Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
M.Tech CSE (BDA/DATA SCIENCE)
SEMESTER- II (2025-27)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS201N	DCC	Advance Computer Architecture & Organization	60	20	20	-	-	3	0	0	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

COURSE OBJECTIVES:
The student will have ability to:
1. Describe current and emerging trends in computer architectures, focusing on performance and the hardware/software interface.
2. Analyzing fundamental issues in architecture design and their impact on application performance.
3. Identify the performance and efficiency in advanced multiple-issue processors
4. Identify and Analyzing various memory models.
5. Describe various techniques to enhance a processors ability.

COURSE OUTCOMES:
Upon completion of the subject, students will be able to:
1. Know the classes of computers, and new trends and developments in computer architecture
2. Understand pipelining, instruction set architectures, memory addressing.
3. Understand the performance metrics of microprocessors, memory, networks.
4. Understand the various techniques to enhance a processors ability to exploit Instruction-level parallelism (ILP), and its challenges.
5. Understand exploiting ILP using dynamic scheduling, multiple issue, and speculation.
6. Understand multithreading by using ILP and supporting thread-level parallelism (TLP).
7. Understand the performance and efficiency in advanced multiple-issue processors.
8. Understand symmetric shared-memory architectures and their performance.
9. Understand multiprocessor cache coherence using the directory based and snooping class of protocols.
10. Understand the various models to achieve memory consistency.

SYLLABUS
UNIT I — 10 HOURS
Overview of Parallel Processing and Pipelining Processing, study and comparison of uni-processors and parallel processors, Evolution of parallel processors, Necessity of high performance, Architectural Classification, Applications of parallel processing, Instruction level Parallelism and Thread Level Parallelism.

UNIT II — 9 HOURS
Principles and implementation of Pipelining, Pipeline Architecture, Study and comparison of processors with and without pipelining, Linear pipeline processor, Nonlinear pipeline processor Instruction pipeline design, Mechanisms for instruction pipelining, pipeline hazards, Dynamic instruction scheduling -score boarding and Tomosulo's algorithm, Branch handling techniques, Arithmetic Pipeline Design, Static arithmetic pipeline, Multifunctional arithmetic pipelines. Superscaler pipeline design, Super pipeline processor design.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
M.Tech CSE (BDA/DATA SCIENCE)
SEMESTER- II (2025-27)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS201N	DCC	Advance Computer Architecture & Organization	60	20	20	-	-	3	0	0	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

UNIT III **8 HOURS**

Study and comparison of Vector and array processors, Vector Processing Principles, Vector instruction types, Vector-access memory schemes. Vector supercomputer architecture, SIMD Computer Organization Masking and Data network mechanism, distributed memory model and shared memory model, Parallel Algorithms For Array Processors: Matrix Multiplication. Sorting, SIMD computer organization, Implementation issues of Matrix multiplication and sorting on array processor and their analysis

UNIT IV **7 HOURS**

Microprocessor Architectures, study and comparison of Loosely and Tightly coupled multiprocessors.Loosely and Tightly coupled multiprocessors, Processor characteristics of multiprocessors, Inter Processor communication network, Time shared bus, Crossbar switch, Multiport Memory Model, Memory contention and arbitration techniques, Cache coherence, Snoopy protocols, Directory based protocols. Message routing schemes in multicomputer network, deadlock and virtual channel.

UNIT V **8 HOURS**

Study of Architecture of Multithreaded processors, Latency hiding techniques, Principles of multithreading, Issues and solutions, Parallel Programming Techniques: Message passing program development, Synchronous and asynchronous message passing, Message passing parallel programming, Shared Memory Programming, Data Parallel Programming. Implementation issues of a multithreaded program.

TEXTBOOKS:

1. Kai Hwang, Faye A. Briggs, "Computer Architecture and Parallel Processing" McGraw- Hill international Edition
2. J.P.Hayes, “computer Architecture and organization”; MGH.

REFERENCE:

1. V.Rajaranam & C.S.R.Murthy, “Parallel computer”; PHI Learning.
2. Kain,”Advance Computer Architecture: - A System Design Approach”, PHI Learning
3. M.J Flynn, “Computer Architecture, Pipelined and Parallel Processor Design”; Narosa Publishing.
4. Hwang and Briggs, “Computer Architecture and Parallel Processing”; MGH.
5. Hwang and Briggs, “Computer Architecture and Parallel Processing”; MGH.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
M.Tech CSE (BDA/DATA SCIENCE)
SEMESTER- II (2024-26)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS202N	DCC	Advance Database Management System	60	20	20	30	20	2	0	2	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

COURSE OBJECTIVES:
The student will have ability to:
1. Describe database management system internals. Understand and describe internal algorithms in detail. Decide on optimization issues given a known database workload, by manipulating indexes, choosing more adequate data types, and modifying queries.
2. Identify opportunities for the use of the object model, and design and code client code to manipulate an object database.
3. Analyze and optimize transactional code, identifying causes of possible anomalies and correct them.
4. Identify and be able to use recent and advanced database techniques (e.g. in concurrency control, buffer management, and recovery).
5. Analyze, describe and use other models than the Relational. Analyze, compare and evaluate alternative database architectures and models in different application contexts. Identify limitations of the standard Relational databases in certain application domains, e.g. for multidimensional data, or unstructured data.

