Anekant Education Society's

Tuljaram Chaturchand College of Arts, Science and Commerce, Baramati

(AUTONOMOUS)

Department of Statistics

Two Years Post Graduate Program

M.Sc. Data Science

Course Structure and Syllabus

Semester- I

Paper Code	Course Title	No. of Credits
PSDS111	Linear Algebra in Matlab	04
PSDS112	Probability Distributions	04
PSDS113	Optimization Techniques	04
PSDS114	Statistical Inference	04
PSDS115	Database Management System	04
PSDS116	Introduction to MATLAB and R	04

Semester- II

Paper Code	Course Title	No. of Credits
PSDS121	Design and Analysis of Experiments	04
PSDS122	Regression Analysis and Predictive Models	04
PSDS123	Statistical Quality Control	04
PSDS124	Computational Statistics	04
PSDS125	Bayesian Inference	04
PSDS126	Python and SQL Programming	04

Semester- III

Paper Code	Course Title	No. of Credits
PSDS231	Stochastic Models and Applications	04
PSDS232	Exploratory Multivariate Data Analysis	04
PSDS233	Time series analysis and Forecasting	04
PSDS234	Artificial Intelligence	04
PSDS235	Text Mining and Natural Language Processing	04
PSDS236	Data Visualization using Tableau	04

Semester- IV

Paper Code	Course Title	No. of Credits
PSDS241	Machine Learning	04
PSDS242	Discrete Data Analysis	04
PSDS243	Supply Chain & Logistics Analytics	04
PSDS244	Deep Learning	04
PSDS245	Thesis	08

Program Outcomes (POs) for M.Sc. Programme

PO1	**Disciplinary Knowledge:** Demonstrate comprehensive knowledge of the discipline that forms a part of a postgraduate programme. Execute strong theoretical and practical understanding generated from the specific programme in the area of work.
PO2	**Critical Thinking and Problem solving:** Exhibit the skill of critical thinking and understand scientific texts and place scientific statements and themes in contexts and also evaluate them in terms of generic conventions. Identify the problem by observing the situation closely, take actions and apply lateral thinking and analytical skills to design the solutions.
PO3	**Social competence:** Exhibit thoughts and ideas effectively in writing and orally; communicate with others using appropriate media, build effective interactive and presenting skills to meet global competencies. Elicit views of others, present complex information in a clear and concise way and help reach conclusions in group settings.
PO4	**Research-related skills and Scientific temper :** Infer scientific literature, build a sense of enquiry and able to formulate, test, analyse, interpret and establish hypothesis and research questions; and to identify and consult relevant sources to find answers. Plan and write a research paper/project while emphasizing on academics and research ethics, scientific conduct and creating awareness about intellectual property rights and issues of plagiarism.
PO5	**Trans-disciplinary knowledge:** Create new conceptual, theoretical and methodological understanding that integrates and transcends beyond discipline-specific approaches to address a common problem.
PO6	**Personal and professional competence:** Perform independently and also collaboratively as a part of a team to meet defined objectives and carry out work across interdisciplinary fields. Execute interpersonal relationships, self-motivation and adaptability skills and commit to professional ethics.
PO7	**Effective Citizenship and Ethics:** Demonstrate empathetic social concern and equity centred national development, and ability to act with an informed awareness of moral and ethical issues and commit to professional ethics and responsibility.
PO8	**Environment and Sustainability:** Understand the impact of the scientific solutions in societal and environmental contexts and demonstrate the knowledge of and need for sustainable development.
PO9	**Self-directed and Life-long learning:** Acquire the ability to engage in independent and life-long learning in the broadest context of socio-technological changes.

SYLLABUS (CBCS) FOR M.Sc. Data Science
(2022 Pattern)
(With effect from Academic Year 2022-2023)

Class : M. Sc. (Semester- III)
Paper Code : PSDS231
Paper : I
Title of Paper : Stochastic Models and Applications
Credit : 4 credits
No. of lectures : 60

Course Objectives:

1. To understand discrete and continuous Markov chains models to compute the probability of events.
2. Formulate and solve problems by computing the long-term probabilities of a Markov chain model.
3. Write Python/R code to simulate Markov chains, and compute probabilities of events that may be difficult to derive by hand.
4. Apply Poisson processes to model the occurrence of events in various applications.

