program in Data Science

(Faculty of Science & Technology)

CBCS Syllabus

M.Sc.(Data Science) Part – II Semester – III

For Department of Statistics

Tuljaram Chaturchand Collegeof Arts, Science and Commerce, Baramati

oice Based Credit System Syllabus (2023 Pattern)

(As Per NEP 2020)

To be implemented from Academic Year 2024-2025

Program Outcomes for M.Sc. (Data Science)

PO1. Comprehensive Knowledge and Understanding:

Postgraduates will possess a profound understanding of their field, encompassing foundational theories, methodologies, and key concepts within a multidisciplinary context.

PO2. Practical, Professional, and Procedural Knowledge:

Postgraduates will acquire practical skills and expertise necessary for professional tasks, including industry standards, regulations, and ethical considerations, with effective application in real-world scenarios.

PO3. Entrepreneurial Mindset, Innovation, and Business Understanding:

Postgraduates will cultivate an entrepreneurial mindset, identify opportunities, foster innovation, and understand business principles, market dynamics, and risk management strategies.

PO4. Specialized Skills, Critical Thinking, and Problem-Solving:

Postgraduates will demonstrate proficiency in technical skills, analytical abilities, effective communication, and leadership, adapting and innovating in response to changing circumstances.

PO5. Research, Analytical Reasoning, and Ethical Conduct:

Postgraduates will exhibit observational and inquiry skills, formulate research questions, utilize appropriate methodologies for data analysis, and adhere to research ethics while effectively reporting findings.

PO6. Communication, Collaboration, and Leadership:

Postgraduates will effectively communicate complex information, collaborate in diverse teams, demonstrate leadership qualities, and facilitate cooperative efforts toward common goals.

PO7. Digital Proficiency and Technological Skills:

Postgraduates will demonstrate proficiency in using ICT, accessing information sources, analyzing data using appropriate software, and adapting to technological advancements.

PO8. Multicultural Competence, Inclusive Spirit, and Empathy:

Postgraduates will engage effectively in multicultural settings, respect diverse perspectives, lead diverse teams, and demonstrate empathy and understanding of others' perspectives and emotions.

PO9. Value Inculcation, Environmental Awareness, and Ethical Practices:

Postgraduates will embrace ethical and moral values, practice responsible citizenship, recognize and address ethical issues, and promote sustainability and environmental conservation.

PO10. Autonomy, Responsibility, and Accountability:

Postgraduates will apply knowledge and skills independently, manage projects effectively, and demonstrate responsibility and accountability in work and learning contexts, contributing to societal well-being.

Anekant Education Society's
Tuljaram Chaturchand College,Baramati
(Autonomous)

Board of Studies (BOS) in Statistics

From 2022-23 to 2024-25

Sr.No.	Name	Designation
1.	Prof. Dr. Vikas C. Kakade	Chairman
2.	Prin. Dr. Avinash S. Jagtap	Member
3.	Dr. Neeta K. Dhane	Member
4.	Dr.Vaishali V. Patil	Member
5.	Mrs. Sarita D. Wadkar	Member(Ad hoc)
6.	Mr. Chandrashekhar P. Swami	Member
7.	Ms. Priti M. Mohite	Member(Ad hoc)
8.	Ms. Nilambari A. Jagtap	Member (Ad hoc)
9.	Miss.Kalyani C. Kale	Member (Ad hoc)
10.	Ms.Pooja S. Zanjurne	Member (Ad hoc)
11.	Dr.Akanksha S. Kashikar	Vice-Chancellor Nominee
12.	Prin. Dr.Rajendra G. Gurao	Expert from other University
13.	Mr. RohanKoshti	Expert from other University
14.	Mr. Saurabh Kadam	Industry Expert
15.	Dr. Jaya L. Limbore	Meritorious Alumni
16.	Miss. Priya N. Rakate	Invitee Member
17.	Ms. Ankita G. Deshmukh	Invitee Member
18.	Ms. Shital B. Choudhar	Invitee Member

19.	Miss. Kiran Banda (M.Sc. II)	Student Representative
20.	Mr. RushikeshPandhare (M.Sc. II)	Student Representative
21.	Mr. Bharat Jambhulkar (TYBSc)	Student Representative
22.	Miss. Prapti Mane (TYBSc)	Student Representative

Credit Distribution Structure for M.Sc.(Data Science)Part-I

Level	Semester	Major		Research Methodology (RM)	OJT/FP	RP	Cum. Cr.	Degree
		Mandatory	Electives					
6.0	Sem-I	DSC-501-MJM: Probability and Statistics for Data Science(Credit 04) DSC-502-MJM: Data Analytics Using R (Credit 04) DSC -503-MJM: Data Science Practical – I (Credit 02) DSC-504-MJM: Data Science Practical – II (Credit 02)	DSC-511-MJE (A): Data Base Management System DSC -511-MJE (B): Stochastic Models and Applications (Credit 04)	DSC -521-RM: Research Methodology (Credit 04)	--	--	20	PG Diploma (after 3 Year Degree)
	Sem-II	DSC -551-MJM: Machine Learning and Artificial intelligence(Credit 04) DSC-552-MJM: Regression Analysis and Predictive Models (Credit 04) DSC-553-MJM: Data Science Practical – III (Credit 02) DSC -554-MJM: Data Science Practical – IV (Credit 02)	DSC -561-MJE (A): Bayesian Inference DSC -561-MJE (B): Computational Statistics (Credit 04)	--	DSC-581-OJT/FP: On Job Training/ Field Project	--	20	

Credit Distribution Structure for M.Sc.(Data Science) Part-II

Level	Semester	Major		Research Methodology (RM)	OJT /FP	RP	Cum. Cr.	Degree
		Mandatory	Electives					
6.5	Sem-III	DSC-601-MJM: Multivariate Analysis (Credit 04)	DSC-611-MJE(A): Business and Project Management (Credit 02)	--	--	DSC-621-RP: Research Project (Credit 04)	20	PG Diploma (after 3 Year Degree)
		DSC-602-MJM: Time Series Analysis and Forecasting (Credit 04)	DSC-611-MJE(B): Text Mining and Natural Language Processing (Credit 02)					
		DSC-603-MJM: Statistics Practical – V (Credit 02)	DSC-612-MJE(A):Practical Based on Business and Project Management (Credit 02)					
		DSC-604-MJM: Statistics Practical – VI (Credit 02)	DSC-612-MJE(B): Practical Based on Text Mining and NLP (Credit 02)					
	Sem-IV	DSC-651-MJM: Introduction to Generative AI and Large Language Models (Credit 04)	DSC-661-MJE (A): Supply Chain and Logistics Analytics (Credit 02)	--	--	DSC-621-RP: Research Project (Credit 06)	20	
		DSC-652-MJM: Advanced Machine Learning and Deep Learning (Credit 04)	DSC-661-MJE(B): Image Processing and Computer Vision (Credit 02)					
		DSC-653-MJM: Data Science Practical – VI (Credit 02)	DSC-662-MJE (A): Introduction to Hadoop (Credit 02)					
			DSC-662-MJE (B): Web Application Development (Credit 02)					

Course Structure for M.Sc. Part-I (Data Science)(2023 Pattern)

Sem	Course Type	Course Code	Course Title	Theory/ Practical	No. of Credits
I	Major (Mandatory)	DSC-501-MJM	Probability Distributions	Theory	04
	Major (Mandatory)	DSC -502-MJM	Statistical Inference	Theory	04
	Major (Mandatory)	DSC -503-MJM	Data Science Practical – I	Practical	02
	Major (Mandatory)	DSC -504-MJM	Data Science Practical – II	Practical	02
	Major (Elective)	DSC-511-MJE (A)	Data Base Management System	Theory	04
		DSC -511-MJE (B)	Stochastic Models and Applications	Theory	
	Research Methodology (RM)	DSC -521-RM	Research Methodology	Theory	04
			Total Credits Semester I		**20**
II	Major (Mandatory)	DSC -551-MJM	Machine Learning and Artificial intelligence	Theory	04
	Major (Mandatory)	DSC-552-MJM	Regression Analysis and Predictive Models	Theory	04
	Major (Mandatory)	DSC-553-MJM	Data Science Practical – III	Practical	02
	Major (Mandatory)	DSC-554-MJM	Data Science Practical – IV	Practical	02
	Major (Elective)	DSC -561-MJE (A)	Bayesian Inference	Theory	04
		DSC -561-MJE (B)	Computational Statistics	Theory	
	On Job Training (OJT)/Field Project (FP)	DSC -581-OJT/FP	On Job Training Field Project	Training/Project	04
			Total Credits Semester-II		**20**
			Cumulative Credits Semester I and II		**40**

Course Structure for M.Sc. Part-II (Data Science) (2023 Pattern)

