Estd. 1962
"A++" Accredited by NAAC (2021)
With CGPA 3.52

SHIVAJI UNIVERSITY, KOLHAPUR
416 004, MAHARASHTRA
PHONE : EPABX – 2609000, BOS Section – 0231-2609094, 2609487
Web : www.unishivaji.ac.in Email: bos@unishivaji.ac.in
शिवाजी विद्यापीठ, कोल्हापूर ४१६ ००४, महाराष्ट्र
दूरध्वनी - इपीबीएक्स - २०६०९०००, अभ्यासमंडळे विभाग : ०२३१- २६०९०९४, २६०९४८७
वेबसाईट : www.unishivaji.ac.in ईमेल : bos@unishivaji.ac.in

SU/BOS/Sci & Tech/ 540 **Date: 08/09/2025**

To,

The Principal,
All Concerned Affiliated Colleges/Institutions
Shivaji University, Kolhapur

I/C- Director,
Yashwantrao Chavan School of Rural Development,
Shivaji University, Kolhapur

Subject: Regarding revised syllabi of MCA & M.Sc Part-I (Sem.I & II) degree programme under the Faculty of Science and Technology as per NEP-2020 (2.0)

Ref: No. SU/BOS/Science/651 Date: 31/08/2023 Letter.

Sir/Madam,

With reference to the subject mentioned above, I am directed to inform you that the university authorities have accepted and granted approval to the syllabi, nature of question paper of MCA & M.Sc. Part-I (Sem.I & II) degree programme under the Faculty of Science and Technology as per NEP-2020 (2.0).

Part - I (Sem- I – II) 2025-26	
1	MCA (YCSRD) (ON Campus)
2	M.Sc. Data Scienc (OFF Campus)

This syllabus, nature of question and equivalence shall be implemented from the academic year 2025-2026 onwards. A soft copy containing the syllabus is attached herewith and it is also available on university website www.unishivaji.ac.in NEP-2020@suk (Online Syllabus)

The question papers on the pre-revised syllabi of above-mentioned course will be set for the examinations to be held in October /November 2025 & March/April 2026. These chances are available for repeater students, if any.

You are, therefore, requested to bring this to the notice of all students and teachers concerned.

Thanking you,

Yours faithfully,

Dr. S. M. Kubal
Dy. Registrar

Encl. : As above.

Copy to: For Information and necessary action.

1	I/c Dean, Faculty of Science & Technology	7	Appointment Section A & B
2	Director, Board of Examinations & Evaluation	8	Affiliation Section (T.1) (T.2)
3	The Chairpersan, Respective Board of Studies	9	P.G.Admission Section,
4	B.Sc. Exam Section	10	Computer Centre / IT Cell
5	Eligibility Section	11	Internal Quality Assorance Cell (IQAC)
6	P.G Seminar Section		

SHIVAJI UNIVERSITY,KOLHAPUR

Established:1962
A^{++}AccreditedbyNAAC(2021)withCGPA3.52

StructureandSyllabusinAccordancewith National Education Policy - 2020

withMultipleEntryandMultipleExit

MasterofScience (DataScience)

(Under Faculty of Science and Technology)

PART-I SEMESTER - I &II

(Syllabus to be implemented from Academic year 2025-26)

1. Preamble

M.Sc. Data Science , an innovative program, is introduced in Smt. KasturbaiWalchand College of Arts and Science, Sangli from June 2022 and opened the doors of ubiquitous technology knowledge. It is an interdisciplinary field of study which uses scientific processes, approaches, methods, system and algorithm to extract insights and information from structured and unstructured data. With the view to provide exposure to the recent technologies of various sectors of the Data Science and to empower the students to make them competent for industrial needs, R & D sectors and self- employment the curriculum is framed. Indeed, the curriculum encompasses knowledge of Data Science. Also this course covers the essentials of statistics and how to prepare data before processing in real time application. With successful completion of this course, the students will be competent for handling and processing of big data and reveal good opportunity as a Data Scientist in renowned companies in corporate field. As a Data Scientist the students will be able to empower the management by helping them to take better decisions, and identifying goals based on recent trends.

2. Duration

The M.Sc. program will be a full-time TWO years i.e. 4 semesters. Pattern of examination will be Semester System.

3. EligibilityforAdmission

- ✓ **B.Sc.**(Statistics / Mathematics / Electronics / Physics / Chemistry/IT) from recognized Indian University.
- ✓ **B.Sc.(**Computer Science Optional), **B.C.S., B.Sc**.(Entire Computer Science),
- ✓ **(BC**A)**Bachelor of Computer Application** from recognized Indian University.
- ✓ **Bachelor of Engineering** (Computer Engineering, Electronics Engineering, Electronics and Telecommunication Engineering, Mechanical Engineering and Electrical Engineering) from recognized Indian University.
- ✓ Admission through College Level Entrance exam only.
- ✓ Only entrance marks should be considered for admission process.
- ✓ Reservation of Seats as per rules of Government of Maharashtra.

4. Medium of Instruction

The medium of Instruction will be English only.

5. ProgramStructure

Structure in Accordance with National Education Policy - 2020
With Multiple Entry and Multiple Exit Options
M.Sc.(DataScience)Part–I(Level-6.0)

	CourseCode	TeachingScheme			ExaminationScheme					
		TheoryandPractical			UniversityAssessment(UA)			InternalAssessment(IA)		
		Lectures+ Tutorial/ (Hours/ week)	Practical (Hours/ week)	Credit	Maximum Marks	Minimum Marks	Exam.Hours	Maximum Marks	Minimum Marks	Exam. Hours
Semester-I										
Major Mandatory	MMT-101	4	--	4	60	24	2	40	16	
	MMT-102	4	--	4	60	24	2	40	16	
	MMPR-103	--	8	4	60	24	2	40	16	
	MMT-104	2	--	2	30	12	2	20	8	
Major Elective	MET-105	4	--	4	60	24	2	40	16	
	MET-106									
Research Methodology	RM-107	4	--	4	60	24	2	40	16	
Total				22	330			220		
Semester-II										
Major Mandatory	MMT-201	4	--	4	60	24	2	40	16	
	MMT-202	4	--	4	60	24	2	40	16	
	MMPR-203	--	8	4	60	24	2	40	16	
	MMT-204	2	--	2	30	12	2	20	8	
Major Elective	MET-205	4	--	4	60	24	2	40	16	
	MET-206									
OJT/FP	OJT-207	--	--	4	60	24	2	40	16	
Total				22	330			220		
Total(Sem I+SemII)				44						