Course Outcomes:

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

CO1. formulate tpm, n-step transition probabilities
CO2. classify of states and perform stochastic simulations.
CO3. familiar with stochastic processes, including Poisson process, Wiener process and Renewal process, etc.
CO4. understand stationary processes and its properties.
CO5. develop problem-solving skills of stochastic processes theory to practical problems.
CO6. explore the ethical implications of using stochastic processes in various fields.
CO7. learn statistical packages for modeling and analyzing stochastic processes.

TOPICS/CONTENTS:

Unit 1

Notion of stochastic processes, Markov chain, one step transition probabilities, Chapman-Kolmogorov equations, evaluation of higher step transition probabilities, classification of states, periodicity of a Markov chain, concept of closed class, minimal closed class, stationary distribution. Some examples such as gamblers ruin problem and one-dimensional random walk. Concept of absorption probabilities, Use of these to compute probability of winning the game by a gambler having initial capital ‘a’

Unit 2

Branching process, classification of states, identification of criticality parameter, extinction probability, relationship between criticality parameter and extinction probability of the process, Expression for mean and variance of the process. Extinction probability, Some epidemiological applications, Introduction to Markov chain in continuous time, concept of intensity rate, relationship between intensity matrix and transition probability matrix. Kolmogorov's forward and backward equations

Unit 3

Introduction to birth process, birth and death process, linear birth and death process, Growth model with immigration and related results, Expression for mean and variance of a birth process and, birth and death process, Applications of these processes.

Unit 4

Poisson process, two definitions and their equivalence, Distribution of inter arrival times, conditional joint distribution of inter arrival times. Compound Poisson process, Some applications. Introduction to renewal process, relationship with Poisson process, key and elementary renewal theorems associated with renewal processes.

Books Recommended

1. Bhat B.R. (2000). Stochastic Models: Analysis and Applications, New Age International.
2. Medhi, J. (2010) Stochastic Processes, New Age Science Ltd.
3. Pinsky M. A. and Karlin, S. (2010). An Introduction to Stochastic Modeling, 4thEdn. Academic Press.
4. Ross, S. (2014). Introduction to Probability Models, 11th Edn. Academic Press.
5. Feller, W. (1972). An Introduction to Probability Theory and its Applications, Vol. 1, Wiley Eastern.
6. Hoel, P.G. Port, S.C. & Stone, C.J. (1972). Introduction to Stochastic Processes, Houghton Mifflin
7. Karlin, S & Taylor, H.M. (1975). A First Course in Stochastic Processes (Second. Edition), Academic Press.
8. Serfozo, R. (2009). Basics of Applied Stochastic Processes, Springer.

Programme Outcomes and Course Outcomes Mapping:

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

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

PO1. Disciplinary Knowledge

CO1. Develop a deep understanding of what stochastic processes are, including their definitions, characteristics, and mathematical representations. (Weightage: 3 - Strongly Related)

Justification: Developing a deep understanding of stochastic processes directly aligns with building disciplinary knowledge in probability and statistics.

CO2. Understand stationary processes and their properties. (Weightage: 3 - Strongly Related)

Justification: Understanding stationary processes contributes to disciplinary knowledge in stochastic processes, covering important properties and concepts.

CO7. Learn about continuous-time stochastic processes, including the Poisson process, Brownian motion, Wiener process, and Renewal process. (Weightage: 3 - Strongly Related)

Justification: Learning about continuous-time stochastic processes enhances disciplinary knowledge, providing a broader understanding of stochastic modeling.

PO2. Critical Thinking and Problem Solving

CO3. Develop problem-solving skills of stochastic processes theory to practical problems. (Weightage: 3 - Strongly Related)

Justification: Problem-solving in stochastic processes involves critical thinking, connecting theoretical knowledge to practical applications.

CO5. Perform stochastic simulations. (Weightage: 2 - Moderately Related)

Justification: Stochastic simulations involve practical problem-solving, making this outcome moderately related to critical thinking.

PO3. Social Competence

CO4. Explore the ethical implications of using stochastic processes in various fields. (Weightage: 2 - Moderately Related)

Justification: Exploring ethical implications has a social dimension, making this outcome moderately related to social competence.