Sem	Course Type	Course Code	Course Title	Theory/ Practical	No. of Credits
III	Major (Mandatory)	DSC-601-MJM	Multivariate Analysis	Theory	04
	Major (Mandatory)	DSC -602-MJM	Time Series Analysis and Forecasting	Theory	04
	Major (Mandatory)	DSC -603-MJM	Data Science Practical – V	Practical	02
	Major (Mandatory)	DSC -604-MJM	Data Science Practical – VI	Practical	02
	Major (Elective)	DSC-611-MJE(A)	DSC-611-MJE(A): Business and Project Management (Credit 02)	Theory	02
		DSC-611-MJE(B)	DSC-611-MJE(B): Text Mining and Natural Language Processing (Credit 02)	Theory	
		DSC-612-MJE(A)	DSC-612-MJE(A):Practical Based on Business and Project Management (Credit 02)	Practical	02
		DSC-612-MJE(B):	DSC-612-MJE(B): Practical Based on Text Mining and NLP (Credit 02)	Practical	
	Research Project (RP)	DSC -621-RP	Research Project	Project	04
				Total Credits Semester III	**20**
IV	Major (Mandatory)	DSC -651-MJM	Introduction to Generative AI and Large Language Models	Theory	04
	Major (Mandatory)	DSC-652-MJM	Advanced Machine Learning and Deep Learning	Theory	04
	Major (Mandatory)	DSC-653-MJM	Data Science Practical – VI	Practical	02
	Major (Elective)	DSC-661-MJE (A)	Supply Chain and Logistics Analytics	Theory	02
		DSC-661-MJE(B)	Image Processing and Computer Vision	Theory	
		DSC-662-MJE (A)	Introduction to Hadoop	Practical	02
		DSC-662-MJE (B)	Web Application Development	Practical	
	Research Project (RP)	DSC -581-RP	Research Project	Project	06

	Total Credits Semester-IV	**20**
	Cumulative Credits Semester III and IV	**40**

CBCS Syllabus as per NEP 2020 for M.Sc. Part-II Data Science (2023 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – II
Semester	: III
Course Type	:Major Mandatory Theory
Course Name	: Multivariate Analysis
Course Code	: DSC-601-MJM
No. of Credits	: 4
No. of Teaching Hours	: 60

Course Objectives:

1. To develop feasible solution of real-life problems, using multivariate methods and techniques.
2. To develop an understanding of appropriate and relevant methods of multivariate data analysis.
3. To use visualization methods adapted to multidimensional exploratory analysis.
4. To recognize the method adapted to the exploration of a dataset according to the nature and structure of the variables.
5. To introduce fundamental concepts and principles of multivariate data analysis.
6. To develop proficiency in performing exploratory data analysis techniques for multivariate data.
7. To develop critical thinking skills for interpreting and communicating results from multivariate analysis.

Course Outcomes:

By the end of the course, students will be able to:

CO1. carry out an extensive exploratory multivariate analysis for a given multivariate data carry out cluster analysis of given multivariate data.

CO2. Apply exploratory multivariate data analysis techniques including graphical representation, mean, variance, covariance, and correlation.

CO3. Understand the properties of multivariate normal distribution including mean, variance, independence, and sampling.apply the concepts of linear and quadratic forms in multivariate normal variables.

CO4. Conduct discriminant analysis for two-group and three-group scenarios.

CO5. Understand the decision process of discriminant analysis including model estimation, interpretation, and validation.

CO6. carry out statistical inference procedures using the data from a multivariate normal distribution.

CO7. Estimate parameters and test linear hypotheses in multivariate linear models using maximum likelihood estimation.

Topics and Learning Points

OPICS/CONTENTS:

Unit1

Exploratory Multivariate Data Analysis & Dimensionality Reduction

Introduction to Multivariate Data: Structure, Pre-processing, and Handling Missing Values Diagrammatic Representations: Scatter plot Matrix, Heatmaps, Pairplots, Sample Mean Vector, Sample Dispersion Matrix, Sample Correlation Matrix, Graphical Representation of Multivariate Data (Bubble Charts, Parallel Coordinates), Effects of Linear Transformations: Mean, Variance, Covariance, and Correlation, Six-Step Approach to Multivariate Model Building

Dimensionality Reduction Techniques:

1) Principal Component Analysis (PCA):Covariance and Correlation-based PCA, Standardized PCA and Interpretation of Components, Applications in Feature Engineering
2) Factor Analysis: Models, Eigen values, Eigenvectors, Rotation Techniques (Varimax, Promax),
3) Canonical Correlation Analysis (CCA):Real-life Applications in Finance, Marketing, and Healthcare. **[20L]**

Unit 2: Clustering & Multivariate Distributions[20L]

Clustering Methods in Data Science: Hierarchical Clustering (Agglomerative, Divisive), Linkage Methods: Single, Complete, Average, Ward's Method, Non-Hierarchical Clustering (K-Means, K-Medoids)

Multivariate Normal Distribution: Singular and Non-Singular Normal Distribution, Mean, Variance, Marginal, and Conditional Distributions, Independence of Variables, Moment Generating Functions (MGF), and Characteristic Functions, Maximum Likelihood Estimation

(MLE) for Multivariate Normal ParametersEstimators and their Sampling Distributions Random Sampling from Multivariate Normal Distributions.

Statistical Tests for Multivariate Data: Test for Multivariate Normality (Shapiro-Wilk, Anderson-Darling, QQ-Plots), Test for Equality of Dispersion Matrices (Box's M Test),

Unit3: Multivariate Hypothesis Testing & MANOVA

Wishart Distribution & Wishart Matrix: Properties and Applications in Covariance Estimation, Hypothesis Testing for Multivariate Data:Tests for Mean Vector of a Multivariate Normal Population, Test for Equality of Dispersion Matrices, Hotelling's T^2 Test:Distribution, Applications in Fraud Detection & Quality Control, Multivariate Analysis of Variance (MANOVA):Methods and Applications, Wilk's Lambda Statistic (without derivation), Applications in Customer Segmentation & Healthcare Studies **[15L]**

Unit4: Discriminant Analysis for Multi-Class Classification

Two-Group Discriminant Analysis (Linear and Quadratic Discriminant Analysis - LDA & QDA), Three-Group Discriminant Analysis and Decision Boundaries, Model Assumptions & Interpretation, Model Validation and Performance Metrics, Real-world Applications in Credit Scoring, Customer Retention, and Image Recognition. **[5L]**

References Books:

1. Anderson, T. W. (1984). Introduction to Multivariate Analysis, John Wiley.
2. Richard A. Johnson and Dean W. Wichern, Applied Multivariate Statistical Analysis,
3. Prentice hall India, 7th Edition, 2019.
4. Fang, K., Kotz, S., Ng K. W. (1990). Symmetric Multivariate and Related Distributions, Chapman and Hall
5. Härdle, W. K. &Simar, L. (2012). Applied Multivariate Statistical Analysis, Springer, New York
6. Härdle, W. K., Hlávka, Z. (2007). Multivariate Statistics: Exercises andSolutions,
7. Springer, New York
8. Kotz, S., Balakrishnan N. and Johnson N. L. (2000). Continuous Multivariate Distributions,Volume 1, Models and Applications, John Wiley & Sons,
9. Kshirsagar, A. M. (1983). Multivariate Analysis, Marcel Dekker
10. Morrison, D.F. (1990). Multivariate Statistical Methods, McGraw Hill Co.

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	Programme Outcomes (POs)									
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	1	3	1	2	2	1	1	1	1	1
CO2	2	3	1	3	2	1	2	1	1	1
CO3	2	1	1	3	2	1	2	1	1	1
CO4	2	2	1	3	2	1	2	1	1	1
CO5	2	2	1	3	2	1	2	1	1	1
CO6	2	1	1	2	2	1	1	1	1	1
CO7	2	3	1	3	2	1	3	1	1	2

Weight: 1 - Partially related2 - Moderately Related 3 - Strongly related

PO1. Comprehensive Knowledge and Understanding:

CO1: Partially Related (1)

CO2: Moderately Related (2)

CO3: Moderately Related (2)

CO4: Moderately Related (2)

CO5: Moderately Related (2)

CO6: Moderately Related (2)

CO7: Moderately Related (2)

Justification: The course outcomes align with PO1 by providing students with comprehensive knowledge and understanding of exploratory multivariate analysis techniques, multivariate normal distribution properties, discriminant analysis, statistical inference procedures, and multivariate linear models. While some outcomes directly address the depth of understanding (e.g., CO1), others contribute indirectly by building foundational knowledge (e.g., CO2-CO7).

PO2. Practical, Professional, and Procedural Knowledge:

CO1: Strongly Related (3)

CO2: Strongly Related (3)

CO3: Partially Related (1)

CO4: Moderately Related (2)

CO5: Moderately Related (2)

CO6: Partially Related (1)

CO7: Strongly Related (3)

Justification: CO1 and CO2 directly contribute to practical, professional, and procedural knowledge by involving hands-on application of exploratory multivariate analysis techniques and cluster analysis. CO3, CO4, CO5, CO6, and CO7, while important for building a deeper understanding, are not as directly tied to practical applications.

PO3. Entrepreneurial Mindset, Innovation, and Business Understanding:

CO1: Partially Related (1)

CO2: Partially Related (1)

CO3: Partially Related (1)

CO4: Partially Related (1)

CO5: Partially Related (1)

CO6: Partially Related (1)

CO7: Partially Related (1)

Justification: While some aspects of the course, such as understanding data patterns and relationships, may indirectly contribute to entrepreneurial mindset and innovation (e.g., CO1, CO2), the direct linkage to business understanding or fostering entrepreneurial mindset is not explicitly emphasized in the course outcomes.