COURSE OUTCOMES:
Upon completion of the subject, students will be able to:
1. Be familiar with a commercial relational database system (Oracle) by writing SQL using the system.
2. Be familiar with the relational database theory, and be able to write relational algebra expressions for queries.
3. Be familiar with basic database storage structures and access techniques: file and page organizations, indexing methods including B-tree, and hashing.
4. Master the basics of query evaluation techniques and and query optimization.
5 Be familiar with the basic issues of transaction processing and concurrency control.

SYLLABUS
UNIT I 10 HOURS
Introduction to Database Systems: Database System Concepts and Architecture, Data Models, Data Independence, SQL: DDL, DML, DCL, Normalization: 1NF, 2NF, 3NF, BCNF, 4NF, 5NF.Query Processing and Optimization: Query Processing, Syntax Analyzer, Query Decomposition, Query Optimization, Heuristic Query Optimization, Cost Estimation, Cost Functions for Select, Join, Query Evaluation Plans.

UNIT II 9 HOURS
Object Oriented and Object Relational Databases Object Oriented Concepts, Object Oriented Data Model, Object Definition Language, Object Query Language, Object Relational Systems, SQL3, ORDBMS Design

UNIT III 8 HOURS

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS202N	DCC	Advance Database Management System	60	20	20	30	20	2	0	2	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

Transaction Processing and Concurrency Control: Transaction Processing Concepts, Concurrency Control Techniques: Two-phase Locking, Timestamp Ordering, Multiversion, Validation, Multiple Granularity Locking.

UNIT IV **7 HOURS**

Backup and Recovery: Types of Database Failures, Types of Database Recovery, Recovery Techniques: Deferred Update, Immediate Update, Shadow Paging, Checkpoints, Buffer Management.

UNIT V **8 HOURS**

Introduction to Data Warehousing and Data Mining: Introduction to OLAP, OLTP, Data Warehouse, Data Marts, Data Mining, Data Mining Process. Distributed Databases: Distributed Database Concepts, Advantages and Disadvantages, Types of Distributed Database Systems, Data Fragmentation, Replication and Allocation Techniques for Distributed Database Design, Five Level Schema Architecture, Query Processing, Concurrency Control and Recovery in Distributed Databases. Commercial Databases: Commercial Database Products, Familiarity with IBM DB2 Universal Database, Oracle, Microsoft SQL Server, MySQL, their features.

TEXTBOOKS:

1. C. J. Date: An Introduction to Database Systems , Addison-Wesley
2. Avi Silberschatz, Henry F. Korth ,S. Sudarshan ,Data Base System Concepts, TMH

REFERENCE:

1. Patrick O'Neil & Elizabeth O'Neil, Database Principles, Programming and Performance,
2. Morgan Kaufmann Hardcourt India
3. Gillenson, Fundamental of Data Base Management Sytem, Willey India
4. Ceri & Pelagatti, Distributed Databases Principles & Systems,TMH
5. Paulraj Ponniah, Data Ware Housing Fundamental, Willey India.
6. Jiawei Han, Data Mining Concept & Techniques, Elsevier Pub.

LIST OF PRACTICALS

1. Distributed Database for Bookstore
2. Deadlock Detection Algorithm for distributed database using wait- for graph
3. Object Oriented Database – Extended Entity Relationship (EER)
4. Parallel Database – University Counselling for Engineering colleges
5. Parallel Database – Implementation of Parallel Join & Parallel Sort
6. Active Database – Implementation of Triggers & Assertions for Bank Database
7. Deductive Database – Constructing Knowledge Database for Kinship Domain (Family Relations)
8. Study and Working of WEKA Tool

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Information Technology

M.Tech CSE (BDA/DATA SCIENCE)

SEMESTER- II (2024-26)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS202N	DCC	Advance Database Management System	60	20	20	30	20	2	0	2	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

9. Query Processing – Implementation of an Efficient Query Optimizer
10. Designing XML Schema for Company Database

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
M.Tech CSE (BDA/DATA SCIENCE)
SEMESTER- II (2024-26)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTRM301	DCC	Research Methodology in Engineering	60	20	20	0	0	3	1	0	4

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

COURSE OBJECTIVES:
The student will have ability to:
1. The ensuing knowledge as property.
2. To plan and design business research using scientific and statistical methods.

COURSE OUTCOMES:
Upon completion of the subject, students will be able to:
1. Demonstrate understanding of research methodology.
2. Apply the statistical concepts in business research.
3. Validate statistical statements relating to business research.

SYLLABUS
UNIT I — 10 HOURS
Business Research
1. An overview: Research process
2. Types of Research - Exploratory Research, Descriptive Research, Causal Research, Analytical Research
3. Problem formulation, Management problem v/s. Research problem
4. Approaches to Research
5. Importance of literature review
6. Business Research Design: Steps involved in a research design

UNIT II — 9 HOURS
Sampling and Data Collection
1. Sampling and sampling distribution: Meaning, Steps in Sampling process
2. Types of Sampling - Probability and Non probability Sampling Techniques
3. Data collection: Primary and Secondary data – Sources – Advantages/Disadvantages
4. Data collection Methods: Observations, Survey, Interview and Questionnaire design, Qualitative Techniques of data collection.