PO4. Research-related Skills and Scientific Temper

All COs (Weightage: 3 - Strongly Related)

Justification: The outcomes align closely with research-related skills and the development of a scientific temper in the context of stochastic processes.

PO5. Trans-disciplinary Knowledge

CO4. Explore the ethical implications of using stochastic processes in various fields. (Weightage: 2 - Moderately Related)

Justification: Ethical considerations in using stochastic processes may have trans-disciplinary implications, making this outcome moderately related.

PO6. Personal and Professional Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes contribute more to technical competence in stochastic processes, with limited direct connections to personal and professional aspects.

PO7. Effective Citizenship and Ethics

CO4. Explore the ethical implications of using stochastic processes in various fields. (Weightage: 2 - Moderately Related)

Justification: Exploring ethical implications is moderately related to effective citizenship and ethical considerations.

PO8. Environment and Sustainability

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more focused on stochastic processes methodologies than on environmental or sustainability aspects.

PO9. Self-directed and Life-long Learning

All COs (Weightage: 3 - Strongly Related)

Justification: Stochastic processes are a dynamic field, and mastering its principles requires ongoing self-directed learning and adaptation.

SYLLABUS (CBCS) FOR M.Sc. Data Science
(2022 Pattern)
(With effect from Academic Year 2022-2023)

Class : M. Sc. Data Science (Semester- III)
Paper Code : PSDS232
Paper : II
Title of Paper : Exploratory Multivariate Data Analysis
Credit : 4 credits
No. of lectures : 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 summaries and synthesize datasets using simple graphs, Make graphical displays of very high dimensional data
4. To use visualization methods adapted to multidimensional exploratory analysis.
5. To recognize the method adapted to the exploration of a dataset according to the nature and structure of the variables.

Course Outcomes:

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. create meaningful graphical representations of multivariate data.
CO3. apply the concepts of linear and quadratic forms in multivariate normal variables.
CO4. solve problems involving multivariate normal distribution evaluate.
CO5. carry out statistical inference procedures using the data from a multivariate normal distribution.
CO6. carry out classification of given multivariate data.
CO7. perform hypothesis tests related to the mean vector of a multivariate normal

TOPICS/CONTENTS:

Unit 1

Multivariate data and their diagrammatic representation, Exploratory multivariate data analysis, sample mean vector, sample dispersion matrix, sample correlation matrix, graphical representation, mean, variance, co-variance, correlation of Linear transformations., six step approach to multivariate model building. Introduction to multivariate regressions models

Principal component Analysis (by using covariance and correlation method, standardized method), Factor analysis (models, rotation types), Canonical correlation with real life examples. **[15L]**

Unit 2

Cluster analysis (Hierarchical and Non hierarchical, Agglomerative, Single, complete, average, Wald's linkage, K mean clustering method, qualitative method clustering).Multivariate normal distribution, Singular and non singular normal distribution, mean, variance of multivariate normal distribution, Random sampling from multivariate normal distributions, independence of variables, M.G.F. Characteristic function, moments, Distribution of linear and quadratic form of normal variables, marginal and conditional distribution, multiple and partial correlation coefficient (3 random variable case) with examples on each of the topic. **[15L]**

Unit 3

Multivariate Linear Model and Analysis of Variance and Covariance: Maximum likelihood estimation of parameters, tests of linear hypothesis, distribution of partial and multiple correlation coefficients and regression coefficients. Multivariate linear regression, multivariate analysis of variance of one and two way classification data (only LR test). Multivariate analysis of covariance. Hotelling $T^2$and Mahalanobis D^2applications in testing and confidence set construction.

[10L]

Unit 4

Logistic Regression model and analysis: regression with a binary dependent variable, representation of the binary dependent variable, estimating the logistic regression model, assessing the goodness of fit of the estimation model, testing for significance of the coefficients, interpreting the coefficients, criteria for evaluation of logistic regression model, KS, Gini, AUC, Precision, Recall F1 score etc. Discriminant model and analysis: a two-group discriminant analysis, a three-group discriminant analysis, the decision process of discriminant analysis (objective, research design, assumptions, estimation of the model, assessing overall fit of a model, interpretation of the results, validation of the results). **[20L]**

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 and Solutions,
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
CO1	3	3	1	3	1	1	1	1	3
CO2	3	3	1	3	1	1	1	1	3
CO3	3	3	1	3	1	1	1	1	3
CO4	3	3	1	3	1	1	1	1	3
CO5	3	3	1	3	1	1	1	1	3
CO6	3	3	1	3	1	1	1	1	3
CO7	3	3	1	3	1	1	1	1	3

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

PO1. Disciplinary Knowledge

All COs (Weightage: 3 - Strongly Related)

Justification: All course outcomes involve the application of statistical techniques to multivariate data, directly contributing to disciplinary knowledge in multivariate analysis.