PO4. Specialized Skills, Critical Thinking, and Problem-Solving:

CO1: Moderately Related (2)

CO2: Strongly Related (3)

CO3: Strongly Related (3)

CO4: Strongly Related (3)

CO5: Strongly Related (3)

CO6: Moderately Related (2)

CO7: Strongly Related (3)

Justification: The course outcomes focus extensively on developing specialized skills in multivariate data analysis, critical thinking, and problem-solving, including applying various techniques, understanding distribution properties, conducting discriminant analysis, and performing statistical inference procedures.

PO5. Research, Analytical Reasoning, and Ethical Conduct:

CO1: Moderately Related (2)

CO2: Moderately Related (2)

CO3: Moderately Related (2)

CO4: Moderately Related (2)

CO5: Moderately Related (2)

CO6: Moderately Related (2)

CO7: Moderately Related (2)

Justification: While the course outcomes involve research and analytical reasoning in the context of multivariate data analysis, they do not explicitly address ethical conduct. However, the skills developed in analyzing data and drawing conclusions contribute to analytical reasoning, which is essential for research.

PO6. Communication, Collaboration, and Leadership:

CO1: Partially Related (1)

CO2: Partially Related (1)

CO3: Partially Related (1)

CO4: Partially Related (1)

CO5: Partially Related (1)

CO6: Partially Related (1)

CO7: Partially Related (1)

Justification: The course outcomes primarily focus on technical skills related to data analysis and statistical modeling, with less emphasis on communication, collaboration, or leadership skills.

PO7. Digital Proficiency and Technological Skills:

CO1: Partially Related (1)

CO2: Moderately Related (2)

CO3: Moderately Related (2)

CO4: Moderately Related (2)

CO5: Moderately Related (2)

CO6: Partially Related (1)

CO7: Strongly Related (3)

Justification: The course outcomes involve the application of statistical techniques using statistical software, contributing to digital proficiency and technological skills. However, there is room for improvement in directly addressing digital proficiency in all outcomes.

PO8. Multicultural Competence, Inclusive Spirit, and Empathy:

CO1: Partially Related (1)

CO2: Partially Related (1)

CO3: Partially Related (1)

CO4: Partially Related (1)

CO5: Partially Related (1)

CO6: Partially Related (1)

CO7: Partially Related (1)

Justification: The course outcomes focus on technical aspects of data analysis, with minimal emphasis on multicultural competence, inclusive spirit, or empathy.

PO9. Value Inculcation, Environmental Awareness, and Ethical Practices:

CO1: Partially Related (1)

CO2: Partially Related (1)

CO3: Partially Related (1)

CO4: Partially Related (1)

CO5: Partially Related (1)

CO6: Partially Related (1)

CO7: Partially Related (1)

Justification: While ethical conduct is important in research and data analysis, the course outcomes do not explicitly address environmental awareness or broader value inculcation.

PO10. Autonomy, Responsibility, and Accountability:

CO1: Partially Related (1)

CO2: Moderately Related (2)

CO3: Moderately Related (2)

CO4: Moderately Related (2)

CO5: Moderately Related (2)

CO6: Moderately Related (2)

CO7: Moderately Related (2)

Justification: The course outcomes involve independent application of statistical techniques and responsibility in interpreting and communicating results, contributing to autonomy,

responsibility, and accountability. However, there is room for improvement in directly addressing these aspects in all outcomes.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-II Data Science (2023 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – II
Semester	: III
Course Type	:Major Mandatory Theory
Course Name	: Time Series Analysis and Forecasting
Course Code	: DSC -602-MJM
No. of Credits	: 4
No. of Teaching Hours	: 60

Course Objectives:

Students successfully completing this course will be able to:

1. The main objective of this course is that students should understand various time series models, estimation of its parameters and be able to make predictions.
2. To learn the concepts like Auto-covariance, auto-correlation function and vector auto regression.
3. Combine AR and MA processes to create and analyze AutoRegressive Moving Average (ARMA) models.
4. Understand the concepts of causality and invertibility in time series models.
5. Understand and apply Seasonal ARIMA (SARIMA) models for seasonal time Series forecasting.
6. Explore the innovation algorithm and its applications in time series analysis.
7. Understand the Box-Jenkins methodology and its application in developing time series forecasting models.

Course Outcomes:

By the end of the course, students will be able to:

CO1. Understand fundamental concepts of time series analysis, including its components and decomposition techniques.

CO2. Perform exploratory data analysis (EDA) on time series data using graphical and statistical methods.

CO3. Implement smoothing techniques such as moving averages, exponential smoothing, and Holt-Winters methods.

CO4. Analyze stationary and non-stationary time series data using AR, MA, and ARIMA models.

CO5. Develop and evaluate forecasting models using AIC, BIC, and AICC selection criteria.

CO6. Utilize machine learning techniques for anomaly detection and predictive modeling in time series data.

CO7. Evaluate the effectiveness of different forecasting models and improve predictions based on real-time data analysis.

Topics and Learning Points

Unit 1

Exploratory analysis of Time Series, Graphical display, classical decomposition model, components and various decompositions of Time Series Models-Numerical description of Time Series: Stationarity, Auto-covariance and Autocorrelation functions, data transformations, Methods of estimation, trend, seasonal and exponential. Smoothing Techniques, Moving Average, exponential smoothing, Holt's and Winter's methods, exponential smoothing techniques for Series with trend and seasonality, basic evaluation of exponential smoothing. **(15)**

Unit 2

Stationary process: General linear process, stationary process and strict stationary process, moving average (MA), Auto Regressive (AR) and autoregressive moving average (ARMA). ARIMA Models: Basic formulation of the ARIMA Model and their statistical properties, Autocorrelation function (ACF), Partial autocorrelation function (PACF)and their standard Errors. **(15)**

Unit 3

Non-stationary: Unit root, non-stationary unit root test, Integrated ARMA (ARIMA) model, Analysis of seasonal models: Parsimonious models for seasonal time series, SARIMA models, forecasting, identification, estimation and diagnosis methods for seasonal time series Forecasting: Nature of Forecasting, Forecasting methods, qualitative and quantitative methods, steps involved in stochastic model building, forecasting model evaluation Model selection techniques: AIC, BIC and AICC – Forecasting model monitoring, ARCH and GARCH properties, examples. **(15)**

Unit 4

Applications of Time Series in Data Science

Financial Time Series Analysis: Stock Market Prediction, Risk Modeling, Portfolio Optimization, Economic and Business Forecasting: Sales Forecasting, Demand Prediction, Revenue Forecasting, Climate and Environmental Time Series: Weather Forecasting, Climate Change Analysis, Anomaly Detection in Time Series Data using Statistical and Machine Learning Techniques, Practical Implementation of Time Series Models using Python/R,Hands-on Sessions using Libraries: Statsmodels, Prophet, TensorFlow/Keras for Time Series Analysis

(15)

Reference Books:

1. Brockwell, P.J. and Davis, R. A. *Introduction to Time Series Analysis*, Springer.
2. Chatfield, C. (2001). *Time Series Forecasting*, Chapmann& hall, London.
3. Fuller, W. A. (1996). *Introduction to Statistical Time Series*, 2nd Ed. John Wiley.
4. Hamilton N. Y. (1994). *Time Series Analysis*. Princeton University press. Princeton.
5. Kendall, Sir Maurice and Ord, J. K. (1990).Time Series (Third Edition), Edward Arnold.

Programme Outcomes and Course Outcomes Mapping:

CO-PO Mapping Table:

Course Outcomes (COs)	Programme Outcomes (POs)									
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	3	2	2	3	2	2	3	1	1	1
CO2	2	3	1	3	2	3	2	1	1	1
CO3	2	2	2	3	2	2	3	1	1	1
CO4	2	3	2	3	2	3	2	1	1	1
CO5	2	3	2	3	3	2	3	1	1	1
CO6	2	3	2	3	2	2	2	1	1	1
CO7	2	3	2	3	2	3	2	1	1	1

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

Justification

PO1: Comprehensive Knowledge and Understanding

CO1: Applying time series models and estimating parameters requires a profound understanding of foundational theories and methodologies in time series analysis. **(3)**

CO2: Analyzing auto-covariance and auto-correlation functions involves deep knowledge of time series dependencies and vector auto-regression models. **(2)**

CO3: Combining AR and MA processes to create ARMA models requires comprehensive knowledge of time series models. **(2)**

CO4: Evaluating causality and invertibility concepts requires a thorough understanding of time series models. **(2)**

PO2: Practical, Professional, and Procedural Knowledge

CO1: Applying time series models to real-world datasets involves practical skills and industry-standard procedures. **(2)**

CO2: Analyzing and implementing vector auto-regression models requires practical and professional knowledge in time series analysis. **(3)**

CO3: Creating ARMA models and analyzing their performance involves practical application and procedural knowledge. **(2)**

CO4: Evaluating causality and invertibility involves practical application to improve model validity. **(3)**

CO5: Implementing SARIMA models for forecasting seasonal patterns involves practical skills in time series analysis. **(3)**

CO6: Applying the innovation algorithm requires practical knowledge and problem-solving skills. **(3)**

CO7: Developing forecasting models using Box-Jenkins methodology involves practical application of this technique. **(3))**

PO3: Entrepreneurial Mindset, Innovation, and Business Understanding

CO1: Applying time series models in business forecasting enhances decision-making skills. **(2)**

CO2: Understanding auto-regressive models helps in financial and economic applications. **(1)**

CO3: Exploring innovative forecasting methods contributes to business intelligence. **(2)**