<table>
<tr><td>• MMT–MajorMandatoryTheory
• MMPR–MajorMandatoryPractical
• MET–MajorElectiveTheory
• MEPR–MajorElectivePractical
• RM-ResearchMethodology
• OJT/FP-OnJobTraining/FieldProject</td><td>• TotalMarksforM.Sc.-I :1100
• TotalCreditsforM.Sc.-I (SemesterI &II): 44
• Separate passing is mandatory for UniversityandInternal Examinations</td></tr>
<tr><td colspan="2">*EvaluationschemeforOJT/FPshallbedecidedbyconcernedBOS</td></tr>
<tr><td colspan="2">• RequirementforEntryatLevel6.0:CompletionofLevel5.5</td></tr>
<tr><td colspan="2">• RequirementforExitafterLevel6.0: StudentscanexitaftercompletionofLevel6.0withPostGraduateDiplomainDataScience</td></tr>
<tr><td colspan="2">• RequirementforEntryatLevel6.5:He/ShehavecompletedMScPart-I(Level6.0)</td></tr>
</table>

Structure in Accordance with National Education Policy - 2020
With Multiple Entry and Multiple Exit Options
M.Sc.(DataScience)Part–II(Level-6.0)

	CourseCode	TeachingScheme			ExaminationScheme					
		TheoryandPractical			UniversityAssessment(UA)			InternalAssessment(IA)		
		Lectures+ Tutorial/ (Hours/ week)	Practical (Hours/ week)	Credit	Maximum Marks	Minimum Marks	Exam.Hours	Maximum Marks	Minimum Marks	Exam. Hours
Semester-III										
Major Mandatory	MMT-301	4	--	4	60	24	2	40	16	
	MMT-302	4	--	4	60	24	2	40	16	
	MMPR-303	--	8	4	60	24	2	40	16	
	MMT-304	2	--	2	30	12	2	20	8	
Major Elective	MET-305	4	--	4	60	24	2	40	16	
	MET-306									
Research Project	RP - 307	4	--	4	60	24	2	40	16	
Total				22	330			220		
Semester-IV										
Major Mandatory	MMT-401	4	--	4	60	24	2	40	16	
	MMT -402	4	--	4	60	24	2	40	16	
	MMPR -403	--	8	4	60	24	2	40	16	
Major Elective	MET -404	4	--	4	60	24	2	40	16	
	MET - 405									
Research Project	RP - 406	--	--	6	60	24	2	40	16	
Total				22	330			220		
Total(Sem III+SemIV)				44						

<table>
<tr><td rowspan="2">• MMT–MajorMandatory Theory
• MMPR–MajorMandatoryPractical
• MET–MajorElective Theory
• MEPR–MajorElective Practical
• RP- Research Project</td><td>• TotalMarksforM.Sc.-I :1100</td></tr>
<tr><td>• TotalCreditsforM.Sc.-I (SemesterI &II): 44
• Separate passing is mandatory for UniversityandInternal Examinations</td></tr>
<tr><td colspan="2">#Evaluation scheme for Research Projectshall be decided by concerned BOS</td></tr>
<tr><td colspan="2">##Evaluation scheme for Research Projectshall be decided by concerned BOS</td></tr>
<tr><td colspan="2">• Requirement for Exit after Level 6.5:
• Students can exit after completion of Level 6.5 with Master of Data Science</td></tr>
</table>

6. Program Educational Objectives (PEO)

1. Apply knowledge of statistics to analyze large dataset, develop predictive models.
2. Demonstrate ethical practices, strong communication and team work skills in the field of data science.
3. Work productively as data scientist professionals in industrial careers, graduate government organization, or academia by conducting research, effective use of oral and written communication, working competently as a member of a team unit, adhering to ethical standards in the profession.

7. Program Outcomes (PO's)

- Apply conceptual knowledge for complex computing problems and arrive at a solution based on user requirements.
- Provide analytical skills necessary for arriving at solutions to socio economic problems.
- Work in multi-disciplinary teams to develop interpersonal skills and project management techniques.
- Apply mathematical foundations and computer science theory in the modelling and design of software systems.
- Integrate the concepts of networks and security issues in the required fields.

8. Program Specific Outcomes (PSO)

- Build foundational understanding of key technologies and tools driving data science to develop machine learning models for businesses at global level.
- Inculcate research skills, develop novel ideas, and discover insights in diverse domains to develop data-driven solutions for sustainable development.
- Apply quantitative modelling techniques and develop skills to present inferences using tools that are adaptable to evolving technologies, laying the foundation for lifelong learning.
- Be acquainted with the contemporary issues, latest trends in technological development and thereby innovate new ideas and solutions to existing problems.