PO2. Critical Thinking and Problem Solving

CO1. Carry out an extensive exploratory multivariate analysis for a given multivariate data, carry out cluster analysis of given multivariate data. (Weightage: 3 - Strongly Related)

CO2. Create meaningful graphical representations of multivariate data. (Weightage: 3 - Strongly Related)

CO3. Apply the concepts of linear and quadratic forms in multivariate normal variables. (Weightage: 3 - Strongly Related)

CO4. Solve problems involving multivariate normal distribution evaluate. (Weightage: 3 - Strongly Related)

CO5. Carry out statistical inference procedures using the data from a multivariate normal distribution. (Weightage: 3 - Strongly Related)

CO6. Carry out classification of given multivariate data. (Weightage: 3 - Strongly Related)

CO7. Perform hypothesis tests related to the mean vector of a multivariate normal. (Weightage: 3 - Strongly Related)

Justification: All outcomes involve critical thinking and problem-solving skills in the context of multivariate analysis.

PO3. Social Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more technically oriented, with less direct relevance to social competence.

PO4. Research-related Skills and Scientific Temper

All COs (Weightage: 3 - Strongly Related)

Justification: The outcomes align closely with research-related skills and the development of a scientific temper in the context of multivariate analysis.

PO5. Trans-disciplinary Knowledge

All COs (Weightage: 1 - Partially Related)

Justification: The focus is primarily on statistical techniques and multivariate analysis, with limited direct connection to trans-disciplinary knowledge.

PO6. Personal and Professional Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes contribute more to technical competence in multivariate analysis, with limited direct connections to personal and professional aspects.

PO7. Effective Citizenship and Ethics

All COs (Weightage: 1 - Partially Related)

Justification: The content is more technical, with less direct emphasis on effective citizenship and ethics.

PO8. Environment and Sustainability

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more focused on statistical techniques and multivariate analysis, with less direct relevance to environmental or sustainability aspects.

PO9. Self-directed and Life-long Learning

All COs (Weightage: 3 - Strongly Related)

Justification: The continuous application of statistical techniques to multivariate data aligns well with self-directed and life-long learning in the context of multivariate analysis.

Top of Form

SYLLABUS (CBCS) FOR M.Sc. Data Science

(2022 Pattern)

(With effect from Academic Year 2022-2023)

Class : M. Sc. Data Science (Semester- III)
Paper Code : PSDS233
Paper : III
Title of Paper : Time Series Analysis and Forecasting
Credit : 4 credits
No. of lectures : 60

Course Objectives:

1. To equip various forecasting techniques and familiarize on modern statistical methods for analyzing time series data.
2. To amalgamate the intellectual facts of the time series data to implement in the field projects scientifically.
3. To link time dependent analytical tools and building the models by extracting real time data.

Course Outcomes

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

CO1. fit the model on time series data like ARMA, ARIMA, SARIMA, ARCH and GARCH properties.

CO2. apply and understand the techniques for estimating parameters of time series models also the role of maximum likelihood estimation in time series modeling.

CO3. perform diagnostic checks on time series models to assess model adequacy.

CO4. identify and address issues such as autocorrelation and heteroscedasticity.

CO5. analyses time series data and use multivariate time series models such as vector auto regression (VAR).

CO6. Gain proficiency in using ITSM, R and Python to fit an appropriate time series model and infer the results.

CO7. effectively interpret the results of time series analyses, both in written reports and oral presentations.

TOPICS/CONTENTS:

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 L)**

Unit 2

Stationary models:

Time series data, Trend, seasonality, cycles and residuals, Stationary, White noise processes, Autoregressive (AR), Moving Average (MA), Autoregressive and Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) processes, Choice of AR and MA periods.