CO4: Identifying model deficiencies and applying corrective measures supports innovation in analytics. **(2)**

CO5: Evaluating predictive accuracy of models fosters an entrepreneurial approach to data-driven decision-making. **(2)**

CO6: Applying statistical tools for business forecasting strengthens industry-relevant innovation. **(2)**

CO7: Developing robust time series models contributes to innovative business solutions. **(3)**

PO4: Research and Problem-Solving Skills

CO1: Applying theoretical and computational methods enhances research capabilities in time series modeling. **(3)**

CO2: Implementing statistical techniques to analyze dependencies fosters critical problem-solving skills. **(3)**

CO3: Developing customized forecasting models improves the ability to solve complex analytical problems. **(3)**

CO4: Applying inferential techniques in time series models enhances research-oriented skills. **(3)**

CO5: Evaluating real-world datasets with innovative techniques improves analytical problem-solving. **(3)**

CO6: Implementing advanced model selection strategies enhances research and problem-solving capabilities. **(3)**

CO7: Applying statistical methods to diverse datasets fosters the ability to approach complex research challenges. **(3)**

PO5: Data-Driven Decision Making

CO1: Applying time series models to decision-making processes enhances the ability to interpret data for strategic purposes. **(2)**

CO2: Evaluating statistical dependencies in forecasting supports data-driven insights. **(2)**

CO3: Utilizing statistical tools to optimize predictive models enhances decision-making skills. **(2)**

CO4: Comparing different forecasting models aids in selecting the most suitable approach for decision-making. **(2)**

CO5: Using SARIMA models for forecasting supports effective planning and strategy. **(3)**

CO6: Applying time series diagnostics enhances the ability to interpret business and economic trends. **(2)**

CO7: Developing advanced forecasting techniques supports strategic decision-making in organizations. **(2)**

PO6: Technical and Software Proficiency

CO1: Implementing time series models using statistical software enhances computational skills. **(2)**

CO2: Applying vector auto-regression in specialized software strengthens technical expertise. **(3)**

CO3: Utilizing software tools for ARMA model development fosters proficiency in data analytics. **(2)**

CO4: Conducting causality tests using statistical packages enhances technical competency. **(3)**

CO5: Implementing SARIMA models in software platforms develops hands-on computational skills. **(2)**

CO6: Using advanced statistical tools for innovation algorithm applications strengthens software proficiency. **(2)**

CO7: Automating forecasting models using programming languages enhances technical expertise. **(3)**

PO7: Ethical and Professional Responsibilities

CO1: Ensuring the ethical use of statistical models in decision-making fosters professional integrity. **(3)**

CO2: Applying responsible data handling practices enhances ethical awareness in analytics. **(2)**

CO3: Ensuring transparency in forecasting methodologies promotes ethical research. **(2)**

CO4: Evaluating model biases and their implications enhances ethical responsibility in data science. **(2)**

CO5: Conducting reliable time series analysis supports professional accountability. **(3)**

CO6: Applying validation techniques ensures ethical research integrity in modeling. **(2)**

CO7: Ensuring reproducibility in forecasting models enhances ethical research standards. **(2)**

PO8: Lifelong Learning and Adaptability

CO1: Adapting to evolving time series methodologies fosters continuous learning. **(1)**

CO2: Exploring new statistical techniques enhances lifelong learning in data analytics. **(1)**

CO3: Engaging with recent developments in forecasting methodologies supports adaptability. **(1)**

CO4: Learning about emerging trends in predictive modeling fosters a growth mindset. **(1)**

CO5: Adapting to new data challenges enhances lifelong analytical skills. **(2)**

CO6: Implementing new forecasting techniques supports continuous professional development. **(2)**

CO7: Keeping up with innovations in statistical modeling fosters adaptability. **(2)**

PO9: Communication and Presentation Skills

CO1: Effectively presenting time series modeling results enhances communication skills in statistical analysis. **(1)**

CO2: Explaining auto-regression and vector auto-regression models improves the ability to convey complex statistical concepts. **(1)**

CO3: Communicating ARMA modeling techniques strengthens professional presentation skills. **(1)**

CO4: Clearly articulating causality and invertibility concepts supports effective knowledge transfer. **(1)**

CO5: Presenting findings from SARIMA models enhances the ability to communicate statistical results. **(1)**

CO6: Explaining the innovation algorithm in forecasting models develops clarity in communication. **(1)**

CO7: Presenting forecasting results using Box-Jenkins methodology improves professional communication. **(2)**

PO10: Leadership and Teamwork

CO1: Collaborating on time series modeling projects fosters teamwork and leadership in statistical research. **(1)**

CO2: Working in teams to analyze auto-regressive models enhances collaborative problem-solving skills. **(1)**

CO3: Leading discussions on ARMA modeling techniques develops leadership abilities. **(1)**

CO4: Contributing to group research on causality and invertibility promotes teamwork in data analysis. **(1)**

CO5: Working collaboratively on SARIMA forecasting projects strengthens cooperative learning. **(1)**

CO6: Engaging in team-based analysis of the innovation algorithm enhances group decision-making. **(1)**

CO7: Leading group projects on forecasting techniques fosters leadership and communication skills. **(2)**

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – II
Semester	: III
Course Type	:Major Mandatory Practical
Course Name	: Data Science Practical – V
Course Code	: DSC-603-MJM
No. of Credits	: 2
No. of Teaching Hours	: 60

Course Objectives:

1. Students will develop a deep understanding of what multivariate data is, including the characteristics and challenges associated with analyzing data with multiple variables.
2. Students will learn techniques for exploring multivariate data, including data visualization, summary statistics, and data transformation.
3. Students will learn how to perform and interpret MANOVA, which extends the analysis of variance to multiple dependent variables.
4. Students will learn how to effectively explore and visualize multivariate data to get a sense of its complexity and characteristics.
5. Students will learn techniques for data cleaning, handling missing values, and transforming variables to make them suitable for multivariate analysis.
6. Students will learn how to apply Exploratory Multivariate Data Analysis (EMDA) techniques to real-world datasets through practical exercises and projects.
7. Students will learn knowledge about multivariate probability distributions, covariance, and correlation, and their significance in analyzing multivariate data.

Course Outcomes:

By the end of the course, students will be able to:

CO1. Understand the link between multivariate techniques and corresponding univariate techniques.

CO2. Analyze multivariate data and the dependence structure of variates to extract the useful information from a massive dataset.

CO3. Apply suitable tools for exploratory data analysis, dimension reduction, and classification to formulate and solve real-life problems.

CO4. Analyze multivariate data using data reduction techniques like principal component analysis, factor analysis.

CO5. Explore methods for reducing the dimensionality of multivariate data while preserving important information. Techniques like Principal Component Analysis (PCA) or Factor analysis are often covered.

CO6. Gain knowledge of discriminant analysis techniques, including Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA), for classification and dimensionality reduction.

CO7. Learning how to use discriminant analysis to classify observations into different groups based on their multivariate characteristics.

Topics and Learning Points

Sr. No.	Title of Experiments
1.	Exploratory multivariate data analysis
2.	Testing multivariate normality
3.	Model sampling from multivariate normal distribution and computation of M.L.E.'s of parameters.
4.	Principal component analysis (PCA)
5.	Factor analysis
6.	Hierarchical clustering analysis
7.	Non-Hierarchical clustering analysis
8.	Canonical correlation techniques
9.	Application of HotellingT^2 statistics
10.	Discriminant analysis
11	Multivariate analysis of variance
12	Mini project on Multivariate data analysis

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	Programme Outcomes (POs)									
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	3	-	-	-	-	-	-	-	-	-
CO2	3	-	-	-	-	-	-	-	-	-
CO3	-	3	-	-	-	-	-	-	-	-
CO4	-	3	-	-	-	-	-	-	-	-
CO5	-	-	3	-	-	-	-	-	-	-
CO6	-	-	3	-	-	-	-	-	-	-
CO7	-	-	-	3	3	-	-	-	-	-

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

PO1. Comprehensive Knowledge and Understanding:

CO1: (Weightage: 3)

Justification: This outcome directly aligns with the goal of gaining comprehensive knowledge and understanding by connecting multivariate techniques with their univariate counterparts, enhancing the depth of understanding of statistical methodologies.

CO2: (Weightage: 3)

Justification: Analyzing multivariate data requires a comprehensive understanding of the dependence structure among variables, demonstrating a high level of knowledge and understanding of complex data structures.

PO2. Practical, Professional, and Procedural Knowledge:

CO3: (Weightage: 3)

Justification: This outcome reflects the practical application of knowledge in using multivariate analysis techniques to address real-world problems, demonstrating practical and procedural knowledge in data analysis.

CO4: (Weightage: 3)

Justification: Utilizing data reduction techniques such as PCA and factor analysis requires practical knowledge in implementing these methods to handle large and complex datasets effectively.

PO4. Specialized Skills, Critical Thinking, and Problem-Solving:

CO5: (Weightage: 3)

Justification: Dimensionality reduction techniques like PCA and factor analysis involve critical thinking and problem-solving skills to balance the reduction of dimensionality while retaining essential information from the data.

CO6: (Weightage: 3)

Justification: Learning discriminant analysis techniques enhances specialized skills in classification and dimensionality reduction, requiring critical thinking to select appropriate methods for specific analytical tasks.