UNIT III — 8 HOURS
Measurement and Scaling Techniques
1. Nominal Scale, Ordinal Scale, Interval Scale, Ratio Scale, Criteria for good measurement
2. Attitude measurement – Likert's Scale, Semantic Differential Scale, Thurstone-equal appearing interval scale

UNIT IV — 10 HOURS

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTRM301	DCC	Research Methodology in Engineering	60	20	20	0	0	3	1	0	4

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

Statistical Tools for Data Analysis

1. Measures of central tendency - Mean, Median, Mode ,Quartiles, Deciles and Percentiles
2. Measures of Dispersion: Standard Deviation – Variance – Coefficient of Variance, Skewness
3. Correlation - Karl Pearson's coefficient of Correlation, Rank Correlation
4. Regression: Method of Least Squares
5. Formulation of hypothesis
6. Testing of hypothesis
7. Type I and Type II Errors.
8. Parametric tests: Z-Test, t-test, F-test, Analysis of Variance – One-Way and Two-way classification.
9. Non parametric tests - Chi-Square test

UNIT V **9 HOURS**

Report writing

1. Reporting Research
2. Types of reports
3. Characteristics of a research report

TEXTBOOKS:

1. MalhotraNaresh K. (2008). Marketing Research. Pearson publishers, Latest Edition.
2. Zikmund, Babin,Carr,Griffin (2003). Business Research Methods. Cengage Learning, India, Latest Edition.

REFERENCE:

1. Cooper Donald R and Schindler Pamela S. (2006). Business Research Methods. McGraw-Hill Education, Latest Edition.
2. Anderson, Sweeney, William, Cam (2014). Statistics for Business and Economics. Cengage Learning, Latest Edition.
3. Krishnaswami O. R., Ranganatham M. (2011). Methodology of Research in Social Sciences. Himalaya Publishing House, Latest Edition.
4. Levin and Rubin (2008). Statistics for Management. Dorling Kindersley Pvt Ltd, Latest Edition.
5. ekaran Uma (2003). Research Methods for Business. Wiley India, Latest Edition.
6. Gupta S. P. (2014). Statistical Methods. Sultan Chand and Sons, Latest Edition.
7. Aczel and Sounderpandian (2008). Complete Business Statistics. Tata-McGraw Hill,Latest Edition.
8. Kothari C. R. (2004). Research Methodology. VishwaPrakashan, Latest Edition.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
M.Tech CSE (BDA/DATA SCIENCE)
SEMESTER- II (2024-26)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS2031	DSE	Optimization Techniques for Analytics	60	20	20	30	20	2	0	2	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

COURSE OBJECTIVES:
The student will have ability to:
1. Develop a Comprehensive Understanding
2. Master Data Preparation and Modeling:
3. Apply Advanced Analytics Techniques:

COURSE OUTCOMES:
Upon completion of the subject, students will be able to:
1. Proficiency in Data Analytics Lifecycle:
2. Hands-On Data Preparation and Analysis Skills:
3. Advanced Analytical Techniques Application:

SYLLABUS
UNIT I — 8 HOURS
Introduction to Data Analytics and Lifecycle : Data Analytics Lifecycle overview: Key Roles for a Successful Analytics, Background and Overview of Data Analytics Lifecycle, Discovery: Learning the Business Domain, Resources Framing the Problem, Identifying Key Stakeholders. Interviewing the Analytics Sponsor, Developing Initial Hypotheses Identifying, Potential Data Sources Project.

UNIT II — 8 HOURS
Data Preparation: Preparing the Analytic Sandbox, Performing ETLT, Learning About the Data, Data Conditioning, Survey and visualize, Common Tools for the Data Preparation Phase.
Model Planning: Data Exploration and Variable Selection, Model Selection, Common Tools for the Model Planning Phase. Model Building: Common Tools for the Model Building Phase

UNIT III — 10 HOURS
Introduction to simple Linear Regression: The Regression Equation, Fitted value and Residuals, Least Square Introduction to Multiple Linear Regression: Assessing the Model, Cross-Validation, Model Selection and Stepwise Regression, Prediction Using Regression. Logistic Regression: Logistic Response function and logit, Logistic Regression and GLM, Generalized Linear model, predicted values from Logistic Regression, Interpreting the coefficients and odds ratios.

UNIT IV — 10 HOURS
Text Analytics: History of text mining, Roots of text mining overview of seven practices of text analytic, Application and use cases for Text mining: extracting meaning from unstructured text, Summarizing Text. Text Analysis Steps, A Text Analysis Example, Collecting Raw Text, Representing Text, Term Frequency—Inverse Document Frequency (TFIDF), Categorizing Documents by Topics, Determining Sentiments, Gaining Insights.