Non- Stationarity models:

Tests for Non-stationarity: Random walk, random walk with drift, Trend stationary, General Unit Root Tests: Dickey Fuller Test, Augmented Dickey Fuller Test. ARIMA Models: Basic formulation of the ARIMA Model and their statistical properties, Autocorrelation function (ACF), Partial autocorrelation function (PACF) and their standard Errors. **(15 L)**

Unit 3

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.

Transfer function and Intervention analysis

Transfer function models, Transfer function, noise models, Cross correlation function, Model Specification, Forecasting with Transfer function, noise models, Intervention analysis.

Unit 4

Spectral analysis

Spectral density function (s. d. f.) and its properties, s. d. f. of AR, MA and ARMA processes, Fourier transformation and period gram. **(10 L)**

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:

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

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

PO1. Disciplinary Knowledge

CO1. Fit the model on time series data like ARMA, ARIMA, SARIMA, ARCH, and GARCH properties. (Weightage: 3 - Strongly Related)

Justification: This directly aligns with building disciplinary knowledge in time series modeling, covering various models and their properties.

CO2. Apply and understand the techniques for estimating parameters of time series models, also the role of maximum likelihood estimation in time series modeling. (Weightage: 3 - Strongly Related)

Justification: Understanding and applying estimation techniques in time series models contribute directly to disciplinary knowledge in this field.

CO3. Perform diagnostic checks on time series models to assess model adequacy.

(Weightage: 3 - Strongly Related)

Justification: Diagnostic checks are an essential part of assessing the validity of time series models, enhancing disciplinary knowledge.

CO4. Identify and address issues such as autocorrelation and heteroscedasticity. (Weightage: 3 - Strongly Related)

Justification: Dealing with issues like autocorrelation and heteroscedasticity is crucial in time series modeling, contributing directly to disciplinary knowledge.

CO5. Analyze time series data and use multivariate time series models such as vector auto-regression (VAR). (Weightage: 3 - Strongly Related)

Justification: Analyzing time series data and using multivariate models extends disciplinary knowledge to a broader and more advanced level.

CO6. Gain proficiency in using ITSM, R, and Python to fit an appropriate time series model and infer the results. (Weightage: 3 - Strongly Related)

Justification: Proficiency in using specific tools and programming languages is integral to disciplinary knowledge in the modern context of time series analysis.

CO7. Effectively interpret the results of time series analyses, both in written reports and oral presentations. (Weightage: 3 - Strongly Related)

Justification: The ability to interpret and communicate results is a critical aspect of disciplinary knowledge, especially in the context of time series analysis.

PO2. Critical Thinking and Problem Solving

All COs (Weightage: 3 - Strongly Related)

Justification: Time series analysis requires critical thinking and problem-solving skills at every stage, from model fitting to interpretation of results.

PO3. Social Competence

CO7. Effectively interpret the results of time series analyses, both in written reports and oral presentations. (Weightage: 1 - Partially Related)

Justification: While interpretation skills are essential, the direct social relevance is limited in this technical context.

PO4. Research-related Skills and Scientific Temper

All COs (Weightage: 3 - Strongly Related)

Justification: Time series analysis is a research-oriented field, and the outcomes align closely with research-related skills and scientific temper.

PO5. Trans-disciplinary Knowledge

CO5. Analyze time series data and use multivariate time series models such as vector auto-regression (VAR). (Weightage: 2 - Moderately Related)

Justification: The use of multivariate models extends the knowledge to a more trans-

disciplinary context, although not fully.

PO6. Personal and Professional Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes contribute more to technical competence, with limited direct connection to personal and professional aspects.

PO7. Effective Citizenship and Ethics

All COs (Weightage: 1 - Partially Related)

Justification: The content is more technical and less directly related to citizenship and ethics.

PO8. Environment and Sustainability

All COs (Weightage: 1 - Partially Related)

Justification: The focus here is more on statistical methodologies and less on environmental or sustainability aspects.

PO9. Self-directed and Life-long Learning

All COs (Weightage: 3 - Strongly Related)

Justification: Time series analysis is a dynamic field, and mastering its principles requires ongoing self-directed learning and adaptation.