9. Course Codes

M.Sc.Semester-I		
Course Code	**Major Mandatory**	
MMT-101	Python Programming (4credits)	**MSU0325MML911G1**
MMT-102	Database Management System(4credits)	**MSU0325MML911G2**
MMPR-103	Practical Lab-I(4credits)	**MSU0325MMP911G1**
MMT-104	Introduction to Statistics (2credits)	**MSU0325MML911G3**
RM-107	Research Methodology(4credits)	**MSU0325RML911G**
	Major Elective	
MET-105	Data Mining (4 Credits)	**MSU0325MEL911G1**
MET-106	Fundamentals of Machine Learning (4credits)	**MSU0325MEL911G2**
M.Sc.Semester-II		
	Major Mandatory	
MMT-201	Advanced Python Programming (4 credits)	**MSU0325MML911H1**
MMT-202	Artificial Intelligence(4credits)	**MSU0325MML1H2**
MMPR-203	Practical Lab-II (4credits)	**MSU0325MMP911H1**
MMT-204	Statistical Data Analysis (2 credits)	**MSU0325MML911H3**
OJT-207	Internship(4credits)	**MSU0325OJP911H**
	Major Elective	
MET-205	Basics of ImageProcessing(4credits)	**MSU0325MEL911H1**
MET-206	Big Data(4credits)	**MSU0325MEL911H2**
M.Sc.Semester-III		
	Major Mandatory	
MMT-301	Data Structure Using Python (4 Credits)	**MSU0325MML911I1**
MMT-302	Advanced Machine Learning(4 credits)	**MSU0325MML911I2**
MMPR-303	Practical Lab -III(4credits)	**MSU0325MMP911I1**
MMT-304	Power BI(2 credits)	**MSU0325MML911I3**
RP-307	Research Project (4credits)	**MSU0325RP911I**
	Major Elective	
MET-305	Cyber Security(4credits)	**MSU0325MEL911I1**
MET-306	Cloud Computing(4credits)	**MSU0325MEL911I2**
M.Sc.Semester-IV		
MMT-401	Internet of Things and Data Science (4credits)	**MSU0325MML911J1**
MMT-402	Data Visualization using Tableau (4credits)	**MSU0325MML911J2**
MMPR-403	Practical Lab -IV(4credits)	**MSU0325MMP911J1**
RP-406	Research Project (6 credits)	**MSU0325RP911J**
	Major Elective	
MET-404	Deep Learning (4credits)	**MSU0325MEL911J1**
MET-405	Natural Language Processing (4 credits)	**MSU0325MEL911J2**

10.Syllabus

M.Sc.(DataScience)(PartI)(Level-6.0)(SemesterI)

Title of Course: Python Programming
Course Code: MMT-101
Internal Marks: 40 External Marks: 60 Theory: 04 hours/week

CourseOutcomes: Upon successful completionof this course, thestudent willbeable to:

1. Interpret the fundamental Python syntax and semantics and be fluent in the use of Python control flow statements.
2. Express proficiency in the handling of strings and functions
3. Determine the methods to create and manipulate Python programs by utilizing the data structures like lists, dictionaries, tuples and sets.
4. Identify the commonly used operations involving file systems and regular expressions

Unit I: (15 Hours)

Introduction to Python Programming :Introduction to Python, Features of Python, Applications of Python, Installation and setup, Writing and running Python scripts, Overview of IDEs,Comments and indentation, Identifiers, Keywords, Naming conventions, data types in python , Implicit type conversion, Explicit type conversion, type(), isinstance(), input() function, print() function, print formatting using format(), f-strings, operators in python

Unit II: (15 Hours)

Control Structures and Looping : if statement, if-else statement, if-elif-else statement, Nested conditionals, while loop, , for loop, for loop using range, Nested loops, Looping patterns, break statement, continue statement, pass statement, Use cases of pass, else with loops, Pattern printing basics

Unit III: (15 Hours)

Functions and Functional Programming Concepts: Introduction to functions, Syntax of defining functions, Calling functions, return statement, Importance of modularity, Code reusability, Built-in functions overview, User-defined functions, Required arguments, Default arguments, Keyword arguments, Variable-length arguments: *args, **kwargs, Introduction to lambda functions, Syntax of lambda, Use-cases of lambda, Comparison of lambda and regular functions, Function annotations, Syntax of annotations, Use cases of annotations, Function practice problems.

Unit IV: (15 Hours)

Introduction to modules, Creating modules, Importing modules, import statement, from-import statement, Built-in modules: math, random, Creating custom modules, Using custom modules, Basics of Python packages, Creating a simple package, init**.**py file usage, Importing from packages, Importing from sub-packages, Organizing code using packages, Introduction to file handling, Opening files, Closing files, Reading files: read(), readline(), readlines(), Writing to files: write(), writelines(), File modes: r, w, a, r+, w+,, seek() function, tell() function, Writing logs to file.

Reference Books:

1. Learning Python, O'Reilly publication
2. Programming Python, O'Reilly publication
3. Python Data Science Handbook, Jake Vander Plas O'Reilly publication

M.Sc.(DataScience)(PartI)(Level-6.0)(SemesterI)

Title of Course: Database Management System
Course Code:MMT-102
Internal Marks: 40 External Marks: 60 Theory: 04 hours/week

CourseOutcomes: Upon successful completionof this course, thestudent willbeable to:

1. Describe different concepts of database management systems.
2. Discuss structure of relational databases and apply relational operations on it.
3. Apply the basic and advanced concepts of SQL language to solve the queries in the databases.
4. Analyze database requirements and determine the entities involved in the system and their relationship.
1. Write the queries to implement different functionalities of SQL language.
2. Write queries related to NoSQL.

Unit I: (15 Hours)

Introduction: Database-system Applications, Purpose of Database Systems, View of Data-Data Abstraction, Instance and Schemas Data Models, Data storage and querying: Storage Manager, The query processor Database Architecture.

Unit II: (15 Hours)

E-R model, Relational Model and SQL Entity Relationship Model Generalization, Specialization Aggregation. Introduction to Relational Model Relational Model and Codd Rules Structure of Relational Databases Anomalies in a Databases, The normalization process: (1NF, 2NF, 3NF, BCNF, 4NF, 5NF) Database Schema, Keys. SQL query language SQL- Basic Types, Basic schema definition, Date and Time in SQL. Integrity constraint- Constraints on a single relation, Not Null Constraint, Unique constraint, referential integrity. Basic structure of SQL queries- Queries on single relation. Additional basic operations Set operations. Aggregate Functions-Basic aggregation, Aggregation and grouping, The Having clause, Nested Queries Modification of the Database-Deletion, Insertion, Updates

Unit III: (15 Hours)

Intermediate and Advanced SQL: Join Expressions- Join conditions, Outer joins, Join types. Views- View definition, using views in SQL queries, views, update a view.
PL/SQL-PL /SQL Block Control Structures, Cursor, Trigger, Procedure, Function.
Transaction and Concurrency Control: Acid Properties, Failure, Classification, Recovery and atomicity Log- based recovery, Concurrency Control.