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS2031	DSE	Optimization Techniques for Analytics	60	20	20	30	20	2	0	2	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

UNIT V 8 HOURS

Optimization: Introduction to Optimization, Unconstrained and Linear Constrained Optimization, Non-linear Constrained Optimization, Robust Optimization Techniques.

TEXTBOOKS:

1. An introduction to Optimization by Edwin P K Chong, Stainslaw Zak
2. Nonlinear Programming by Dimitri Bertsekas

REFERENCE:

1. Richard W. Cottle and Mukund N. Thapa, Linear and Nonlinear Optimization, ISBN 978-1-4939-7053-7, Springer New York (2017).
2. Gerard Cornuejols and Reha Tut¨ unc ¨ u¨, Optimization Methods in Finance, Fourth Printing, Cambridge University Press (2013).

LIST OF PRACTICALS

1. Conduct mock interviews with stakeholders to frame a business problem and develop initial hypotheses for a data analytics project.
2. Prepare an analytic sandbox, perform ETLT (Extract, Transform, Load, and Transform), and visualize the dataset using common tools.
3. Explore a given dataset, select relevant variables for model planning, and document the rationale behind your selections.
4. Implement a simple linear regression model using a provided dataset, analyze the fitted values and residuals, and interpret the results.
5. Develop a multiple linear regression model, assess the model, perform cross-validation, and use stepwise regression for variable selection.
6. Text Analytics Application
7. Solve an unconstrained and linear constrained optimization problem using robust optimization techniques.
8. Optimization Problem Solving: Solve an unconstrained and linear constrained optimization problem using robust optimization techniques.
9. Comprehensive Data Analytics Project

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
M.Tech CSE (BDA/DATA SCIENCE)
SEMESTER- II (2024-26)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS2032	DSE	Data Management for Machine Learning	60	20	20	30	20	2	0	2	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

COURSE OBJECTIVES:
The student will have ability to:
1. Identify and describe various data sources, data preprocessing techniques, and data visualization methods.
2. Explain the concepts and functions of data structures, algorithms, and database management systems.
3. Apply machine learning algorithms and performance metrics to analyze and visualize datasets.

COURSE OUTCOMES:
Upon completion of the subject, students will be able to:
1. Understand and explain the principles of data collection, preprocessing, and visualization techniques.
2. Analyze data structures and algorithms, interpreting their complexity and efficiency in real-world data processing tasks.
3. Apply machine learning techniques and evaluate their performance using appropriate metrics, addressing issues of overfitting and dimensionality reduction.

SYLLABUS
UNIT I — 10 HOURS
Data Analysis and Visualization Techniques Data Collection: Various sources and types of data: text, video, audio, biology etc; Data Preprocessing: Cleaning data, missing data imputation, noise elimination, feature selection and dimensionality reduction (PCA, tSNE, UMAP), normalisation. Data Visualization: histogram, pie chart, area-plot, box plot, scatter-plot, bubble plot, waffle charts

UNIT II — 9 HOURS
Data Structure and Algorithm Data Structure: Stack, Queue, Linked List, Graphs. Algorithms. Searching, Sorting, Graph traversal, Complexity (Time).

UNIT III — 8 HOURS
Database Management System Database: Schema, ER diagram, SQL (Postgresql), database normalization (1NF,2NF,3NF and BCNF), indexing (B+tree), transaction concept and simple transaction model, serializability concept, concurrency control.

UNIT IV — 7 HOURS
Foundations of machine learning: Regression Vs. Classification, Generalization, Training, Validation and Testing, Problem of Overfitting and mechanisms to overcome overfitting, Bias Vs. Variance, vector independence, matrices and vector spaces, probability recap – conditional and joint.

UNIT V — 8 HOURS

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
M.Tech CSE (BDA/DATA SCIENCE)
SEMESTER- II (2024-26)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS2032	DSE	Data Management for Machine Learning	60	20	20	30	20	2	0	2	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

Performance metrics: Classification, Regression & Clustering, Data Dimensionality – Curses of dimensionality, dimensionality reduction techniques. Machine learning algorithms applied for signal and image analysis.

TEXTBOOKS:
1. Data Engineering: A Novel Approach to Data Design. Brian Shive. Technics Publications. ISBN-13: 978-1935504603
2. Python Data Science Handbook: Essential Tools for Working with Data. Joel Grus. O'Reilly. ISBN-13: 978-9352134915

REFERENCE:
1. Introduction to Algorithms. Cormen, Leiserson, Rivest, Stein. MIT Press 3ed. ISBN-13: 978-0262533058
2. Database System Concepts. Silberschatz, Korth, Sudarshan. McGraw Hill Education; Sixth edition. ISBN-13: 978-9332901384
3. Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools. Cielen, Meysman,Ali. Dreamtech Press. ISBN-13: 978-9351199373