SYLLABUS (CBCS) FOR M.Sc. Data Science

(2022 Pattern)

(With effect from Academic Year 2022-2023)

Class : M. Sc. Data Science (Semester- III)
Paper Code : PSDS234
Paper : IV
Title of Paper : Artificial Intelligence
Credit : 4 credits
No. of lectures : 60

Course Objectives:

1. Students will be familiar with basic principles of AI.
2. Students will be capable of using heuristic searches.

Course Outcome:

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

CO1. develop a solid understanding of the fundamental concepts and principles of artificial intelligence.
CO2. explore techniques for processing.
CO3. aware of knowledge-based systems.
CO4. use fuzzy logic and neural networks.
CO5. learn a variety of AI algorithms and techniques applicable to different domains, and understand the strengths and limitations of various approaches.
CO6. explore the ethical considerations and societal impacts of AI technologies.
CO7. apply AI techniques to real-world problems in different industries, and understand how AI is used in research and development.

Unit 1

Fundamentals of Artificial Intelligence: Introduction, A.I. Representation, Non-AI & AI Techniques, Representation of Knowledge, Knowledge Base Systems, State Space Search, Production Systems, Problem Characteristics, types of production systems, Intelligent Agents and Environments, concept of rationality, the nature of environments, structure of agents, problem solving agents, problem formulation. **(15)**

Unit 2

Uninformed Search Strategies: Formulation of real-world problems, Breadth First Search, Depth First Search, Depth Limited Search, Iterative Deepening Depth First Search, Bidirectional Search, Comparison of Uninformed search Strategies, Searching with partial information, Sensor-less problems, Contingency problems. **(15)**

Unit 3

Knowledge Representation Knowledge based agents, Wumpus world. Propositional Logic: Representation, Inference, Reasoning Patterns, Resolution, Forward and Backward Chaining. First order Logic: Representation, Inference, Reasoning Patterns, Resolution, Forward and Backward Chaining. Basics of PROLOG: Representation, Structure, Backtracking. Expert System: Case study of Expert System in PROLOG. **(15 L)**

Unit 4

Introduction to Planning and ANN: Blocks world, STRIPS, Implementation using goal stack, Introduction to Neural networks: - basic, comparison of human brain and machine, biological neuron, general neuron model, activation functions, Perceptron learning rule, applications and advantages of neural networks. Brief introduction to single layer and multiplayer networks.

(15 L)

Reference Book:

1. Ivan Bratko : "Prolog Programming For Artificial Intelligence" , 2nd Edition Addison Wesley, 1440.
2. Eugene, Charniak, Drew Mcdermott: "Introduction to Artificial Intelligence.", Addison Wesley
3. Patterson: ―Introduction to AI and Expert Systems‖, PHI
4. Nilsson: ―Principles of Artificial Intelligence‖, Morgan Kaufmann.
5. Carl Townsend, ―Introduction to turbo Prolog‖, Paperback, 1483 6. Jacek M. Zurada, Introduction to artificial neural systems, Jaico Publication.
6. Elaine Rich and Kevin Knight: "Artificial Intelligence." Tata McGraw Hill
7. Stuart Russell & Peter Norvig: "Artificial Intelligence: A Modern Approach", Pearson Education, 2nd Edition.

Programme Outcomes and Course Outcomes Mapping:

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

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

PO1. Disciplinary Knowledge

CO1. Develop a solid understanding of the fundamental concepts and principles of artificial intelligence. (Weightage: 3 - Strongly Related)

Justification: Developing a solid understanding of AI principles directly contributes to disciplinary knowledge in artificial intelligence.

PO2. Critical Thinking and Problem Solving

All COs (Weightage: 3 - Strongly Related)

Justification: All outcomes involve critical thinking and problem-solving skills in the context of artificial intelligence.

PO3. Social Competence

CO6. Explore the ethical considerations and societal impacts of AI technologies. (Weightage: 3 - Strongly Related)

Justification: Exploring the ethical considerations and societal impacts of AI technologies contributes to social competence.

PO4. Research-related Skills and Scientific Temper

CO7. Apply AI techniques to real-world problems in different industries, and understand how AI is used in research and development. (Weightage: 3 - Strongly Related)

Justification: Applying AI techniques to real-world problems aligns closely with research-related skills and the development of a scientific temper in the context of AI.

