SYLLABUS
M. Sc. (Computer Science with Specialization in Data Science)
Outlines of Text Syllabi and Courses of Reading.

For Examination of Session 2023-24 and 2024-25

1st Semester

Paper Code	Title of Paper
MDSM1101T	Probability and Statistical Techniques
MDSM1102T	Introduction to Data Science and AI
MDSM1103T	Programming in Python
MDSM1104T	Database Management Systems
MDSM1105P	Software Lab – I)Based on MDSM1103T)
MDSM1106P	Software Lab – II (Based on MDSM1104T using SQL)
	Total

2nd Semester

Paper Code	Title of Paper
MDSM1201T	Mathematics for Machine Learning
MDSM1202T	Advanced Statistical Techniques
MDSM1203T	Data Analytics and Visualization
MDSM1204T	Web Development using Python and Jango
MDSM1205P	Software Lab – III (Based on MDSM1203T)
MDSM1206P	Software Lab – IV (Based on MDSM1204T)
	Total

3rd Semester

Paper Code	Title of Paper
MDSM2101T	Soft Skills and Technical Communication
MDSM2102T	Software Engineering for Data Scientists
MDSM2103T	Business Intelligence
MDSM2104E	Elective – I*
MDSM2105P	Software Lab – V (Based on MDSM2103T)
MDSM2106P	Software Lab – VI (Based on MDSM2104E)
	Total

4th Semester

Paper Code	Title of Paper
MDSM2201T	MOOC Courses (Elective 1)
MDSM2202T	MOOC Courses (Elective 2)
MDSM2203P	Dissertation
	Total

List of MOOC Courses (Electives)

Sl. No.	Electives	MOOC Course
1	MOOC courses for Elective 1 and 2	Artificial Intelligence: Knowledge, Representation and Reasoning
2		Blockchain and its Application

3		Business Intelligence and Analytics
4		Computer Vision and Image Processing- Fundamentals and Applications
5		Natural Language Processing

Core Subject Syllabus

MDSM1101T: Probability and Statistical Techniques

Course Objectives: The objective of this course is to provide an understanding for student on statistical concepts to include measurements of location and dispersion, probability, random variables and expectation, probability. By completing this course the student will learn to perform the following:

- How to calculate and apply measures of location and measures of dispersion -- grouped and ungrouped data cases.
- How to apply discrete and continuous probability
- Learn various probability distributions

Course contents

SECTION A

Descriptive Statistics: Variables, Raw data, Graphical plots and charts, Frequency distribution, Histogram and Frequency polygons, Relative frequency distributions, Cumulative frequency distributions, Frequency curves and their types.

Measures of Central Tendency: Mean, Median, Mode.

Measures of Dispersion: Range, Standard Deviation, Quartile deviation, Mean and Median absolute deviation, Moments, Measures of Skewness and Kurtosis.

Introduction of Probability: Algebra of events, Concept of probability, Axioms of probability, Conditional probability, Independent and compound events, Bayes' Rule (without proof) and applications.

SECTION B

Random Variables and Expectations: Introduction of random variables and types, Independent random variables, Bivariate random variables, Joint probability distribution functions: Marginal and Conditional Distributions, distribution function technique, transformation technique: one variable and several variables, Expectation, Variance, and Co-variance of random variables, Conditional expectation and conditional variance.

Generating Functions and Probability Distributions: Discrete and Continuous random variables, Probability distributions functions for discrete and continuous random variables, Probability mass and density function, Probability generating function (p. g. f.), Moment generating function (m.g.f.), Characteristic function (c. f.)-Properties and Applications, Binomial, Poisson and Normal Distributions-Properties, Relationship between Binomial and

Normal, Poisson and Normal distributions, Uniform, Exponential, Gamma and Beta Distributions, Chi-square, student's t, and F Distributions, Central Limit Theorem.

Reference Books:

1. Gupta, S.C. and Kapoor, V.K. (1997) Fundamentals of Mathematical Statistics. Sultan Chand and Sons, New Delhi, 11.23-12.23.
2. Spiegel, M. R., Schiller, J. J., and Alu Srinivasan, R. (2013). Probability and Statistics, Fourth Edition, Schaum's Outline Series, McGraw Hill Companies, Inc
3. Rohatgi, V.K. and Ebsanes Saleh, A.K.Md., An introduction to Probability and Statistics, 2nd Ed., John Wiley & Sons,2002.
4. Shanmugam, R., Chattamvelli, R. Statistics for scientists and engineers, John Wiley, 2015.