Unit IV: (15 Hours)

Introduction To Nosql: Objectives, Characteristics, NoSQL Storage types, Advantages and Disadvantages of NoSQL, SQL and NoSQL, NoSQL Products, CRUD operations using MongoDB, Introduction to Casandra.

Reference Books:

1. Ramakrishnan, Gehrke, Database Management Systems, McGrawHill International Edition, ThirdEdition
2. Peter Rob, Carlos Coronel, Database System Concepts, Cengage Learning, India Edition
3. S.K.Singh, "Database Systems Concepts, Design and Applications", First Edition, Pearson Education,2006
4. Redmond,E. & Wilson, Seven Databases in Seven Weeks: A Guide to Modern Databases and theNoSQL Movement Edition:1st Edition
5. KristinaChodorow,"MongoDB-TheDefinitiveGuide",SecondEdition,O'Reilly,2013

M.Sc.(DataScience)(PartI)(Level-6.0)(SemesterI)

Title of Course: Practical Lab -I
Course Code: MMPR-103
Internal Marks: 40 **External Marks: 60** **Lab: 08 hours/week**

CourseOutcomes: Upon successful completionof this course, thestudent willbeable to:

1. Create, modify and apply operations on table
2. Handling different keys on database
3. Handling different types of SQL queries on database
4. Handling different types of mathematical and statistical functions
5. Handling and write code using PL/SQL
6. Handling cursor in in PL/SQL.
7. Programs based on Python programming.

Lab assignments based on Course Code **MMT -101** and **MMT-102**.

M.Sc.(DataScience)(PartI)(Level-6.0)(SemesterI)

Title of Course: Introduction to Statistics
Course Code: MMT-104

Internal Marks: 20 **External Marks: 30** **Theory: 02 hours/week**

CourseOutcomes: Upon successful completionof this course, thestudent willbeable to:

1. To accurately compute and interpret the Measures of central tendency and dispersion with 100% correct calculations.
2. To analyze the data and classify them using skewness and kurtosis methods.
3. To define and apply fundamental probability concepts to solve theoretical and practical problems, achieving at least 85% accuracy in both interpretation and calculation.
4. To apply probability concepts to solve real-world problems using discrete and continuous probability distributions with at least 75% accuracy in problem-solving assessments.
5. To define and distinguish between discrete and continuous random variables, compute probability mass and density functions, to real-life problems, achieving at least 85% accuracy in interpretation and problem-solving.

UNIT I: **(15 Hours)**

Measures of Central Tendency, Measures of Dispersion, Skewness, Kurtosis, BoxPlot. Introduction to Probability: Definitions of probability: classical approach, axiomatic approach, probability of an event. Addition Theorem, Concepts and definitions of conditional probability, multiplication theorem $P(A \cap B) = P(A).\ P(B|A)$, Bayes' theorem, Application of Bayes' theorem, Concept and definition of independence of two events.

UNIT-II **(15Hours)**

Introduction to Random Variables: Definition of discrete random and continuous random variable. Concept of Discrete and Continuous probability distributions. (p.m.f. and p.d.f.), Distribution function, Expectation and variance, Numerical problems related to real life situations. Special Distributions: Binomial Distribution, Poisson Distribution, Normal Distribution, Exponential Distribution.

Recommended Book:

1. Fundamentals of Mathematical Statistics (3rd Edition),Gupta and Kapoor, S.Chand and Sons, New Delhi,1987
2. Fundamentals of Applied Statistics (3rd Edition),Gupta and Kapoor, S.Chand and Sons, New Delhi,1987

Reference Books:

1. Statistical Methods, G.W. Snedecor, W.G. Cochran, John Wiley & sons,1989.
2. Probability, Statistics, Design of Experiments and Queuing theory with applications Computer Science, Trivedi K.S., Prentice Hall of India, NewDelhi,2001.
3. Statistics Using R, SudhaPurohit, S.D.Gore, ShailajaDeshmukh, Narosa, Publishing Company

M.Sc.(Data Science)(Part I)(Level-6.0)(SemesterI)

Title of Course: Data Mining
Course Code: MET-105
Internal Marks: 40 **External Marks: 60** **Theory: 04 hours/week**

CourseOutcomes: Upon successful completionof this course, thestudent willbeable to:

1. Ability to understand the types of the data to be mined and present a general classification of tasks and primitives to integrate a data mining system.
2. Apply preprocessing methods for any given raw data.
3. Extract interesting patterns from large amounts of data
4. Discover the role played by data mining in various fields.

Unit I: **(15 Hours)**

Introduction Data Mining, Types of Data Mining- Relational Database, DataWarehouse,Data Repositories, Object-Relational Database, Transactional Database. Advantages of Data Mining, Disadvantages of Data Mining. Data Mining Application- Data Mining in Healthcare, Market Basket Analysis, Education, Manufacturing Engineering, CRM, Fraud Detection, Lie Detection, Financial Banking Challenges of Implementation in Data Mining Data Mining Techniques, Data Mining Implementation Process Data Mining Models, Data Mining Tools.

Unit II: (**15 Hours**)

Classification: Preliminaries: Definition, Types of Classification, Key Concepts, general approach to solve classification problem, Decision tree induction: Concept, Attribute Selection Measures, Advantages, Disadvantages, Rule-based classifier, Bayesian Classifiers: Bayes theorem, types, Support Vector Machine: Hyperplane, support vectors, margins, kernel tricks.

Unit III: (**15 Hours)**

Association analysis: Problem definition, Frequent Itemset Generation, Apriori Principle and algorithm, Maximal Frequent itemset, closet frequent itemset. FP- growth algorithm, Sequential Patterns, Infrequent Patterns.

Unit IV: (**15 Hours**)

Cluster analysis: Introduction, Types of clustering, Types of clusters. K- Means algorithm, Agglomerative Hierarchical clustering, DBSCAN, Prototype based clustering and Density based clustering.