PO5. Trans-disciplinary Knowledge

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are primarily focused on AI techniques and principles, with limited direct connection to trans-disciplinary knowledge.

PO6. Personal and Professional Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes contribute more to technical competence in artificial intelligence, with limited direct connections to personal and professional aspects.

PO7. Effective Citizenship and Ethics

CO6. Explore the ethical considerations and societal impacts of AI technologies. (Weightage: 3 - Strongly Related)

Justification: Exploring ethical considerations and societal impacts aligns with effective citizenship and ethical awareness.

PO8. Environment and Sustainability

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more focused on AI techniques and principles, with less direct relevance to environmental or sustainability aspects.

PO9. Self-directed and Life-long Learning

All COs (Weightage: 3 - Strongly Related)

Justification: Continuous exploration of AI techniques and principles aligns well with self-directed and life-long learning in the context of artificial intelligence.

SYLLABUS (CBCS) FOR M.Sc. Data Science

(2022 Pattern)

(With effect from Academic Year 2022-2023)

Class : M. Sc. (Semester- III)
Paper Code : PSDS235
Paper : V
Title of Paper : Text Mining and Natural Process Language
Credit : 4 credits
No. of lectures : 60

Course Objectives:

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

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.

Unit 1 **(15 L)**

Introduction to text data, structure of text data, Working with Text Data, Character Encodings, Tokenization, Parsing, Stemming, APIs, Web Scraping, Regular Expressions, Spelling Correction, representation of the unstructured text documents with appropriate format and structure to support later automated text mining algorithms.

Unit 2 **(15 L)**

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.

Unit 3 **(15 L)**

Text clustering: introduction, typical types of clustering algorithms: connectivity-based clustering (hierarchical clustering) and centroid-based clustering (e.g., k-means clustering).

Unit 4 **(15 L)**

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. Social media and network analysis: characteristic of social network (inter-connectivity, and introduce Google's winning algorithm Page Rank), social influence analysis and social media analysis.

References Books:

1. Mining Text Data. Charu C. Aggarwal and Cheng Xiang Zhai, Springer, 2012.
2. Speech & Language Processing. Dan Jurafsky and James H Martin, Pearson Education India, 2000.
3. Introduction to Information Retrieval. Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schuetze, Cambridge University Press, 2007.
4. Foundations of Statistical Natural Language Processing by Christopher Manning and Hinrich Schütze.
5. Natural Language Processing with Pythonby Steven Bird, Ewan Klein and Edward Loper.
6. Survey of Text Mining Clustering, 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**
CO1	3	3			1	1		1	3
CO2		3			1	1		1	3
CO3		3			1	1		1	3
CO4		3		3	1	1		1	3
CO5		3			1	1		1	3
CO6		3			1	1		1	3
CO7		3	3		1	1	3	1	3

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

PO1. Disciplinary Knowledge

CO1. Grasp the significance of natural language processing in solving real-world problems. (Weightage: 3 - Strongly Related)

Justification: Grasping the significance of natural language processing contributes directly to disciplinary knowledge in this field.

PO2. Critical Thinking and Problem Solving

All COs (Weightage: 3 - Strongly Related)

Justification: All outcomes involve critical thinking and problem-solving skills in the context of natural language processing.

PO3. Social Competence

CO7. Effectively communicate the results of text mining and NLP analyses in written reports and oral presentations. (Weightage: 3 - Strongly Related)

Justification: Effective communication of NLP analyses results contributes to social competence.

PO4. Research-related Skills and Scientific Temper

CO4. Comprehend the state-of-the-art advanced NLP research articles and present them to an audience. (Weightage: 3 - Strongly Related)

Justification: Comprehending advanced NLP research articles aligns closely with research-related skills and the development of a scientific temper in the context of NLP.

PO5. Trans-disciplinary Knowledge

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are primarily focused on NLP techniques and principles, with limited direct connection to trans-disciplinary knowledge.

PO6. Personal and Professional Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes contribute more to technical competence in natural language processing, with limited direct connections to personal and professional aspects.

PO7. Effective Citizenship and Ethics

CO7. Effectively communicate the results of text mining and NLP analyses in written reports and oral presentations. (Weightage: 3 - Strongly Related)

Justification: Effective communication aligns with effective citizenship and ethical considerations.