Pedagogy:

The Instructor is expected to use leading pedagogical approaches in the class room situation, research-based methodology, innovative instructional methods, extensive use of technology in the class room, online modules of MOOCS, and comprehensive assessment practices to strengthen teaching efforts and improve student learning outcomes.

The Instructor of class will engage in a combination of academic reading, analyzing case studies, preparing the weekly assigned readings and exercises, encouraging in class discussions, and live project based learning.

Case/Class Discussion Assignments:

Students will work in groups of up to four to prepare a brief write-up due before the start of each class covering the case study or class material to be discussed in the next session. Questions may include a quantitative analysis of the problem facing the decision-maker in the case.

Class Participation:

Attendance will be taken at each class. Class participation is scored for each student for each class.

Scheme of Examination

- English will be the medium of instruction and examination.
- Written Examinations will be conducted at the end of each Semester as per the Academic Calendar notified in advance
- Each course will carry 100 marks of which 30 marks shall be reserved for the continuous internal assessment and the remaining 70 marks for written examination to be held at the end of each semester.
- The duration of written examination for each paper shall be three hours.
- The minimum marks for passing the examination for each semester shall be 40% in aggregate as well as a minimum of 40% marks in the semester-end examination in each paper.
- A minimum of 75% of classroom attendance is required in each subject

Instructions for the Paper Setter

The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 12 marks for each question. Section C will have 11 short-answer type questions carrying total of 22 marks, which will cover the entire syllabus uniformly.

Instructions for the Candidates

1. Candidates are required to attempt two questions each from the sections A & B of the question paper and the entire section C.
2. Use of non-programmable scientific calculator is allowed.

MDSM1102T: Introduction to Data Science and AI

Course Objectives: The basic objective of the course is to introduce the students with various concepts of Artificial Intelligence and Data Science. By the end of this course, a student would be expected to be able to:

- Understand concept of Artificial Intelligence and Intelligent agents
- Understand reasoning and learning in knowledge-base systems
- Outline the roles and requirements of artificial intelligence in various applications of AI in general and specific.
- Discuss and illustrate the problems in terms of data exploration and visualization
- Understand the basics of generative modelling and predictive modelling

Course contents

SECTION A

Data Science: Definition and scope of data science, Role of data scientists in industry and research, Applications of data science in various domains, Challenges and opportunities of big data.

Sources of data: structured, unstructured, and semi-structured, Data collection methods: surveys, experiments, web scraping, Data quality assessment and Data Pre-processing, EDA using statistics and visualization. Data-Driven Decision Making.

Introduction to Data Mining, Data Warehousing, Data Analysis, Role of Probability and Statistics in Data Science. Data Science and Business Intelligence

SECTION B

Artificial Intelligence: Definition and history of AI, AI applications and impact on society, Major subfields of AI, Ethical considerations in AI.

Problem Solving and Search: Problem-solving agents, State space search, Heuristic search techniques, Adversarial search, Knowledge Representation and Reasoning: Predicate logic, Propositional logic, Inference rules and mechanisms, Probabilistic knowledge representation and inference, Markov decision problems and reinforcement learning.

Approaches of AI: Turing Test, Cognitive Modelling, Laws of Thought, Rational agent, Machine Learning, Neural Networks and Fuzzy logic.

Applications of AI: Natural Language Processing, Gaming, Speech Recognition, computer Vision Systems, Healthcare, Automotive, Machine Learning, Natural language processing and Robotics.

References:

1. Cathy O'Neil, Rachel Schutt, Doing Data Science, Straight Talk from The Frontline. O'Reilly, 2013.
2. Introducing Data Science, Davy Cielen, Arno D. B. Meysman, Mohamed Ali, Manning Publications Co., 1st edition, 2016
3. Data Science from Scratch: First Principles with Python, Joel Grus, O'Reilly, 1st edition, 2015.
4. Doing Data Science, Straight Talk from the Frontline, Cathy O'Neil, Rachel Schutt, O' Reilly, 1st edition, 2013.
5. Artificial Intelligence: Concepts and Applications by Lavika Goel
6. Artificial Intelligence: A Modern Approach by Russell/Norvig

Pedagogy:

The Instructor is expected to use leading pedagogical approaches in the class room situation, research-based methodology, innovative instructional methods, extensive use of technology in the class room, online modules of MOOCS, and comprehensive assessment practices to strengthen teaching efforts and improve student learning outcomes.