LIST OF PRACTICALS
1. Collect and analyze data from various sources, including text, video, audio, and biological data.
2. Perform data cleaning, handle missing data imputation, and eliminate noise from a given dataset.
3. Apply PCA, tSNE, and UMAP for feature selection and dimensionality reduction on a sample dataset.
4. Create various data visualizations such as histograms, pie charts, area plots, box plots, scatter plots, bubble plots, and waffle charts.
5. Implement and manipulate basic data structures such as stacks, queues, linked lists, and graphs.
6. Implement searching, sorting, and graph traversal algorithms and analyze their time complexity.
7. Design a database schema, create an ER diagram, normalize a database, and execute SQL queries using PostgreSQL.
8. Develop and evaluate regression and classification models, addressing issues of overfitting, bias, and variance.
9. Evaluate the performance of classification, regression, and clustering models using appropriate metrics.
10. Apply machine learning algorithms to analyze signal and image data, interpreting the results for practical applications.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
M.Tech CSE (BDA/DATA SCIENCE)
SEMESTER- II (2024-26)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS2033	DSE	Probabilistic Graphical Models	60	20	20	30	20	2	0	2	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

COURSE OBJECTIVES:
The student will have ability to:
1. Develop a Thorough Understanding of Probability and Statistics
2. Master Probabilistic Graphical Models and Mathematical Modelling
3. Apply Advanced Regression, Hypothesis Testing, and Time Series Analysis Techniques

COURSE OUTCOMES:
Upon completion of the subject, students will be able to:
1. Proficiency in Probability and Estimation Theory
2. Competency in Graphical Models and Mathematical Modelling
3. Expertise in Regression, Hypothesis Testing, and Time Series Analysis

SYLLABUS
UNIT I — 10 HOURS
Introduction: Samples, Events, Event space, Probability Space, Random Variables, Independence and Conditional Independence, Conditional Probability, Joint Probability, Bayes' theorem Joint and Marginal Probability, Estimation Theory - Maximum Likelihood Estimators.

UNIT II — 9 HOURS
Probabilistic Graphical Models: Direct and undirected model, Inference from Direct and undirected graphical model, Structured and Unstructured graphical models, Partition Function, D-Separation, Decision Analysis, Decision Trees, Influence Diagrams, Factor Graphs, Sampling from Graphical Models. Markov Process and Markov Chain.

UNIT III — 8 HOURS
Introduction to Mathematical Modelling: Linear system of equations, Eigenvalues and Eigenvectors, Stability of Systems using Eigen value analysis, Taylor series, Numerical differentiation, Numerical integration, Higher order accuracy schemes for integration and differentiation, Ordinary differential equations, Partial Differential Equations, Iterative methods.

UNIT IV — 7 HOURS
Probability: Random Variables & Probability Distributions. Sampling, analysis of sample dataEmpirical Distributions, Sampling from a Population Estimation, confidence intervals, point estimation--Maximum Likelihood, Probability mass functions, Modeling distributions, Hypothesis testing- Z, t, Chi-Square. ANOVA & Designs of Experiments - Single, Two factor ANOVA, Factorials ANOVA models.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
M.Tech CSE (BDA/DATA SCIENCE)
SEMESTER- II (2024-26)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS2033	DSE	Probabilistic Graphical Models	60	20	20	30	20	2	0	2	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

UNIT V 8 HOURS

Regression: Correlation & Regression Models-linear regression methods, Ridge regression, LASSO, univariate and Multivariate Linear Regression, probabilistic interpretation, Regularization, Logistic regression, locally weighted regression. Exploratory data analysis, Time series analysis, Analytical methods – ARIMA and SARIMA.

TEXTBOOKS:

1. Koller, Daphne, and Nir Friedman. Probabilistic graphical models: principles and techniques. MIT press, 2009.
2. Practical Statistics for Data Scientists, by Peter Bruce and Andrew Bruce, O'REILLY'

REFERENCE:

1. Think Stats 2e, Exploratory Data Analysis in Python, Allen B. Downey, O'REILLY', Green Tea Press.
2. Julian J. Faraway. Extending the Linear Model with R – Generalized Linear, Mixed Effects and Nonparametric Regression Models, Second Edition, CRC Press 2016
3. Michael Friendly and David Meyer. Discrete Data Analysis with R – Visualization and Modeling Techniques for Categorical and Count Data, CRC Press Dec 2015.

LIST OF PRACTICALS

1. Use real-world scenarios to apply Bayes' theorem and calculate conditional probabilities.
2. Develop maximum likelihood estimators for different probability distributions and validate their effectiveness with sample data.
3. Construct both directed and undirected probabilistic graphical models using given datasets.
4. Perform inference on graphical models, both structured and unstructured, and analyze the results.
5. olve linear systems of equations using different numerical methods, and analyze the stability of systems using eigenvalue analysis.
6. Perform sampling from different populations, analyze the sample data, and construct empirical distributions.
7. Conduct hypothesis tests (Z, t, Chi-Square) and ANOVA for single and two-factor experiments to analyze the significance of results.
8. Develop linear, ridge, LASSO, and logistic regression models, and compare their performances using univariate and multivariate data.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
M.Tech CSE (BDA/DATA SCIENCE)
SEMESTER- II (2024-26)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS2041	DSE	Introduction to Statistical Methods	60	20	20	30	20	2	0	2	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

COURSE OBJECTIVES:
The student will have ability to:
1. Understanding Probability and Distributions
2. Mastering Descriptive Statistics and Estimation Techniques
3. Applying Regression and Correlation Analysis