PO8. Environment and Sustainability

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more focused on NLP techniques and principles, with less direct relevance to environmental or sustainability aspects.

PO9. Self-directed and Life-long Learning

All COs (Weightage: 3 - Strongly Related)

Justification: Continuous exploration of NLP techniques and principles aligns well with self-directed and life-long learning in the context of natural language processing.

SYLLABUS (CBCS) FOR M.Sc. Data Science

(2022 Pattern)

(With effect from Academic Year 2022-2023)

Class : M. Sc. (Semester- III)
Paper Code : PSDS236
Paper : VI
Title of Paper : Data Visualization Using Tableau
Credit : 4 credits
No. of lectures : 60

Course Objectives:

1. The main focus of Tableau software is for you to better understand your datasets, especially large datasets.
2. Able to handle 'big' data.
3. To implement the best design practices, and use the most appropriate chart for a particular situation.
4. To build interactive Tableau dashboards and construct a data story using Tableau Story point

Course Outcomes:

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

CO1. understand the importance of data visualization in conveying complex information.
CO2. learn how to connect Tableau to various data sources.
CO3. gain proficiency in using the Tableau software.
CO4. create fundamental visualizations, including bar charts, line charts, scatter plots, and pie charts.
CO5. develop skills in designing interactive dashboards..
CO6. gain knowledge of working with real-time data in Tableau.
CO7. explore ethical considerations in data visualization.

TOPICS/CONTENTS:

1. **Understanding Data:** What is data, Foundations for building Data Visualizations, Getting started with Tableau Software, Using Data file formats, Connecting Data to Tableau, Creating basic charts (line, bar charts, Treemaps), Using the Show me panel.
2. **Tableau Calculations:** Overview of SUM, AVR, and Aggregate features, Creating custom calculations and fields, Applying new data calculations to your visualization.
3. **Formatting Visualizations:** Formatting Tools and Menus, Formatting specific parts of the view, Editing and Formatting Axes.

4. **Manipulating Data in Tableau:** Cleaning-up the data with the Data Interpreter, Structuring your data, Sorting and filtering Tableau data, Pivoting Tableau data.
5. **Advanced Visualization Tools:** Using Filters, Using the Detail panel, Using the Size panels, Customizing filters, Using and Customizing tooltips, Formatting your data with colours.
6. **Creating Dashboards & Stories:** Using Storytelling, Creating dashboard and Story, Design for different displays, Adding interactivity to Dashboard.
7. **Distributing & Publishing Visualization:** Tableau file types, Publishing to Tableau Online, Sharing visualization.

Programme Outcomes and Course Outcomes Mapping:

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

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

PO1. Disciplinary Knowledge

CO1. Understand the importance of data visualization in conveying complex information. (Weightage: 3 - Strongly Related)

Justification: Understanding the importance of data visualization directly contributes to disciplinary knowledge in the field.

PO2. Critical Thinking and Problem Solving

All COs (Weightage: 3 - Strongly Related)

Justification: All outcomes involve critical thinking and problem-solving skills in the context of data visualization.

PO3. Social Competence

CO7. Explore ethical considerations in data visualization. (Weightage: 3 - Strongly Related)

Justification: Exploring ethical considerations in data visualization aligns with social competence, considering the impact and implications of visualizations on the audience.

PO4. Research-related Skills and Scientific Temper

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more focused on practical skills in data visualization rather than in-depth research-related skills or scientific temper.

PO5. Trans-disciplinary Knowledge

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are primarily focused on data visualization techniques, with limited direct connections to trans-disciplinary knowledge.

PO6. Personal and Professional Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes contribute more to technical competence in data visualization, with limited direct connections to personal and professional aspects.

PO7. Effective Citizenship and Ethics

CO7. Explore ethical considerations in data visualization. (Weightage: 3 - Strongly Related)

Justification: Exploring ethical considerations aligns with effective citizenship and ethical considerations in the context of data visualization.

PO8. Environment and Sustainability

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more focused on data visualization techniques, with less direct relevance to environmental or sustainability aspects.

PO9. Self-directed and Life-long Learning

All COs (Weightage: 3 - Strongly Related)

Justification: Continuous exploration of data visualization techniques aligns well with self-directed and life-long learning in the context of this field.