The Instructor of class will engage in a combination of academic reading, analyzing case studies, preparing the weekly assigned readings and exercises, encouraging in class discussions, and live project based learning.

Case/Class Discussion Assignments:

Students will work in groups of up to four to prepare a brief write-up due before the start of each class covering the case study or class material to be discussed in the next session. Questions may include a quantitative analysis of the problem facing the decision-maker in the case.

Class Participation:

Attendance will be taken at each class. Class participation is scored for each student for each class.

Scheme of Examination

- English will be the medium of instruction and examination.
- Written Examinations will be conducted at the end of each Semester as per the Academic Calendar notified in advance
- Each course will carry 100 marks of which 30 marks shall be reserved for the continuous internal assessment and the remaining 70 marks for written examination to be held at the end of each semester.
- The duration of written examination for each paper shall be three hours.
- The minimum marks for passing the examination for each semester shall be 40% in aggregate as well as a minimum of 40% marks in the semester-end examination in each paper.
- A minimum of 75% of classroom attendance is required in each subject

Instructions for the Paper Setter

The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 12 marks for each question. Section C will have 11 short-answer type questions carrying total of 22 marks, which will cover the entire syllabus uniformly.

Instructions for the Candidates

1. Candidates are required to attempt two questions each from the sections A & B of the question paper and the entire section C.
2. Use of non-programmable scientific calculator is allowed.

MDSM1103T: Programming in Python

Course Objectives: The main goal is to formulate a solid foundation in programming using the Python programming language, whether the student is a beginner or has some prior programming experience, this course is designed to equip students with the essential knowledge and skills to become a proficient Python developer. Acquire problem-solving skills: Programming is all about solving problems, and this course aims to sharpen students' problem-solving abilities. You will learn how to approach programming challenges, break them down into smaller, manageable tasks, and devise efficient solutions using Python.

Course Contents

SECTION A

Introduction to Python: Technical Strength of Python, Introduction to Python Interpreter and program execution, Using Comments, Literals, Constants, Python's Built-in Data types, Numbers (Integers, Floats, Complex Numbers, Real, Sets), Strings (Slicing, Indexing, Concatenation, other operations on Strings), Input and Output console operations.

Operators, Expressions and Python Statements: Assignment statement, expressions, Arithmetic, Relational, Logical, Bitwise operators and their precedence, Conditional statements; Notion of iterative computation and control flow–range function, Loops, Statements including break, continue, pass and assert.

Sequence Data Types: Lists, tuples, and dictionary, (Slicing, Indexing, Concatenation, other operations on Sequence datatype), the concept of mutability, Examples to include finding the maximum, minimum, mean; linear search on list/tuple of numbers, and counting the frequency of elements in a list using a dictionary, Using list as Stack and Queues, List comprehensions.

Basic and Advance Functions: Top-down approach of problem solving, Modular programming and functions, Function parameters, Local variables, the Return statement, Doc Strings, global statement, Default argument values, keyword arguments, VarArgs parameters. Library function-input(), eval(),print(), String Functions, String: Slicing, Membership, Pattern Matching, Numeric Functions, Date & Time Functions, Recursion, Lambda(),map(), Reduce(), Filter(), zip().

File Processing: Concept of Files, File opening in various modes and closing of a file, Reading from a file, Writing onto a file, File functions-open (), close (), read (), readline(), readlines(), write(), writelines(),tell(),seek(), Command Line arguments.

SECTION B

Scope and Modules: Scope of objects and Names, LEGB Rule Module Basics, Module Files as Namespaces, Import Model, Reloading Modules.

NumPy Basics and Advance: Introduction to NumPy ndarray, datatypes, array attributes, array creation routines, Array from existing data, Array from Numerical Ranges, Indexing & Slicing, Linear algebra on n-dimensional arrays.

Data Visualization: Introduction to Matplotlib, Scatter plot, Line plot, Bar chart, Histogram, Box plot

Image processing using Python: Introduction to digital image processing, Basic operations on an image including Crop, Scale, Rotate, Flip, Changing contrast, brightness and colour, Edge detection, blur, sharpening.

Object-Oriented Programming: Classes and Objects, Creating Classes in Python, Creating Objects in Python, The Constructor Method, Classes with Multiple Objects, Class Attributes versus Data Attributes, Encapsulation, Inheritance, The Polymorphism.