PO5. Research, Analytical Reasoning, and Ethical Conduct:

CO7: (Weightage: 3)

Justification: Applying discriminant analysis involves analytical reasoning to classify observations ethically and accurately, aligning with the principles of ethical conduct in research and analysis.

PO6. Communication, Collaboration, and Leadership:

(No direct mapping for this learning outcome)

PO8. Multicultural Competence, Inclusive Spirit, and Empathy:

(No direct mapping for this learning outcome)

PO9. Value Inculcation, Environmental Awareness, and Ethical Practices:

(No direct mapping for this learning outcome)

PO10. Autonomy, Responsibility, and Accountability:

(No direct mapping for this learning outcome)

These mappings demonstrate how each course outcome (CO) relates to specific program outcomes (PO) with varying degrees of strength, emphasizing the alignment between the course curriculum and the overarching goals of the program.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-II Data Science (2023 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – II
Semester	: III
Course Type	:Major Mandatory Practical
Course Name	: Data Science Practical – VI
Course Code	: DSC-604-MJM
No. of Credits	: 2
No. of Teaching Hours	: 40

Course Objectives:

1. To understand the fundamental concepts of time series analysis and explore its practical applications.
2. To decompose time series data into trend, seasonality, and residual components for better interpretation.
3. To apply smoothing techniques such as moving averages and exponential smoothing for noise reduction.
4. To understand and implement AR, MA, ARMA, ARIMA, and SARIMA models for time series forecasting.
5. To perform model selection using ACF/PACF, AIC, and BIC criteria.
6. To construct and evaluate forecasting models, including Holt-Winters and ARIMA-based models.
7. To apply theoretical concepts in a practical setting through a mini-project involving real-world datasets.

Course Outcomes:

By the end of the course, students will be able to:

CO1. Perform exploratory data analysis (EDA) on time series data to understand its structure and key patterns.

CO2. Decompose time series data into its trend, seasonality, and residual components for better interpretability.

CO3. Apply statistical tests to check for stationarity and transform data using differencing and Box-Cox techniques.

CO4. Use smoothing techniques such as moving averages and exponential smoothing to remove noise in time series.

CO5. Interpret and analyze ACF/PACF plots to determine time series dependencies and residual behavior.

CO6. Understand stationarity, causality, and invertibility conditions for ARMA models.

CO7. Select appropriate time series models using ACF/PACF, AIC, and BIC criteria.

Topics and Learning Points

Sr. No.	Title of Experiments
1.	Exploratory Data Analysis of Time Series.
2.	Decomposing Time Series Data into Trend, Seasonality, and Residual Components
3.	Testing and Ensuring Stationarity in Time Series Data
4.	Smoothing Time Series Using Moving Averages and Exponential Smoothing.
5.	Transforming Time Series Using Box-Cox, Differencing, and Checking Stationarity & Normality.
6.	ACF/PACF of series and residual analysis.
7.	Stationarity, causality and invertibility Of ARMA model
8.	Order selection in time series: use of ACF/PACF and AIC, BIC.
9.	Fitting of AR, MA models (conditional least squares or maximum likelihood)
10.	Forecasting using fitted linear models (recursively), Holt -Winters forecasts construction of forecast intervals
11.	Fitting heteroscedastic models: checking for heteroscedasticity from residuals, ARCH, GARCH modelling.
12.	Mini Project (3 Practical)

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes (COs)	Programme Outcomes (POs)									
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	3	2	1	3	3	2	2	1	1	2
CO2	3	2	1	3	3	2	2	1	1	2
CO3	3	2	1	3	3	2	3	2	2	2

Course Outcomes (COs)	Programme Outcomes (POs)									
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO4	3	3	2	3	3	3	3	2	2	2
CO5	3	3	2	3	3	2	3	2	2	2
CO6	3	3	2	3	3	2	3	2	2	2
CO7	3	3	2	3	3	3	3	2	2	2

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

Programme Outcomes (POs) and Justification:

PO1: Comprehensive Knowledge and Understanding

CO1, CO2, CO3, CO4, CO5, CO6, CO7, (3 - Strongly Related)

Justification: Understanding time series concepts, decomposition techniques, stationarity checks, and model fitting require a strong theoretical foundation. These topics develop knowledge of statistical methods, probability, and forecasting models, forming the core of time series analysis.

PO2: Practical, Professional, and Procedural Knowledge

CO1, CO2, CO3, CO4, CO5, CO6, CO7 (3 - Strongly Related)

Justification: The course covers practical aspects of time series modeling, including data preprocessing, decomposition, and forecasting techniques, which are widely applied in business, finance, and industry.

PO3: Analytical and Problem-Solving Skills

CO4, CO5, CO6, CO7, (2 - Moderately Related)

Justification: Students apply analytical thinking in model selection, residual analysis, and forecasting accuracy assessment. The ability to choose appropriate models based on statistical properties strengthens problem-solving skills.

PO4: Application of Computational Tools and Techniques

CO1, CO2, CO3, CO4, CO5, CO6, CO7 (3 - Strongly Related)

Justification: The implementation of time series models using computational tools such as R and Python enhances students' ability to apply software techniques in real-world scenarios.

PO5: Research and Development in Data Science

CO1, CO2, CO3, CO4, CO5, CO6, CO7 (3 - Strongly Related)

Justification: Understanding stationarity, causality, and heteroscedasticity contributes to research in statistical modeling, making students proficient in advanced data science applications.

PO6: Forecasting and Predictive Modeling Expertise

CO4, CO5, CO6, CO7 (3 - Strongly Related)

Justification: The ability to forecast future values using Holt-Winters, ARIMA, and SARIMA models is a critical skill in predictive analytics, which is applicable in various industries.

PO7: Ethical and Professional Responsibility

CO3, CO5, CO7 (2 - Moderately Related)

Justification: Ethical considerations in handling time series data, selecting appropriate forecasting models, and making responsible business decisions are emphasized throughout the course.

PO8: Lifelong Learning and Adaptability

CO4, CO5, CO6, CO7, (2 - Moderately Related)

Justification: Time series analysis is an evolving field, and students develop the ability to adapt to new trends, tools, and techniques in forecasting.

PO9: Communication and Interpretation of Results

CO1, CO2, CO3, CO4, CO5, CO6, CO7 (2 - Moderately Related)

Justification: Effective communication of analytical results, including interpretation of model accuracy, forecast intervals, and diagnostic tests, is essential for decision-making.

PO10: Teamwork and Project Management

CO1, CO2, CO3, CO4, CO5, CO6, CO7 (2 - Moderately Related)

Justification: The mini-project involves collaborative work, requiring students to apply time series analysis concepts to real-world problems, fostering teamwork and project management skills.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-II Data Science (2023 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – II
Semester	: III
Course Type	: Major Elective Theory
Course Name	: Business and Project Management
Course Code	: DSC-611-MJE(A)
No. of Credits	: 2
No. of Teaching Hours	: 30

CBCS Syllabus as per NEP 2020 for M.Sc. Part-II Data Science (2023 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – II
Semester	: III
Course Type	: Major Elective Theory
Course Name	: Business and Project Management
Course Code	: DSC-611-MJE(A)
No. of Credits	: 2
No. of Teaching Hours	: 30

Course Objectives:

1. Understand fundamental business concepts and their applications in data-driven decision-making.
2. Develop essential project management skills applicable to data science projects.
3. Implement project planning and resource management techniques.
4. Apply agile methodologies for effective management of data science projects.
5. Enhance communication skills to present data insights to business stakeholders.
6. Analyze and interpret business cases using data science approaches.
7. Execute a complete data science project from planning to presentation.

Course Outcomes:

By the end of the course, students will be able to:

CO1. Demonstrate understanding of business fundamentals and their relevance to data science.

CO2. Apply project management principles to design and manage data science projects.

CO3. Develop project scopes, timelines, and resource plans.

CO4. Implement agile and iterative development processes in data science applications.

CO5. Communicate data insights effectively using data storytelling techniques.

CO6. Prepare comprehensive business reports and project documentation.

CO7. Present and justify data science project plans and outcomes to stakeholders.

Topics and Learning Points

Unit 1 **(15 L)**

Introduction to Business Management (Business Fundamentals, Understanding Business Functions, Data-Driven Decision Making),

Project Management Principles (Introduction to Project Management, Project Management Methodologies, Defining Project Scope and Objectives, Stakeholder Management),

Planning and Execution of Data Science Projects(Project Planning, Resource Management, Risk Management, Budgeting and Cost Estimation).

Unit 2 **(15 L)**

Agile Data Science Project Management (Agile in Data Science, Sprints and Iterative Development, Monitoring Progress, Continuous Improvement)

Business Communication and Reporting (Effective Communication, Data Storytelling, Writing Reports and Business Proposals)

Capstone Project and Case Studies (Case Study Analysis, Capstone Project, Presentation)

References:

References Books:

1. *"The Data Science Handbook"* by Field Cady
2. *"Project Management for the Unofficial Project Manager"* by Kory Kogon
3. Agile Manifesto and Scrum Guides
4. Harvard Business Review Case Studies

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	Programme Outcomes (POs)									
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	3	3	-	3	-	-	-	-	-	-
CO2	3	3	-	3	-	-	3	-	-	-
CO3	2	2	-	-	-	-	-	-	-	1
CO4	2	-	-	3	3	3	-	-	-	-
CO5	2	-	-	3	3	-	-	-	-	-
CO6	2	2	-	-	3	-	3	-	-	2
CO7	2	2	-	-	-	3	3	-	-	-

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

PO1. Comprehensive Knowledge and Understanding:

CO1: (Weightage: 3)

CO2: (Weightage: 3)

CO3: (Weightage: 2)

CO4: (Weightage: 3)

CO5: (Weightage: 2)

CO6: (Weightage: 2)

CO7: (Weightage: 2)

CO1, CO2, CO3, CO4, CO5, CO6, CO7 are aligned with PO1 as they involve comprehensive understanding of business concepts, project management principles, and agile methodologies in data science.