Recommended Book:

1. Data Mining concepts and techniques --- Jiawei Han and MichelineKamber , Elsevier

Reference Books:

1. Introduction to Data Mining – Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Pearson education.
2. Data Mining: Introductory and Advanced Topics - Margaret H. Dunham, Pearson education

M.Sc.(DataScience)(PartI)(Level-6.0)(SemesterI)

Title of Course: Fundamentals of Machine Learning
Course Code: MET-106
Internal Marks: 40 External Marks: 60 Theory: 04 hours/week

CourseOutcomes: Upon successful completionof this course, thestudent willbeable to:

1. Able to generate, analyze and interpret data summaries.
2. Able to carry out analysis on machine learning algorithms.
3. Able to design and implement frequent itemset mining systems.
4. Explain and utilize concepts of machine learning for data science.

UNIT-I (15Hours)

What is Machine Learning, Types of Machine Learning (Supervised, Unsupervised, Reinforcement Learning), Key Terminology (Features, Labels, Training set, Test set, Overfitting, Underfitting), Workflow of Machine Learning, Applications and Use Cases of Machine Learning, Key Tasks of Machine Learning, Steps in Developing a Machine Learning Algorithm, Basics Introduction to Models in Machine Learning

UNIT-II (15Hours)

Introduction to Supervised Learning, k-Nearest Neighbors (k-NN) Concept, Distance Intuition, Classification Process, Introduction to Classification, Logistic Regression (Binary Classification, Sigmoid Function, Decision Boundary), Evaluation Metrics for classification.

UNIT-III (15Hours)

Introduction to Decision Trees , Tree Construction ,Visualizing Decision Trees, Evaluating Decision Trees ,Naïve Bayes Classifier , Introduction to Support Vector Machines (SVM) :Margin, Hyperplane, Support Vectors, Kernel Functions Overview :Polynomial and RBF Kernels, Advantages and Limitations of SVM, Model Tuning Basics Regularization, Hyperparameter Overview.

UNIT-IV (15Hours)

What are Recommender Systems, Use Cases and Types, Content-Based Filtering , Collaborative Filtering , Introduction to Ensemble Methods -Bagging, Random Forest, Overfitting and Underfitting in Models, Bias-Variance Tradeoff , Cross-validation Techniques. Implementation using Python.

References:

1. Max Kuhn , Kjell Johnson, “Feature Engineering and Selection: A Practical Approach for Predictive Models” 1st Edition, Chapman & Hall/CRC Data Science Series, ISBN 13- 978-1-138-07922-9. 0
2. Sinan Ozdemir, Divya Susarla, “Feature Engineering Made Easy”, Packet Publishing, ISBN 978-1-78728- 760-0
3. Alice Zheng & Amanda Casari, “Feature Engineering for Machine Learning: Principles and Techniques

M.Sc.(DataScience)(PartI)(Level-6.0)(SemesterI)

Title of Course: Research Methodology
Course Code: RM-107
Internal Marks: 40 External Marks: 60 Theory: 04 hours/week

CourseOutcomes: Upon successful completionof this course, thestudent willbeable to:

1. Identify appropriate topics for research work.
2. Carry out Literature Survey
3. Select and define appropriate research problem and parameters
4. Design the use of major experimental methods for research
5. Use appropriate tools, techniques, and processes of doing research in Data science
6. Become aware of the ethics in research, academic integrity and plagiarism
7. Write a research report and thesis

UNIT-I (15Hours)

Introduction to Research Methodology: Research: Definition, Characteristics, Objectives, Research, Role of research in data science lifecycle and

Scientific method Types of Research: Descriptive vs. Analytical Research, Applied vs. Fundamental Research, Quantitative vs. Qualitative Research, Conceptual vs. Empirical Research

Research Problem: Research Problem, Selecting the problem, Necessity of Defining the problem, Technique Involved in Defining a Problem.

UNIT-II (15Hours)

Research Design, and Data Collection Research Design: Meaning, Need, Features of Good Design, Concepts Relating to Research Design, Different Research Design, Basic Principle of Experimental Designs.

Data Collection: Observation Method, Interview Method, Questionnaires, Case Study Method.

Survey Methods: Definition and Purpose of a Survey, Types of Surveys: Cross- sectional vs. Longitudinal Surveys, Online, Telephonic, Face-to-Face, Email Surveys, Steps in Survey Research.

UNIT-III (15Hours)

Sampling Design and Scaling Techniques Sampling Design: Sample Survey, Implication of a Sample Design Steps in Sampling Design, Criteria of Selecting a Sampling Procedure, Characteristics of a Good Sample Designs, Different Types of Sample Designs.

Scaling Techniques: Scaling, Scale Classification Bases, Important Scaling Techniques, Scale Construction Techniques.

UNIT-IV (15Hours)

Hypotheses and Report Writing Hypothese: What is Hypotheses?, Basic Concepts Concerning Testing of Hypotheses, Procedure for Hypothesis Testing, Test of Hypotheses, Important Parametric Tests.

Report Writing: Significance of Report Writing, Different Steps in Writing Report, Layouts of Research Report, Types of Reports, Oral Presentation, Mechanics and Precautions for Writing Research Report.

Recommended Books:

1. Kothari C.R., "Research Methodology. New Age International, 2004, 2nd Ed; ISBN:13: 978-81- 224- 1522-3.

References Books:

1. David V Thiel, "Research Methods- for Engineers", Cambridge University Press, ISBN:978-1- 107- 61019-4
2. Caroline Whitbeck, "Ethics in Engineering Practice and Research", 2nd Ed., Cambridge University Press; ISBN :978-1-107-66847-8

3. Gordana DODIG-CRNKOVIC, “Scientific Methods in Computer Science”, Department of Computer Science Malardalen University, Vasteas, Sweden;
ISBN: 91-26-97860-1

M.Sc.(DataScience) (PartI) (Level-6.0) (SemesterII)

Title of Course: Advanced Python Programming
Course Code: MMT-201
Internal Marks: 40 External Marks: 60 Theory: 04 hours/week

CourseOutcomes: Upon successful completionof this course, thestudent willbeable to:

1. Demonstrate the ability to implement robust Python programs using exception handling constructs such as try, catch and except, and custom exceptions.
2. Apply core Object-Oriented Programming concepts like encapsulation, inheritance, polymorphism, and abstraction to design modular and reusable Python code.
3. Utilize built-in and standard Python libraries such as math , date-time, OS , sys and random to simplify problem-solving and enhance program functionality.
4. Analyze and integrate advanced modules from the Python Standard Library, including to write efficient and optimized Python applications.