References:

1. Fluent Python by Luciano Ramalho
2. Head-First Python by Paul Barry
3. Think Python by Allen B. Downey
4. Python Cookbook by David Beazley and Brian K. Jones

Pedagogy:

The Instructor is expected to use leading pedagogical approaches in the class room situation, research-based methodology, innovative instructional methods, extensive use of technology in the class room, online modules of MOOCS, and comprehensive assessment practices to strengthen teaching efforts and improve student learning outcomes.

The Instructor of class will engage in a combination of academic reading, analyzing case studies, preparing the weekly assigned readings and exercises, encouraging in class discussions, and live project based learning.

Case/Class Discussion Assignments:

Students will work in groups of up to four to prepare a brief write-up due before the start of each class covering the case study or class material to be discussed in the next session. Questions may include a quantitative analysis of the problem facing the decision-maker in the case.

Class Participation:

Attendance will be taken at each class. Class participation is scored for each student for each class.

Scheme of Examination

- English will be the medium of instruction and examination.
- Written Examinations will be conducted at the end of each Semester as per the Academic Calendar notified in advance
- Each course will carry 100 marks of which 30 marks shall be reserved for the continuous internal assessment and the remaining 70 marks for written examination to be held at the end of each semester.
- The duration of written examination for each paper shall be three hours.
- The minimum marks for passing the examination for each semester shall be 40% in aggregate as well as a minimum of 40% marks in the semester-end examination in each paper.
- A minimum of 75% of classroom attendance is required in each subject

Instructions for the Paper Setter

The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 12 marks for each question. Section C will have 11 short-answer type questions carrying total of 22 marks, which will cover the entire syllabus uniformly.

Instructions for the Candidates

1. Candidates are required to attempt two questions each from the sections A & B of the question paper and the entire section C.
2. Use of non-programmable scientific calculator is allowed.

MDSM1104T: Database Management Systems

Course Objectives: This course main goal is to give students the knowledge necessary to create, implement, and use database management systems. Despite being a key component of all data science programmes, database technology has taken on a special significance as a result of the highly competitive climate due to the rapid growth of company-to-business and client-to-business internet services, and the requirement to store more business data. This course aims to give students a thorough understanding of the concepts and procedures involved in successfully maintaining databases, and provide the information and abilities needed to create, use, and maintain databases that are appropriate for organisations.

Course Contents

SECTION A

Introduction to Databases and Transactions: Database, characteristics of data in database, DBMS, advantages of DBMS, file-oriented approach versus Database-oriented approach to Data Management, disadvantages of file- oriented approach, Database System Applications, Purpose of Database Systems, View of Data, Data Abstraction, Instances and Schemas.

Data Models: ER Model, Relational Model, Other Models, Database Languages, DDL, DML, database Access for applications Programs, data base Users and Administrator, Transaction Management, data base Architecture, Storage Manager, the Query Processor Data base design and ER diagrams, ER Model - Entities, Attributes and Entity sets, Relationships and Relationship sets, ER Design Issues, Concept Design, Conceptual Design for University Enterprise. Introduction to the Relational Model, Structure, Database Schema, Keys, Schema Diagrams

Relational databases: Relational Query Languages, Relational Operations. Relational Algebra, Selection and projection set operations, Renaming, Joins, Division, Examples of Algebra Overviews, Relational calculus, Tuple relational Calculus, Domain relational calculus. Overview of the SQL Query Language, Basic Structure of SQL Queries, Set Operations, Aggregate Functions, GROUP BY, HAVING, Nested Sub queries, Views, Triggers.

SECTION B

Normalization: Introduction, Non-loss decomposition and functional dependencies, First, Second, and third normal forms, dependency preservation, Boyee/Codd normal form. Higher Normal Forms - Introduction, Multi-valued dependencies and Fourth normal form, Join dependencies and Fifth normal form

Transaction Concept: Transaction State- Implementation of Atomicity and Durability, Concurrent, Executions, Serializability- Recoverability, Implementation of Isolation, Testing for serializability- Lock – Based Protocols, Timestamp Based Protocols, Validation- Based Protocols, Multiple Granularity. Recovery and Atomicity, Log, Based Recovery, Recovery with Concurrent Transactions, Buffer Management, Failure with loss of non-volatile storage-Advance Recovery systems- Remote Backup systems

References:

1. Introduction to Database Systems by B.C. Desai
2. Database Management Systems by Jerry Post
3. Ramakrishnan, Gehrke, Database Management Syste

4. Database System Concepts by Avi Silberschatz, Henry F. Korth, S Sudarshan.
5. Fundamental of Database Systems- Elmasari, Navathe

Pedagogy:

The Instructor is expected to use leading pedagogical approaches in the class room situation, research-based methodology, innovative instructional methods, extensive use of technology in the class room, online modules of MOOCS, and comprehensive assessment practices to strengthen teaching efforts and improve student learning outcomes.