PO2. Practical, Professional, and Procedural Knowledge:

CO1: (Weightage: 3)

CO2: (Weightage: 3)

CO3: (Weightage: 2)

CO6: (Weightage: 3)

CO7: (Weightage: 2)

CO1, CO2, CO3, CO6, and CO7 are directly linked to practical knowledge by applying project management and agile practices in real-world scenarios.

PO4. Specialized Skills, Critical Thinking, and Problem-Solving:

CO1: (Weightage: 3)

CO2: (Weightage: 3)

CO4: (Weightage: 3)

CO5: (Weightage: 3)

CO1, CO2, CO4, and CO5 develop specialized problem-solving skills through planning, risk

management, and agile implementation.

PO5. Research, Analytical Reasoning, and Ethical Conduct:

CO4: (Weightage: 3)

CO5: (Weightage: 3)

CO7: (Weightage: 3)

CO4, CO5, and CO7 align with research and analytical reasoning as students analyze case studies, propose project plans, and present findings with ethical considerations.

PO6. Communication, Collaboration, and Leadership:

CO4: (Weightage: 3)

CO7: (Weightage: 3)

CO4 and CO7 emphasize communication and collaboration through effective presentations, stakeholder management, and team coordination.

PO7. Digital Proficiency and Technological Skills:

CO2: (Weightage: 3)

CO6: (Weightage: 3)

CO7: (Weightage: 3)

CO2, CO6, and CO7 foster technological proficiency using project management software and agile tools.

PO8. Multicultural Competence, Inclusive Spirit, and Empathy:

Not directly aligned with the course outcomes as the course focuses on technical and management aspects.

PO9. Value Inculcation, Environmental Awareness, and Ethical Practices:

Not directly aligned since the primary focus is on project management and business strategy.

PO10. Autonomy, Responsibility, and Accountability:

CO3: (Weightage: 1)

CO6: (Weightage: 2)

CO3 and CO6 demonstrate accountability and independence in decision-making while handling projects.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-II Data Science (2023 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – II
Semester	: III
Course Type	: Major Elective Theory
Course Name	: Text Mining and Natural Language Processing
Course Code	: DSC-611-MJE(B)
No. of Credits	: 2
No. of Teaching Hours	: 30

Course Objectives:

1. Describe the fundamental concepts and techniques of natural language processing.
2. Distinguish among the various techniques, taking into account the assumptions, strengths, and weaknesses of each,
3. Use appropriate descriptions, visualizations, and statistics to communicate the problems and their solutions.
4. Analyze large volume text data generated from a range of real-world applications.

Course Outcomes:

By the end of the course, students will be able to:

CO1. grasp the significance of natural language processing in solving real-world problems.

CO2. map the appropriate processing technique to a problem and implement the technique.

CO3. demonstrate required design skills for large collection sets.

CO4. comprehend the state-of-the-art advanced nlp research articles and present them to an audience.

CO5. propose extension of existing nlp techniques for solving a range of problems.

CO6. gain practical experience using text mining tools and libraries.

CO7. effectively communicate the results of text mining and NLP analyses in written reports and oral presentations.

Topics and Learning Points

Unit 1

Introduction to text data, structure of text data, Working with Text Data, Character Encodings, Tokenization, Parsing, Stemming, APIs, Spelling Correction, representation of the unstructured text documents with appropriate format and structure to support later automated text mining algorithms. **(15 L)**

Unit 2

Probabilistic models for text mining: Naïve Bayes, basic supervised text categorization algorithms: k Nearest Neighbor (kNN) and Logistic Regression, Support Vector Machines and Decision Trees. Text clustering: introduction, typical types of clustering algorithms: connectivity-based clustering (hierarchical clustering) and centroid-based clustering (e.g., k-means clustering). **(10 L)**

Unit 3

Sentiment Analysis: Introduction of sentiment analysis, task of extracting subjective information in source materials, problems in sentiment analysis: sentiment polarity prediction, review mining, and aspect identification. **(05 L)**

References:

References Books:

1. Mining Text Data. Charu C. Aggarwal and ChengXiangZhai, Springer, 2012.
2. Speech & Language Processing. Dan Jurafsky and James H Martin, Pearson Education India, 2000.
3. Introduction to Information Retrieval. Christopher D. Manning, PrabhakarRaghavan, and HinrichSchuetze, Cambridge University Press, 2007.
4. Foundations of Statistical Natural Language Processing by Christopher Manning and HinrichSchütze.
5. Natural Language Processing with Pythonby Steven Bird, Ewan Klein and Edward Loper.
6. Survey of Text MiningClustering, Classification, and Retrieval by Michael W. Berry

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	Programme Outcomes (POs)									
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	3	3	-	3	-	-	-	-	-	-
CO2	3	3	-	3	-	-	3	-	-	-
CO3	2	2	-	-	-	-	-	-	-	1
CO4	2	-	-	3	3	3	-	-	-	-
CO5	2	-	-	3	3	-	-	-	-	-
CO6	2	2	-	-	3	-	3	-	-	2
CO7	2	2	-	-	-	3	3	-	-	-

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

PO1. Comprehensive Knowledge and Understanding:

CO1: (Weightage: 3)

CO2: (Weightage: 3)

CO3: (Weightage: 2)

CO4: (Weightage: 3)

CO5: (Weightage: 2)

CO6: (Weightage: 2)

CO7: (Weightage: 2)

Justification: These outcomes align strongly with PO1 as they require a comprehensive understanding of natural language processing techniques, their applications, and the ability to apply them effectively to solve real-world problems.

PO2. Practical, Professional, and Procedural Knowledge:

CO1: (Weightage: 3)

CO2: (Weightage: 3)

CO3: (Weightage: 2)

CO6: (Weightage: 3)

CO7: (Weightage: 2)

Justification: These outcomes are directly related to practical, professional, and procedural knowledge as they involve applying NLP techniques to real-world problems, implementing solutions, gaining practical experience, and effectively communicating results.

PO4. Specialized Skills, Critical Thinking, and Problem-Solving:

CO1: (Weightage: 3)

CO2: (Weightage: 3)

CO4: (Weightage: 3)

CO5: (Weightage: 3)

Justification: These outcomes are strongly related to specialized skills, critical thinking, and problem-solving as they require understanding complex NLP concepts, analyzing research articles, proposing innovative solutions, and applying them effectively to solve problems.

PO5. Research, Analytical Reasoning, and Ethical Conduct:

CO4: (Weightage: 3)

CO5: (Weightage: 3)

CO7: (Weightage: 3)

Justification: These outcomes are strongly related to research and analytical reasoning as they involve understanding research articles, proposing innovative solutions, and communicating findings effectively. Ethical conduct is implicitly addressed through the responsible handling and communication of results.

PO6. Communication, Collaboration, and Leadership:

CO4: (Weightage: 3)

CO7: (Weightage: 3)

Justification: These outcomes are strongly related to communication skills as they involve presenting complex NLP concepts and findings to various audiences through written reports and oral presentations. Collaboration and leadership aspects are not explicitly addressed in these outcomes.

PO7. Digital Proficiency and Technological Skills:

CO2: (Weightage: 3)

CO6: (Weightage: 3)

CO7: (Weightage: 3)

Justification: These outcomes are strongly related to digital proficiency and technological skills as they involve implementing NLP techniques using tools and libraries, as well as effectively communicating results using digital mediums.

PO8. Multicultural Competence, Inclusive Spirit, and Empathy:

No direct alignment with the provided course outcomes.

Justification: The provided outcomes focus primarily on technical aspects related to NLP and do not directly address multicultural competence, inclusive spirit, or empathy.

PO9. Value Inculcation, Environmental Awareness, and Ethical Practices:

No direct alignment with the provided course outcomes.

Justification: The provided outcomes focus primarily on technical aspects related to NLP and do not directly address value inculcation, environmental awareness, or ethical practices.

PO10. Autonomy, Responsibility, and Accountability:

CO3: (Weightage: 1)

CO6: (Weightage: 2)

Justification: These outcomes are partially related to autonomy, responsibility, and accountability as they involve demonstrating design skills and gaining practical experience independently, albeit under the guidance of the course.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-II Data Science (2023 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – II
Semester	: III
Course Type	: Major Elective Practical
Course Name	: Practical Based on Business and Project Management
Course Code	: DSC-612-MJE(A)
No. of Credits	: 2
No. of Teaching Hours	: 60

Course Objectives:

1. To provide a comprehensive understanding of business management concepts and data-driven decision-making.
2. To develop practical skills in planning, executing, and managing data science projects using project management methodologies.
3. To enhance the ability to apply Agile methodologies for iterative project development and continuous improvement.
4. To build effective communication and reporting skills for presenting data-driven insights to stakeholders.
5. To provide hands-on experience in budgeting, risk management, and stakeholder management in data science projects.
6. To encourage critical thinking and problem-solving through case study analysis and capstone project execution.
7. To foster collaborative and leadership abilities for managing cross-functional data science teams.