Unit I: (15 Hours)

Exception Handling and Object-Oriented Programming:
Introduction to errors and exceptions, Types of exceptions in Python, Syntax of try and except block, try-except-else structure, finally block and its use, Catching multiple exceptions, Catching specific exceptions, Raising exceptions manually, Creating custom exceptions, Introduction to OOP, Defining classes and objects, Class variables and methods, Access modifiers, Inheritance in Python, Method overriding, Polymorphism with functions and classes, Abstract classes, Encapsulation and its importance.

Unit II: (15 Hours)

NumPy and Data Manipulation with Pandas:Introduction to NumPy, Creating NumPy arrays, Data types in NumPy, Array indexing and slicing, Array shape and reshaping, Aggregation functionmBroadcasting in NumPy, Computation on NumPy arrays ,Sorting and filtering NumPy arrays, Introduction to Pandas, Series creation and operations, DataFrame creation, Indexing and re-indexing in Pandas, Basic DataFrame operations, Handling missing data, Filtering data using conditions, Data transformation, Pivot tables and reshaping, Data aggregation in Pandas, Summary and mini-project using NumPy and Pandas.

Unit III: (15 Hours)

Data Visualization with Matplotlib and Seaborn: Data Visualization with Matplotlib and Seaborn Introduction to data visualization, Importance and types of plots, Introduction to Matplotlib, Structure of a Matplotlib plot , Creating line plots, Customizing line styles and markers, Creating scatter plots, Bar plots and pie charts, Creating histograms, Density plots and binning, Subplots and figure layout, Plotting multiple plots in one figure, Adding legends, grid, text annotations, Saving plots to files.

Unit IV: (15 Hours)

Introduction to Scikit-learn and Seaborn: Introduction to machine learning concepts, Overview of Scikit-learn library, Setting up and installing Scikit-learn, Loading datasets using Scikit-learn, Data preprocessing using Scikit-learn: train_test_split, Introduction to Seaborn, Seaborn vs Matplotlib, Basic plotting functions,Customizing themes and styles in Seaborn.

References

1. Fluent Python, 2nd Edition by Luciano Ramalho
2. "Learning Python" by Mark Lutz (O'Reilly)
3. Ultimate Python Libraries for Data Analysis and Visualization by Abhinaba Banerjee
4. The Python Library Reference: Release 3.6.4 - Book 2 of 2 by Guido Van Rossum and Python Development Team

M.Sc.(DataScience)(PartI)(Level-6.0)(SemesterII)

Title of Course: Artificial Intelligence
Course Code: MMT-202
Internal Marks: 40 External Marks: 60 Theory: 04 hours/week

CourseOutcomes: Upon successful completionof this course, thestudent willbeable to:

1. Apply problem solving by intelligent search approach.
2. Represent knowledge using AI knowledge representation techniques.
3. Design Machine Learning solution to real life problems.
4. Derive solutions for problems with uncertainty using Fuzzy theory.
5. Define a NLP problem and find a suitable solution to it.
6. To develop a good understanding of all aspects of Natural Language Processing (NLP) and Genetic algorithm.

UNIT I (15 Hours)

Introduction to Artificial Intelligence and Problem Solving: Introduction of Artificial Intelligence. Different Forms of AI. Applications of AI. Artificial Intelligence in Data Science Role of Artificial Intelligence in Data Science. Comparison of AI and Data Science. Heuristic approach in problem solving. AI Problems and Expert problems. Search Algorithm in AI, Properties of Search Algorithm, Types of Search Algorithm: BFS, DFS, DLS, Bi Directional Search, Greedy Search, A* Search, AO* Search.

UNIT-II (15Hours)

Artificial Neural Networks: concept and ANN architectures, Training and implementation of neural network, Model of an Artificial Neuron: Deterministic, Stochastic. Network Architectures: Single-Layer Feed forward Networks, Multi-Layer Feed forward Networks, Recurrent Networks.
Introduction to **Fuzzy Set**: Fuzzy System and ANN: Predicate Logic, Fact- Table, Predicate Calculus, WFF, Fuzzy logic, Design of fuzzy rulebase. Fuzzy Set Theory, Fuzzy Operations.

UNIT III (15Hours)

Knowledge Representation: Procedural Vs Declarative Knowledge, Representations & Approaches to Knowledge Representation, Forward Vs Backward Reasoning, Matching Techniques, Partial Matching, Fuzzy Matching Algorithms and RETE Matching Algorithms; Logic Based Programming-AI Programming languages: Overview of LISP, Search Strategies in LISP, Pattern matching in LISP, An Expert system Shell in LISP, Over view of Prolog, Production System using Prolog

Genetic Algorithm: History and evolution of G.A, Modeling a problem for the application of G.A.- Representation of data in chromosomes, Fitness Function, Comparison of ANN and GA, Application of G.A.

UNIT IV (15Hours)

Natural Language Processing: Natural Language Processing: Text categorization, text summarization, vision and perception, pattern matching.

Experts Systems: Overview of an Expert System, Structure of an Expert Systems, Different Types of Expert Systems-Rule Based, Model Based, Case Based and Hybrid Expert Systems, Knowledge Acquisition and Validation Techniques, Black Board Architecture, Knowledge Building System Tools, Expert System Shells, Fuzzy Expert systems.