COURSE OUTCOMES:
Upon completion of the subject, students will be able to:
1. Proficiency in Probability Concepts and Distributions
2. Competency in Descriptive Statistics and Estimation
3. Expertise in Regression and Correlation Analysis

SYLLABUS
UNIT I — 10 HOURS
Introduction to probability: Probability, conditional probability, independence, Bayes' theorem. Random variables: discrete and continuous, distribution functions and their properties, probability mass and density functions, expectation & moments, moment generating function & its properties. Multiple random variables: joint distributions, marginal and conditional distributions.
Discrete probability distributions: Bernoulli, Binomial, Geometric, Negative Binomial, Hypergeometric and Poisson distribution. Continuous probability distributions: Uniform, Exponential, Gamma, Normal & Log-normal distribution.

UNIT II — 9 HOURS
Descriptive Statistics: Inferential statistics, population, sample, parameter, statistic, random sample, sampling techniques. Summarizing and Exploring Data: Concept of frequency distribution, measures of central tendency, moments, measures of dispersion/variability, measures of skewness and kurtosis.

UNIT III — 8 HOURS
Estimation: Sampling distributions, basic concepts of inference (estimation & hypothesis testing), point estimation & interval estimation

UNIT IV — 10 HOURS
Testing of Hypothesis: Null and alternate hypothesis, simple & composite hypotheses, critical region, N-P lemma, tests for mean, variance and proportion in one and two sample problems. Chi-square goodness of fit test. Introduction to nonparametric test, Contingency table, test of independence

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
M.Tech CSE (BDA/DATA SCIENCE)
SEMESTER- II (2024-26)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS2041	DSE	Introduction to Statistical Methods	60	20	20	30	20	2	0	2	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

UNIT V **8 HOURS**

Regression and Correlation: Simple linear regression, least squares fit and correlation analysis. Tests for slope & correlation, prediction problem, residual plots. Multiple linear regression. Analysis of Variance.

TEXTBOOKS:
1. *Probability and Statistics in Engineering* by Hines, Montgomery, Goldsman & Borror. Wiley Student Edition.
2. *An Introduction to Probability and Statistics* by Rohatgi and Saleh. Wiley

REFERENCE:
1. *Introduction to Probability Models* by Sheldon M. Ross, Academic Press.
2. *Probability and Statistics* by Spiegel, Schiller and Srinivasan. Tata McGraw-Hill Pub. Co. Ltd.
3. *Miller and Freund's Probability and Statistics for Engineers* by Johnson/Miller, Pearson Education India.
4. *Introduction to Probability and Statistics* by J. Susan Milton & J.C. Arnold, 4th Ed.,

LIST OF PRACTICALS
1. :Conduct experiments to calculate probabilities using classical, frequency, and axiomatic approaches. Apply Bayes' theorem to real-world scenarios.
2. Implement and analyze the Bernoulli, Binomial, Geometric, Negative Binomial, Hypergeometric, and Poisson distributions using sample data.
3. Develop and interpret continuous probability distributions
4. Write Python scripts to perform descriptive statistical analysis and visualize various probability distributions.
5. Calculate point estimates and construct confidence intervals for different parameters using sample data. Perform estimation using Python.
6. Hypothesis Testing for Means and Variances
7. Chi-Square Goodness of Fit and Independence Tests
8. Simple and Multiple Linear Regression
9. ANOVA and Correlation Analysis

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
M.Tech CSE (BDA/DATA SCIENCE)
SEMESTER- II (2024-26)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS2042	DSE	DATA WAREHOUSING	60	20	20	30	20	2	0	2	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

COURSE OBJECTIVES:
The student will have ability to:
1. Understand the components and architecture of data warehousing and data mining systems.
2. Implement data preprocessing, cleaning, integration, and transformation techniques.
3. Analyze and evaluate clustering methods, classification algorithms, and regression models for data analysis.

COURSE OUTCOMES:
Upon completion of the subject, students will be able to:
1. Describe the architecture and functionalities of data warehousing and data mining systems.
2. Implement data extraction, cleanup, and transformation processes using appropriate tools.
3. Evaluate various clustering, classification, and regression methods, assessing their effectiveness in different scenarios.

SYLLABUS
UNIT I
Data Warehousing and Business Analysis: Data warehousing Components –Building a Data warehouse –Data Warehouse Architecture – DBMS Schemas for Decision Support – Data Extraction, Cleanup, and Transformation Tools –Metadata – reporting – Query tools and Applications – Online Analytical Processing (OLAP) – OLAP and Multidimensional Data Analysis.

UNIT II
Data Mining: Data Mining Functionalities – Data Preprocessing – Data Cleaning – Data Integration and Transformation – Data Reduction – Data Discretization and Concept Hierarchy Generation- Architecture Of A Typical Data Mining Systems- Classification Of Data Mining Systems.
Association Rule Mining: - Efficient and Scalable Frequent Item set Mining Methods – Mining Various Kinds of Association Rules – Association Mining to Correlation Analysis – Constraint-Based Association Mining.