Course Outcomes:

CO1: Demonstrate understanding of business management principles and data-driven decision-making.

CO2: Apply project management methodologies to plan, execute, and monitor data science projects.

CO3: Implement Agile practices for iterative development and effective progress monitoring.

CO4: Develop risk management and budgeting strategies for data science projects.

CO5: Communicate project outcomes effectively using reports, presentations, and storytelling techniques.

CO6: Apply stakeholder management techniques to ensure alignment with project goals.

CO7: Analyze real-world case studies and execute a capstone project to demonstrate practical application of project management concepts.

Topics and Learning Points

Sr. No.	Experiment Title
1	Analyze a company's business functions using data insights.
2	Develop a project scope and objective document for a data science project.
3	Create a stakeholder management plan for a sample project.
4	Perform risk management analysis for a data science project.
5	Build a project budget and cost estimation using available data.
6	Develop a Gantt chart for a data science project timeline.
7	Monitor and report progress using agile tracking tools.
8	Create a comprehensive business report using data storytelling techniques.
9	Develop a project presentation for stakeholders using visualization tools.
10	Conduct a case study analysis on a failed or successful data science project.
11	Propose a resource management strategy for a data science project.
12	Prepare a capstone project with complete documentation

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	1	2	1	2	1	1	3	1	1	2
CO2	2	3	1	2	1	1	3	1	1	2
CO3	3	3	1	3	3	1	3	1	1	2
CO4	3	3	1	3	3	1	3	1	1	2
CO5	3	3	1	3	3	1	3	1	1	2
CO6	2	3	1	3	2	1	3	1	1	2
CO7	3	3	1	3	3	1	3	1	1	2

Weightage Key:1 - Partially related2 - Moderately Related3 - Strongly related

PO1. Comprehensive Knowledge and Understanding:

- CO1: Partially Related (1)
- CO2: Moderately Related (2)
- CO3: Strongly Related (3)

- CO4: Strongly Related (3)
- CO5: Strongly Related (3)
- CO6: Moderately Related (2)
- CO7: Strongly Related (3)

Justification: The practicals on stakeholder management, risk analysis, and budget estimation provide comprehensive knowledge essential for effective project management.

PO2. Practical, Professional, and Procedural Knowledge:

- CO1: Moderately Related (2)
- CO2: Strongly Related (3)
- CO3: Strongly Related (3)
- CO4: Strongly Related (3)
- CO5: Strongly Related (3)
- CO6: Strongly Related (3)
- CO7: Strongly Related (3)

Justification: Practicals like developing a Gantt chart, resource management strategy, and monitoring progress enhance professional and procedural understanding.

PO3. Entrepreneurial Mindset, Innovation, and Business Understanding:

- All COs: Partially Related (1)

Justification: The experiments foster an entrepreneurial mindset by providing insights into business functions and decision-making through case studies and scenario analysis.

PO4. Specialized Skills, Critical Thinking, and Problem-Solving:

- CO1: Moderately Related (2)
- CO2: Moderately Related (2)
- CO3: Strongly Related (3)
- CO4: Strongly Related (3)
- CO5: Strongly Related (3)
- CO6: Strongly Related (3)
- CO7: Strongly Related (3)

Justification: The practicals on risk management, agile implementation, and business decision-making involve applying critical thinking and problem-solving skills.

PO5. Research, Analytical Reasoning, and Ethical Conduct:

- CO1: Partially Related (1)
- CO2: Partially Related (1)
- CO3: Strongly Related (3)

- CO4: Strongly Related (3)
- CO5: Strongly Related (3)
- CO6: Moderately Related (2)
- CO7: Strongly Related (3)

Justification: Practical applications of case study analysis and scenario evaluation enhance research and analytical reasoning. Ethical considerations are addressed through data interpretation and reporting.

PO6. Communication, Collaboration, and Leadership:

- All COs: Partially Related (1)

Justification: While collaboration and communication are indirectly developed through presentations and reports, no direct leadership focus is embedded.

PO7. Digital Proficiency and Technological Skills:

- All COs:. Strongly Related (3)

Justification: Practicals on using visualization tools, agile tracking tools, and data analysis software ensure digital proficiency.

PO8. Multicultural Competence, Inclusive Spirit, and Empathy:

- All COs: Partially Related (1)

Justification: While not directly addressed, case study analysis may provide cultural insights in the business context.

PO9. Value Inculcation, Environmental Awareness, and Ethical Practices:

- All COs: Partially Related (1)

Justification: While ethical practices in data analysis are reinforced, no direct focus on environmental awareness or values inculcation exists.

PO10. Autonomy, Responsibility, and Accountability:

- All COs: Moderately Related (2)

Justification: Independent work on capstone projects and assignments fosters autonomy and accountability in completing real-world tasks.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2023 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – II
Semester	: III
Course Type	: Major Elective Practical
Course Name	: Practical Based on Text Mining and NLP
Course Code	: DSC-612-MJE(B)
No. of Credits	: 2
No. of Teaching Hours	: 60

Course Objectives:

1. Gain a conceptual understanding of text mining and NLP techniques, including data preprocessing, feature extraction, and text analysis methods.
2. Explore advanced techniques such as topic modeling, text summarization, and text generation to extract meaningful insights from unstructured text data.
3. Learn techniques for text preprocessing, tokenization, stemming/lemmatization, and feature engineering to prepare text data for analysis.
4. Gain skills in evaluating the performance of text mining models using appropriate metrics and interpreting the results to derive actionable insights.
5. Develop skills in communicating findings and insights from text mining projects through written reports, presentations, and visualizations, targeting diverse audiences.
6. Collaborate effectively with experts from other fields to apply text mining techniques to interdisciplinary problems, fostering innovation and cross-disciplinary insights.
7. Develop a mindset for continuous learning and professional development in the rapidly evolving field of text mining and NLP, staying updated with the latest advancements and best practices.

Course Outcomes:

By the end of the course, students will be able to:

CO1. Students will demonstrate a comprehensive understanding of text mining and NLP techniques, including data preprocessing, feature extraction, and text analysis methods.

CO2. Students will gain practical experience in implementing machine learning algorithms for text classification, sentiment analysis, and named entity recognition tasks.

CO3. Students will engage in research-oriented tasks such as text summarization, topic

modeling, and text similarity analysis, fostering analytical reasoning and the ability to draw actionable insights from text data.

CO4. Students will demonstrate proficiency in using text processing tools and libraries (e.g., NLTK, spaCy, scikit-learn) and statistical programming languages (e.g., Python, R) for implementing text mining and NLP techniques.

CO5. Students will adhere to ethical guidelines and practices in handling text data, respecting privacy, confidentiality, and intellectual property rights, and ensuring transparency and fairness in their analysis and interpretation

CO6. Students will take ownership of their text mining projects, demonstrating autonomy, responsibility, and accountability in collecting, preprocessing, analyzing, and interpreting text data, and adhering to project timelines and deliverables.

CO7. Students will collaborate with experts from other fields (e.g., social sciences, humanities, healthcare) to apply text mining and NLP techniques to interdisciplinary problems, potentially leading to innovative solutions and insights

Topics and Learning Points

Sr.No.	Title of Experiment	Weightage
1.	Data Collection and Preprocessing	(2)
2.	Exploratory Data Analysis (EDA) for Text Data	(1)
3.	Text Classification	(1)
4.	Text Similarity Analysis	(1)
5.	Sentiment Analysis	(1)
6	Named Entity Recognition (NER)	(1)
7	Text Summarization	(1)
8	Text Similarity and Clustering	(1)
9	Text Classification with Deep Learning	(1)
10	Mini project	(2)

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	Programme Outcomes (POs)									
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	1	2	-	3	2	-	3	-	-	-
CO2	2	2	-	3	-	-	3	-	-	-

CO3	2	2	-	3	2	-	3	-	-	-
CO4	2	2	-	3	-	-	3	-	-	-
CO5	1	2	-	3	2	-	3	-	2	-
CO6	1	2	-	3	-	-	3	-	-	3
CO7	2	2	2	3	-	3	3	2	-	-

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

PO1. Comprehensive Knowledge and Understanding:

CO1: Partially Related (Weightage: 1)

Justification: While CO1 focuses on text mining and NLP techniques, it doesn't directly address comprehensive knowledge and understanding in broader domains. However, understanding these techniques contributes to a broader understanding of data analysis methodologies.

CO2: Moderately Related (Weightage: 2)

Justification: CO2 involves practical implementation of machine learning algorithms for text classification and sentiment analysis, which contributes to understanding practical applications of data science. However, it focuses specifically on text data, which is only one aspect of comprehensive knowledge.

CO3: Moderately Related (Weightage: 2)

Justification: CO3 involves research-oriented tasks in text mining such as text summarization and topic modeling, which contribute to analytical reasoning. While these tasks enhance understanding of text data analysis, they don't directly address comprehensive knowledge in other domains.

CO4: Moderately Related (Weightage: 2)

Justification: CO4 emphasizes proficiency in using text processing tools and statistical programming languages, which contributes to understanding data analysis techniques. However, it's specific to text mining and may not cover broader aspects of comprehensive knowledge.