References

1. Neural networks, fuzzy logic and genetic algorithms, synthesis and applications by S. Rajsekaran, G.A. VijayalaxmiPai (EEE)
2. Genetic algorithms by David Goldberg (Addison and Wesley)
3. Principles of AI and Expert system development by David Rolston (MGH)
4. Artificial Intelligence by E. Ritch and K. Knight (MGH)
5. Artificial Intelligence: A New Synthesis Nilsson Elsevier Publication 3 Artificial Intelligence with Python. Prateek Joshi Packt Publishing Ltd
6. Artificial Intelligence a Modern Approach Russel and Norvig Pearson Education, 2nd
7. Artificial Intelligence– A Practical Approach Patterson Tata McGraw Hill, 3rd

M.Sc.(DataScience)(PartI)(Level-6.0)(SemesterII)

Title of Course:Practical Lab-II
Course Code: MMPR-203

Internal Marks: 40 **External Marks: 60** **Lab: 08 hours/week**

CourseOutcomes: Upon successful completionof this course, thestudent willbeable to:

1. Implement and manage exception handling techniques to python programs using constructs such as try, except, finally and user-defined exceptions.
2. Design and develop real-world applications using advanced object-oriented programming features such as inheritance, polymorphism, encapsulation, and abstraction in Python.
3. Perform numerical operations, array manipulations, and linear algebra tasks using NumPy for efficient scientific computing.
4. Visualize data effectively using libraries like Matplotlib and Seaborn to create various types of plots such as line plots, bar charts, histograms, scatter plots, and heatmaps for data analysis.
5. Implement programs on search algorithms using AI.

Practical's will be based on MMT-201 and MMT-202

M.Sc.(DataScience)(PartI)(Level-6.0)(SemesterII)

Title of Course: Statistical Data Analysis
Course Code: MMT-204

Internal Marks: 20 **External Marks: 30** **Theory: 02 hours/week**

CourseOutcomes: Upon successful completionof this course, thestudent willbeable to:

1. Identify sampling methods from the pattern of the observed data to Predict the future behavior of the time series data.
2. Analyze sample data and identify the parameters and their probability distributions.
3. To formulate hypotheses and perform hypothesis testing by applying suitable statistical tests under given problem statements.
4. Hypothesize and test an assumption regarding population parameters using Sample data.
5. To perform different Hypothesis Testing for statistical analysis of data.

UNIT-I **(15Hours)**

Introduction to Sampling, Advantages and Disadvantages, types of sampling techniques: probability sampling technique, Non-probability sampling technique, concept of sampling error. Introduction to Sampling distributions, Student's t distribution, Chi-square distribution, Snedecor's F distribution, Interrelations among t, chi-square and F distributions, Central Limit Theorem (Various Versions) and its applications.

UNIT-II **(15Hours)**

Hypothesis Testing: Introduction, Concept of null hypothesis and alternative hypothesis, critical region, level of significance, Power of the test, p-value, Types of errors, one sided and two- sided tests. Large sample tests: Single mean, Difference of two means, Single proportion, Difference of two proportions. Small sample tests: t-test for single mean, difference between two means, F-test for equality of two population variances, Chi-square test: for single mean, for goodness of fit. ANOVA.

Recommended Book:

1. Fundamentals of Applied Statistics (3rd Edition), Gupta and Kapoor, S.Chand and Sons, New Delhi, 1987.
2. Time Series Methods, Brockell and Devis, Springer, 2006.

References Books:

1. Probability, Statistics, Design of Experiments and Queuing theory with applications Computer Science, Trivedi K.S., Prentice Hall of India, New Delhi,2001.
2. Common Statistical Tests, Kulkarni M.B., Ghatpande S.B., Gore S.D., Satyajeet Prakashan,Pune, 1999.
3. Probability and Statistical Inference, 9th Edition, Robert Hogg, Elliot Tanis, Dale Zimmerman, Pearson education Ltd, 2015
4. A Beginners Guide to R, Alain Zuur, Elena Leno, Erik Meesters, Springer, 2009
5. Statistics Using R, Sudha Purohit, S.D.Gore, Shailaja Deshmukh, Narosa, Publishing Company

M.Sc.(Data Science)(PartI)(Level-6.0)(SemesterII)

Title of Course: Basics of Image Processing
Course Code: MET-205

Internal Marks: 40 **External Marks: 60** **Theory: 04 hours/week**

CourseOutcomes: Upon successful completionof this course, thestudent willbeable to:

1. Understand the basic principles and concepts of digital image processing.
2. Gain knowledge of different image representations and color models.
3. Learn how to pre-process and enhance images using various techniques.
4. Exploreimage filteringtechniquesfornoise reductionand feature enhancement.
5. Understand the concept of image segmentation and different segmentation algorithms.
6. Analyzing general terminology of digital image processing.

UNIT-I **(15Hours)**

Introduction, Digital Image Fundamentals: elements of visual perception, light and electromagnetic spectrum, image sensing and acquisition, image sampling and quantization, some basic relationship between pixels. Representationofdigitalimageinspatialdomainaswellas in matrix form. Block diagram of fundamentals steps in digital image processing, application of digital image processing system, Elements of Digital Image, Processing systems, structure of the Human, Image Formation in the Eye, Brightness Adaptation and Discrimination.

UNIT-II **(15Hours)**

Introduction to image processing: basic concepts and applications, Image acquisition and representation, Imagefile formats and colormodels, Image enhancement: Point operations, Contrast stretching, intensity level slicing, log transformation, power log transformation, bit plane slicing, Unnormalized and Normalized Histogram, Histogram Equalization and spatial domain techniques, Noise reduction: spatial andfrequency domain filtering, Image restoration: degradation model, inverse filtering, and Wiener filtering, Image sharpening techniques, Image segmentation: thresholding, region- based segmentation, and edge detection, Contour detection and boundary extraction,

UNIT-III **(15Hours)**

Image Restoration: Noise models, band reject, band pass and notch filter Image Compression: Fundamentals, Models, Error free and lossy compression, Standards.

UNIT-IV **(15Hours)**

Morphological Image Processing: Boundary extraction, Region filtering, connected component extraction, Thinning, thickening, pruning, image segmentation.

References

1. Digital ImageProcessingbyRafaelC.Gonzalez
2. Principles of Digital Image Processing Core Algorithms by Wilhelm Burgerand Mark J.Burge
3. Fundamentals of Digital Image Processing by Annadurai
4. FundamentalsofDigitalImage ProcessingbyJain A K

M.Sc.(DataScience)(PartI)(Level-6.0)(SemesterII)

Title of Course: Big Data
Course Code: MET-206
Total Credits: 04
Internal Marks: 40 External Marks: 60 Theory: 04 hours/week

Course Outcomes: Upon successful completion of this course, the student will be able to:

1. Identify Big Data and its Business Implications
2. Develop Big Data Solutions using Yarn
3. Develop Big Data Solutions using Hive
4. Develop Big Data Solutions using Pig

Unit I: (15 Hours)

Big Data Introduction: What is Big Data?, Evolution of Big Data, Benefits of Big Data, characteristics of Bog Data, Types of Big Data, Sources of Big Data, Architecture of Big Data, Operational vs Analytical Big Data, Need for Big Data Analytics, Big Data Challenges.