UNIT III
Classification and Prediction: Issues Regarding Classification and Prediction – Classification by Decision Tree Introduction – Bayesian Classification – Rule Based Classification – Classification by Back propagation – Support Vector Machines – Associative Classification – Lazy Learners – Other Classification Methods – Prediction – Accuracy and Error Measures – Evaluating the Accuracy of a Classifier or Predictor – Ensemble Methods – Model Section.

UNIT IV
Cluster Analysis: Types of Data in Cluster Analysis – A Categorization of Major Clustering Methods –

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS2042	DSE	DATA WAREHOUSING	60	20	20	30	20	2	0	2	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

Partitioning Methods – Hierarchical methods – Density-Based Methods – Grid-Based Methods – Model-Based Clustering Methods – Clustering High-Dimensional Data – Constraint-Based Cluster Analysis – Outlier Analysis

UNIT V

Mining Object, Spatial, Multimedia, Text and Web Data: Multidimensional Analysis and Descriptive Mining of Complex Data Objects – Spatial Data Mining – Multimedia Data Mining – Text Mining – Mining the World Wide Web.

TEXTBOOKS:

1. Jiawei Han, Micheline Kamber and Jian Pei“Data Mining Concepts and Techniques”, Third Edition, Elsevier, 2011.
2. Alex Berson and Stephen J. Smith “Data Warehousing, Data Mining & OLAP”, Tata McGraw – Hill Edition, Tenth Reprint 2007.

REFERENCE:

1. K.P. Soman, Shyam Diwakar and V. Ajay “Insight into Data mining Theory and Practice”, Easter Economy Edition, Prentice Hall of India, 2006.
2. G. K. Gupta “Introduction to Data Mining with Case Studies”, Easter Economy Edition, Prentice Hall of India, 2006.
3. Pang-Ning Tan, Michael Steinbach and Vipin Kumar “Introduction to Data Mining”, Pearson Education, 2007.

LIST OF PRACTICALS

1. Build and implement a data warehouse architecture using a sample dataset.
2. Perform data extraction, cleanup, and transformation using ETL tools.
3. Design and execute OLAP queries for multidimensional data analysis.
4. Conduct data preprocessing, cleaning, and integration on raw data.
5. Implement frequent item set mining and association rule mining techniques.
6. Classify data using decision tree and Bayesian classification methods.
7. Evaluate the accuracy and error measures of different classifiers.
8. Apply partitioning, hierarchical, and density-based clustering methods to a dataset.
9. Analyze multimedia and spatial data using data mining techniques.
10. Perform text mining and web mining on real-world data sources.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
M.Tech CSE (BDA/DATA SCIENCE)
SEMESTER- II (2024-26)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS2043	DSE	Ethics for Data Science	60	20	20	30	20	2	0	2	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

COURSE OBJECTIVES:
The student will have ability to:
1. Understand the Fundamentals of Ethics
2. Analyze Ethical Implications of AI and Technology
3. Apply Ethical Principles in Real-world Scenarios

COURSE OUTCOMES:
Upon completion of the subject, students will be able to:
1. Comprehend Key Ethical Concepts
2. Evaluate Ethical Challenges in AI and Technology
3. Implement Ethical Practices

SYLLABUS
UNIT I
Introduction to Ethics: What are Ethics?, History, Concept of Informed Consent, Data Ownership, Privacy, Anonymity, Data Validity, Algorithmic Fairness, Societal Consequences, Code of Ethics, Attributions.

UNIT II
Ethics in the age of AI: Artificial Intelligence Algorithms Models and Limitations, Artificial Intelligence Data Fairness and Bias, Artificial Intelligence Privacy and Convenience, Artificial Intelligence Ethics in Action, The Complex World of Data, The Challenges & Strategies of Putting Ethics into Practice.

UNIT III
Ethics, Technology, and Engineering – Eindhoven: Introduction, Codes of Conduct, Normative Ethics, The Ethical Cycle, Ethical Questions in the Design of Technology, Designing Morality, Ethical Aspects of Technological Risks, Distribution of Responsibility.

UNIT IV
Data Ethics, AI, and Responsible Innovation: Getting Started and Big Data Opportunities, Big Data Limitations, Research Ethics, Law and Ethics, Crime and Justice, Home and City, Money and Markets, Life and Health.

UNIT V
Ethical Implications of Emerging Technologies: New and upcoming technologies and their potential impact on society. Ethical Challenges and Considerations, Case Studies of Emerging Technologies, Regulation and Policy, Future Directions in Technology Ethics.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
M.Tech CSE (BDA/DATA SCIENCE)
SEMESTER- II (2024-26)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS2043	DSE	Ethics for Data Science	60	20	20	30	20	2	0	2	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

TEXTBOOKS:
1. American Civil Liberties Union. (2004). "Scary Pizza." (video.) 01:42. Available online: https://www.youtube.com/watch?v=33CIVjvYyEk
2. Raji, Deborah. (December 10, 2020). "How our data encodes systematic racism." MIT Technology Review.