The Instructor of class will engage in a combination of academic reading, analyzing case studies, preparing the weekly assigned readings and exercises, encouraging in class discussions, and live project based learning.

Case/Class Discussion Assignments:

Students will work in groups of up to four to prepare a brief write-up due before the start of each class covering the case study or class material to be discussed in the next session. Questions may include a quantitative analysis of the problem facing the decision-maker in the case.

Class Participation:

Attendance will be taken at each class. Class participation is scored for each student for each class.

Scheme of Examination

- English will be the medium of instruction and examination.
- Written Examinations will be conducted at the end of each Semester as per the Academic Calendar notified in advance
- Each course will carry 100 marks of which 30 marks shall be reserved for the continuous internal assessment and the remaining 70 marks for written examination to be held at the end of each semester.
- The duration of written examination for each paper shall be three hours.
- The minimum marks for passing the examination for each semester shall be 40% in aggregate as well as a minimum of 40% marks in the semester-end examination in each paper.
- A minimum of 75% of classroom attendance is required in each subject

Instructions for the Paper Setter

The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 12 marks for each question. Section C will have 11 short-answer type questions carrying total of 22 marks, which will cover the entire syllabus uniformly.

Instructions for the Candidates

1. Candidates are required to attempt two questions each from the sections A & B of the question paper and the entire section C.
2. Use of non-programmable scientific calculator is allowed.

MDSM1105P: Software Lab – I

This laboratory course will mainly comprise of exercise based on subject MDSM1103T: Programming in Python

MDSM1106P: Software Lab – II

This laboratory course will mainly comprise of exercise based on subject MDSM1104T: Database Management Systems

MDSM1201T: Mathematics for Machine Learning

Course Objectives: The basic objective of the course is to promote mathematical skills and knowledge for their intrinsic beauty, effectiveness in developing proficiency in analytical reasoning, and utility in modeling and solving real world problems. On the successful completion of the course, the students would be able to:

- Apply the knowledge of set theory and probability theory in various practical applications.
- Understand differentiation and integration and their applications in different fields.
- Apply the concept of Matrices & Determinants in computer science to do computations.
- Have necessary understanding of algebra later required for data science.

Course contents

SECTION A

Linear Algebra: Represent data as vectors and matrices and identify their properties using concepts of singularity, rank, and linear independence, etc. Apply common vector and matrix algebra operations like dot product, inverse, and determinants, Express certain types of matrix operations as linear transformations, apply concepts of eigenvalues and eigenvectors to machine learning problems.

SECTION B

Multivariate Calculus: Analytically optimize different types of functions commonly used in machine learning using properties of derivatives and gradients, Approximately optimize different types of functions commonly used in machine learning using first-order (gradient descent) and second-order (Newton's method) iterative methods, Visually interpret differentiation of different types of functions commonly used in machine learning, Perform gradient descent in neural networks with different activation and cost functions

Principal Component Analysis (PCA): To minimize the average squared reconstruction error between data points.

References:

1. Mathematics for Machine Learning by Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong
2. Essential Math for Data Science by O'reilly
3. Practical Mathematics for AI and Deep Learning by Tamoghna Ghosh, Shravan Kumar Belagal
4. Matrix Methods in Data Mining and Pattern Recognition by Lars Eldén
5. Mathematical Foundations for Data Science by JeanPaul Vert, Leon Bottou, and Stéphane
6. Pattern Recognition and Machine Learning by Christopher M. Bishop

Pedagogy:

The Instructor is expected to use leading pedagogical approaches in the class room situation, research-based methodology, innovative instructional methods, extensive use of technology in the class room, online modules of MOOCS, and comprehensive assessment practices to strengthen teaching efforts and improve student learning outcomes.

The Instructor of class will engage in a combination of academic reading, analyzing case studies, preparing the weekly assigned readings and exercises, encouraging in class discussions, and live project based learning.

Case/Class Discussion Assignments:

Students will work in groups of up to four to prepare a brief write-up due before the start of each class covering the case study or class material to be discussed in the next session. Questions may include a quantitative analysis of the problem facing the decision-maker in the case.

Class Participation:

Attendance will be taken at each class. Class participation is scored for each student for each class.