Unit II: (15 Hours)

Yarn: Introduction to Yarn, Importance of Yarn, Advantages of Yarn, Yarn Architecture, Resource Manager, Node Manager, Application Master, Application submission in Yarn, Yarn Applications.

Unit III: (15 Hours)

Hive: Introduction to Hive, Features of Hive, Architecture of Hive, Components of Hive, Hive Data Models, Hive DDL statements, Hive Query Language.

Unit IV: (15 Hours)

Pig: Introduction to Pig, Running Pig, Getting Started with Pig Latin, Working with operators in Pig, Working with functions in Pig.

Recommended Books:

1. BIG DATA Covers Hadoop 2, MapReduce, Hive, YARN, Pig, R and Data Visualization
2. Pete Warden, “Big Data Glossary”, O’Reily, 2011

References Books:

1. Big Data: Concepts, Technology and Architecture- NandiniAbirami R, BalamaruganBalusamy, Seifedine Kadry and Amir Gandomi.

M.Sc.(Data Science)(Part I)(Level-6.0)(SemesterII)

TitleofCourse:Internship
Course Code: OJT-207

Internal Marks: 40 **External Marks: 60** **Theory: 04 hours/week**

CourseOutcomes: Upon successful completionof this course, thestudent willbeable to:

1. Gain industrial experience.
2. Gain Hands on Skill and build professional network.
3. Learn to work in team and how to communicate.
4. Reflect and evaluate on experiences about area of interest and abilities.
5. Demonstrate and apply research skills to complete a project
6. Apply the theoretical concepts to solve industrial problems with teamwork and multidisciplinary approach.

Student is supposed to carry out on job training during his/her semester vacation.

Scheme of Teaching

1. Each contact session for teaching or practical should be of 60 minutes each.
2. 60 lectures should be conducted for each course of 4 credits and 30 lectures for 2 credits.
3. One Practical Batch should be of 30 students.
4. Practical evaluation should be conducted before the commencement of University examination.

11. Examination Pattern

Theory:

- **For 4 credit course-** University examinations: 60 marks, Internal evaluation: 40 marks
    - Two tests should be conducted of MCQ type questions. Each test will be of 10 marks
    - Assignments/Seminar/Activities carries 20 marks
- **For 2 credit course-** University examinations: 30 marks, Internal evaluation: 20 marks
    - One test should be conducted of MCQ type questions of 10 marks.
    - Assignments/ Seminar carries 10 marks
- The internal marks will be communicated to the University at the end of each semester,but before the semester end examinations. These marks will be considered for thedeclarationofthe results.

Practical:

Practical evaluation will be through university appointed panels of external and internal examiners.

Project:

- **For 4 credit course-** University examinations: 60 marks, Internal evaluation: 40 marks
    - Project viva by university appointed external and internal examiners.
    - Internal evaluation will be carried out by internal guide.

12. Nature of Question Paper and Scheme of Marking

Theory:

1) There will be Seven (7) questions of 12 Marks and out of which Four (4) to be attempted from question no 2 to 7.
2) Question No.1 is compulsory and is of multiple choice questions. There will be 6 multiple choice question each carries 2 marks
3) Question No.2 to Question No. 6 should consist 2 sub question each carries 6 marks
4) Question No. 7 should be short note types consisting of 4 sub questions out of which 2 needs to be attempted each carries 6 marks.

Practical:

1) Duration of Practical Examination: 2.30 Hrs

2) Nature of Question paper: There will be three questions out of which any two questions to be attempted and each question carries 30 Marks.

3) Thefinalpracticalexaminationwillbeconductedbytheuniversityappointedexaminers both internal as well as external at the end of semester for each lab courseandmarkswillbesubmittedtotheuniversity by thepanel.ThepatternoffinalPractical Examination willbeas follows;

1	CodingandExecutionof Program	60Marks
2	Viva-voce	20Marks
3	Journal	20Marks
4	Total	100Marks

The practical examination will be conducted semester wise in order to maintain therelevanceofthe respectivetheorycoursewith laboratorycourse.

On Job Training:

Student has to make a presentation of the work carried out during On Job Training in front of a panel external and internal examiners. He/she has to submit the report of work carried out as part of On Job Training.

Standard of Passing:

Internal as well as external examination will be held at the end of semester. The candidate must score 40% marks in each head of internal as well as external Examination.

Backlog

Student should not have more than 6 backlogs for 2nd year admission.

13.Equivalence of courses

M.Sc.PartI(Semester I and II)

OldCourse				Equivalent Course		
Sem No.	**Course Code**	**TitleofOldCourse**	**Credit**	**CourseCode**	**TitleofNew Course**	**Credit**
I	**CC 101**	**BasicStatisticsfor Data Science**	**4**	**MMT–101**	**Introductionto Statistics**	**4**
I	**CC102**	**Feature Engineering-I**	**4**	**---**	***Noequivalence**	**--**
I	**CC103**	**DataStructure using Python**	**4**	**MMT- 104**	**PythonProgramming**	**2**
I	**CC104**	**Introductionto Database Management System**	**4**	**MMT–102**	**Database ManagementSystem**	**4**
II	**CC201**	**AdvancedStatistics for Data Science**	**4**	**MMT–201**	**Statistics–II**	**4**
II	**CC202**	**Feature Engineering-II**	**4**	**---**	***Noequivalence**	**--**
II	**CC203**	**Machine Learning**	**4**	**MMT–301**	**MachineLearning**	**4**
II	**CC204**	**Designand Analysisof Algorithm**	**4**	**---**	***Noequivalence**	**--**

***Two more chances are given to the student.**