CO5: Partially Related (Weightage: 1)

Justification: CO5 focuses on ethical guidelines in handling text data, which is important but doesn't directly contribute to comprehensive knowledge in various domains.

CO6: Partially Related (Weightage: 1)

Justification: CO6 emphasizes ownership, responsibility, and accountability in text mining projects, which are valuable skills but not directly related to comprehensive knowledge in other domains.

CO7: Moderately Related (Weightage: 2)

Justification: CO7 involves collaboration with experts from other fields to apply text mining techniques, which broadens the scope of application but may not directly contribute to comprehensive knowledge in those fields.

PO2. Practical, Professional, and Procedural Knowledge:

All COs are Moderately Related (Weightage: 2)

Justification: Each CO involves practical implementation of text mining and NLP techniques, which contributes to practical knowledge in data science. Additionally, CO6 emphasizes autonomy, responsibility, and accountability, which are essential for professionalism.

PO3. Entrepreneurial Mindset, Innovation, and Business Understanding:

CO7: Moderately Related (Weightage: 2)

Justification: CO7 involves collaboration with experts from other fields, potentially leading to innovative solutions and insights, which aligns with an entrepreneurial mindset and innovation.

PO4. Specialized Skills, Critical Thinking, and Problem-Solving:

All COs are Strongly Related (Weightage: 3)

Justification: Each CO involves specialized skills in text mining, critical thinking in analyzing text data, and problem-solving in implementing NLP techniques.

PO5. Research, Analytical Reasoning, and Ethical Conduct:

CO1, CO3, CO5: Moderately Related (Weightage: 2)

Justification: CO1 and CO3 involve research-oriented tasks in text mining and analytical reasoning in drawing insights from text data. CO5 emphasizes ethical conduct in handling text data.

PO6. Communication, Collaboration, and Leadership:

CO7: Strongly Related (Weightage: 3)

Justification: CO7 emphasizes collaboration with experts from other fields, which requires effective communication and potentially leadership skills.

PO7. Digital Proficiency and Technological Skills:

All COs are Strongly Related (Weightage: 3)

Justification: Each CO involves proficiency in using text processing tools, statistical programming languages, and implementing NLP techniques, which are essential for digital proficiency.

PO8. Multicultural Competence, Inclusive Spirit, and Empathy:

CO7: Moderately Related (Weightage: 2)

Justification: CO7 involves collaboration with experts from diverse fields, which promotes multicultural competence and inclusive spirit.

PO9. Value Inculcation, Environmental Awareness, and Ethical Practices:

CO5: Moderately Related (Weightage: 2)

Justification: CO5 emphasizes ethical practices in handling text data, which aligns with value inculcation and ethical practices.

PO10. Autonomy, Responsibility, and Accountability:

CO6: Strongly Related (Weightage: 3)

Justification: CO6 emphasizes autonomy, responsibility, and accountability in text mining projects, which directly aligns with PO10.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-II Data Science (2023 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – II
Semester	: III
Course Type	: Research Project (RP)
Course Name	: Research Project
Course Code	: DSC-621-RP
No. of Credits	: 4
No. of Teaching Hours	: 60

Course Objectives:

1. To develop proficiency in using statistical software packages like R, SPSS, Matlab or Python for data analysis and visualization.
2. To acquire skills in data collection, data cleaning, and data transformation.
3. To improve the ability to communicate statistical findings effectively through written reports and presentations.
4. To apply advanced statistical techniques to analyze the research data and draw meaningful conclusions.
5. To interpret the results of the analysis and discuss their implications in the context of the project questions/objectives.
6. To present the project findings in a clear and concise manner, both in written form and through oral presentations.
7. To develop the ability to critically evaluate existing statistical literature and research studies in the field.

Course Outcomes:

By the end of the course, students will be able to:

CO1. Students will be able to gain practical experience in data collection, data cleaning, and data imputation, which are essential skills in statistics, data analytics and data science.

CO2. Gaining expertise in statistical software packages like R, SAS, or Python and using these tools is valuable for future career opportunities in IT industry and many more filed.

CO3. MSc project serves as a valuable stepping stone, demonstrating research capabilities.

CO4. Statistical analysis may provide insights that can inform policy or decision-making in these areas in a specific social issue or problem, such as healthcare, education, or environmental sustainability.

CO5. MSc projects can identify actionable insights; consider providing recommendations or guidelines for addressing the social issue that were studied.

CO6. Collaborate with experts from other fields (e.g., biology, economics, psychology, garniture, manufacturing industry) to apply statistical methods to interdisciplinary problems, potentially leading to innovative solutions and insights.

CO7. Successful MSc projects can open doors to consulting opportunities where students can apply statistical methods to solve practical problems for businesses or organizations.

Topics and Learning Points

This part of the course consist summary of research articles, data analysis and report in dissertation form.

1. **Summary of Research Articles**

Students are expected to choose her/his own project topic and read some (not less than 5) articles (exact number of articles will be decided by the supervisor) on a selected topic or theme, summarize and write a comprehensive report and present the summary of the articles.

2. **Data Analysis**

Students are expected to analyze data pertaining to certain theme using a variety of statistical tools that they have studied so far.

Note:

1. Students have to prepare project report and have to submit one copy for the assessment.

2. Data analysis project can be done in a group (at the most 3 students).

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	Programme Outcomes (POs)									
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	2	3	-	3	-	-	-	-	-	-
CO2	-	3	-	-	-	-	3	-	-	-
CO3	-	-	-	-	-	-	-	-	-	2
CO4	3	-	-	-	3	-	-	-	2	-
CO5	-	-	-	-	-	-	-	-	-	-
CO6	3	-	-	3	-	2	-	-	-	-
CO7	-	-	2	-	-	-	-	3	-	-

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

PO1. Comprehensive Knowledge and Understanding:

CO1: Moderately related (Weightage: 2)

Justification: CO1 involves practical experience in data collection, cleaning, and imputation, which contribute to comprehensive knowledge and understanding of data handling processes in statistics, data analytics, and data science.

CO4: Strongly related (Weightage: 3)

Justification: CO4 emphasizes the application of statistical analysis to inform policy or decision-making in various domains, which requires a comprehensive understanding of statistical methods and their implications.

CO6: Strongly related (Weightage: 3)

Justification: CO6 involves collaboration with experts from other fields to apply statistical methods to interdisciplinary problems, which enhances understanding of how statistical techniques can be integrated into diverse domains.

PO2. Practical, Professional, and Procedural Knowledge:

CO1: Strongly related (Weightage: 3)

Justification: CO1 directly involves gaining practical experience in data collection, cleaning, and imputation, which are essential components of practical, professional, and procedural knowledge in statistics and data science.

CO2: Strongly related (Weightage: 3)

Justification: CO2 highlights the importance of gaining expertise in statistical software packages like R, SAS, or Python, which is directly linked to practical, professional, and procedural knowledge required in data analysis and IT industry.

PO3. Entrepreneurial Mindset, Innovation, and Business Understanding:

CO7: Moderately related (Weightage: 2)

Justification: CO7 mentions consulting opportunities where students can apply statistical methods to solve practical problems for businesses or organizations, aligning with the entrepreneurial mindset and business understanding.

PO4. Specialized Skills, Critical Thinking, and Problem-Solving:

CO1: Strongly related (Weightage: 3)

Justification: CO1 involves practical experience in data collection, cleaning, and imputation, which requires specialized skills, critical thinking, and problem-solving abilities to address data challenges effectively.

CO6: Strongly related (Weightage: 3)

Justification: CO6 emphasizes collaboration with experts from other fields to apply statistical methods to interdisciplinary problems, demonstrating critical thinking and problem-solving skills in integrating statistical techniques into diverse contexts.

PO5. Research, Analytical Reasoning, and Ethical Conduct:

CO4: Strongly related (Weightage: 3)

Justification: CO4 involves using statistical analysis to inform policy or decision-making, which requires strong analytical reasoning and adherence to ethical conduct in handling and interpreting data.

PO6. Communication, Collaboration, and Leadership:

CO6: Moderately related (Weightage: 2)

Justification: CO6 involves collaboration with experts from other fields, which enhances communication and collaboration skills, but may not directly address leadership aspects.

PO7. Digital Proficiency and Technological Skills:

CO2: Strongly related (Weightage: 3)

Justification: CO2 emphasizes gaining expertise in statistical software packages like R, SAS, or Python, directly contributing to digital proficiency and technological skills required in data analysis and IT industry.

PO8. Multicultural Competence, Inclusive Spirit, and Empathy:

CO6: Strongly related (Weightage: 3)

Justification: CO6 involves collaboration with experts from diverse fields, fostering multicultural competence, inclusive spirit, and empathy by respecting diverse perspectives and understanding others' viewpoints.

PO9. Value Inculcation, Environmental Awareness, and Ethical Practices:

CO4: Moderately related (Weightage: 2)

Justification: CO4 involves using statistical analysis to inform policy or decision-making, which may include considerations for ethical practices, but may not directly address value inculcation and environmental awareness aspects.

PO10. Autonomy, Responsibility, and Accountability:

CO3: Moderately related (Weightage: 2)

Justification: CO3 mentions MSc projects serving as a valuable stepping stone, demonstrating research capabilities, which indirectly relates to autonomy, responsibility, and accountability in managing research projects.

Top of Form