REFERENCE:
1. Loukides, Mike, Hilary Mason, and DJ Patil. 2018. Ethics and Data Science. Sebastopol, CA: O'Reilly Media. Chapter 1, "Doing Good Data Science"
2. Loukides, Mike, Hilary Mason, and DJ Patil. 2018. Ethics and Data Science. Sebastopol, CA: O'Reilly Media. Chapter 3, "The Five Cs."
3. Zook, Matthew, Solon Barocas, Kate Crawford, Emily Keller, Alyssa Goodman, Rachelle Hollander, Barbara A. Koenig, Jacob Metcalf, Arvind Narayanan, Alondra Nelson, and Frank Pasquale. 2017.

LIST OF PRACTICALS
1. **Facebook-Cambridge Analytica Scandal**: This case study explores the misuse of data by Cambridge Analytica, which harvested personal data from millions of Facebook users without their consent. It raises issues of data ownership, privacy, informed consent, and the societal consequences of data misuse.
2. **COMPAS Recidivism Algorithm**: This case study examines the use of the COMPAS algorithm in the US criminal justice system to predict recidivism rates. It highlights issues of algorithmic fairness, bias, and the ethical implications of using AI in decision-making processes that affect people's lives.
3. **Volkswagen Emissions Scandal**: This case study discusses the ethical breaches by Volkswagen when they installed software in diesel engines to cheat emissions tests. It covers topics such as codes of conduct, normative ethics, the ethical cycle, and the distribution of responsibility in technological design.
4. **Google Street View Privacy Concerns**: This case study explores the privacy issues raised by Google's Street View project, which involved capturing images of streets and homes worldwide. It addresses the balance between innovation and privacy, research ethics, and the societal impact of big data projects.
5. **CRISPR Gene Editing**: This case study examines the ethical challenges and considerations surrounding the use of CRISPR technology for gene editing. It discusses the potential impact on society, regulation and policy, and the future directions in technology ethics.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
M.Tech CSE (BDA/DATA SCIENCE)
SEMESTER- II (2024-26)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS207	SEC	PYTHON FOR DATA SCIENCE LAB	0	0	0	0	50	0	0	2	1

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

COURSE OBJECTIVES:
The student will have ability to:
1. To explore the programming skills relevant to data science and to gain knowledge of various libraries and packages like NumPy, Pandas and Matplotlib required for data analysis, data visualization, natural language processing and machine learning.

COURSE OUTCOMES:
Upon completion of the subject, students will be able to:
1. To understand data types in python and to apply array concepts using NumPy.
2. Structuring data using NumPy and manipulating the data using Pandas.
3. Using Pandas to analyze and work with data sets.
4. Graphical visualization of data using Matplotlib.

SYLLABUS
UNIT I
Compilation v/s Interpretation · Script mode and Interactive mode · Command Line Arguments, Data Types: Numbers, string, bool, Advanced Data Types: list, tuple set, dictionary, Type casting: implicit, explicit.

UNIT II
Function: User defined function, Built in functions, Lambda Function: filter, reduce, map, reverse function, doc string, types of arguments: positional, default, keyword, variable length, variable length keyword.

UNIT III
Exception Handling: User defined exception handler, disadvantages of having single except block, grouping multiple exception in single except block, printing the default message of exception by aliasing, optional else block, propagation of exception object.

UNIT IV
File handling: Reading a file, modes of file, closing a file, context managers in python, reading file contents line by line, reading a single line in a file, reading a multiple lines in a file, reading a character in a line, readline()

UNIT V
OOPS concepts: Encapsulation, polymorphism, inheritance, abstraction, operator overloading and magical methods, multi threading in python

TEXTBOOKS:
1. Python Data Science Handbook: Essential Tools for Working with Data, by Jake VanderPlas, O'reilly Media, 2017.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
M.Tech CSE (BDA/DATA SCIENCE)
SEMESTER- II (2024-26)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
MTCS207	SEC	PYTHON FOR DATA SCIENCE LAB	0	0	0	0	50	0	0	2	1

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

REFERENCE:

1. Python for Data Science by Mohd. Abdul Hameed, May 2021, Wiley.
2. Python for Data Science: The Ultimate Step-by-Step Guide to Python Programming by Daniel, March 2021, O'reilly.
3. Python for Data Science: A Crash Course for Data Science and Analysis, Python Machine Learning and Big Data by Computer Science Academy

LIST OF PRACTICALS
1. Write a program to demonstrate a) Different numeric data types and b) To perform different Arithmetic Operations on numbers in Python
2. Write a program to create, append, and remove lists in Python.
3. Write a program to demonstrate working with tuples in Python
4. Write a program to demonstrate working with dictionaries in Python.
5. Write a program to demonstrate a) arrays b) array indexing such as slicing, integer array indexing and Boolean array indexing along with their basic operations in NumPy.
6. Write a program to compute summary statistics such as mean, median, mode, standard deviationand variance of the given different types of data.
7. Write a script named copyfile.py. This script should prompt the user for the names of two text files. The contents of the first file should be the input that to be written to the second file.
8. Write a program to demonstrate Regression analysis with residual plots on a given data set.
9. Write a program to demonstrate the working of the decision tree-based ID3 algorithm.
10. Write a program to implement the Naïve Bayesian classifier for a sample training data set stored as a .CSV file.