Scheme of Examination

- English will be the medium of instruction and examination.
- Written Examinations will be conducted at the end of each Semester as per the Academic Calendar notified in advance
- Each course will carry 100 marks of which 30 marks shall be reserved for the continuous internal assessment and the remaining 70 marks for written examination to be held at the end of each semester.
- The duration of written examination for each paper shall be three hours.
- The minimum marks for passing the examination for each semester shall be 40% in aggregate as well as a minimum of 40% marks in the semester-end examination in each paper.
- A minimum of 75% of classroom attendance is required in each subject

Instructions for the Paper Setter

The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 12 marks for each question. Section C will have 11 short-answer type questions carrying total of 22 marks, which will cover the entire syllabus uniformly.

Instructions for the Candidates

1. Candidates are required to attempt two questions each from the sections A & B of the question paper and the entire section C.
2. Use of non-programmable scientific calculator is allowed.

MDSM1202T: Advanced Statistical Techniques

Course Objectives: The objectives of the course if to enhance the ability of the students to:

- summarize and present data numerically and visually.
- Determine which statistical methods to use in which situations
- think critically about data-based claims and quantitative arguments
- learn new statistical analysis techniques on your own

Course Contents

SECTION A

Sampling and Statistical Estimation Theory: Population and Sample, Random Samples, Sampling With and Without Replacement, Sampling Distributions, Sampling distributions of Mean, Proportionand Difference of Means, Standard Error, Parameter, Estimation of Parameters, Estimator, Properties of Estimators:(Unbiasedness, Consistency, Efficiency, Sufficiency), Point and Interval Estimates, Methods of point estimation, Method of Moments, Properties of Moment Method Estimators, Principal of Maximum Likelihood: properties andapplications, Method of least squares, Confidence Interval and confidence limits of Population Parameters Based on Normal, t and Chi-square Distributions.

Statistical Inference: Statistical Decisions, Statistical Hypothesis, Null and alternate hypothesis, simple & composite hypotheses, Tests of statistical Hypothesis and Significance, critical region, Types of errors, Level of significance, Likelihood Ratio tests: Description and properties.

SECTION B

Test of Sampling Distributions: Test of significance of Large sample, Test of single proportion and difference of proportions, Test of significance of Single Mean and difference of Means. Chi-square Test for population variance, Chi-square test of goodness of fit, Student t-test, t-test for single mean and difference of means, F-test for equality of population variances, F-test for equality of several means.

Correlation and Regression: Karl Pearson Coefficient of correlation, Limits of Correlation coefficient, Shortcut and Step deviation method, Correlation of Ranks, Regression, Regression Coefficient, Properties of Regressions, Lines of Regression.

References:

1. Gupta, S.C.andKapoor, V.K.(1997) Fundamentals of Mathematical Statistics. Sultan Chand and Sons, New Delhi, 11.23-12.23.
2. Spiegel, M. R., Schiller, J. J., and Alu Srinivasan, R. (2013). Probability and Statistics, Fourth Edition, Schaum's Outline Series, McGraw Hill Companies, Inc
3. Manoj Kumar Srivastava and Namita Srivastava, Statistical Inference–Testing of Hypotheses, Prentice Hall of India, 2014.
4. Bansi lal, Sanjay Arora and Sudha Arora, Introducing Probability and Statistics, 2/e, Satya Prakash Publications, 2006.

Pedagogy:

The Instructor is expected to use leading pedagogical approaches in the class room situation, research-based methodology, innovative instructional methods, extensive use of technology in the class room, online modules of MOOCS, and comprehensive assessment practices to strengthen teaching efforts and improve student learning outcomes.

The Instructor of class will engage in a combination of academic reading, analyzing case studies, preparing the weekly assigned readings and exercises, encouraging in class discussions, and live project based learning.

Case/Class Discussion Assignments:

Students will work in groups of up to four to prepare a brief write-up due before the start of each class covering the case study or class material to be discussed in the next session. Questions may include a quantitative analysis of the problem facing the decision-maker in the case.

Class Participation:

Attendance will be taken at each class. Class participation is scored for each student for each class.

Scheme of Examination

- English will be the medium of instruction and examination.
- Written Examinations will be conducted at the end of each Semester as per the Academic Calendar notified in advance
- Each course will carry 100 marks of which 30 marks shall be reserved for the continuous internal assessment and the remaining 70 marks for written examination to be held at the end of each semester.
- The duration of written examination for each paper shall be three hours.
- The minimum marks for passing the examination for each semester shall be 40% in aggregate as well as a minimum of 40% marks in the semester-end examination in each paper.
- A minimum of 75% of classroom attendance is required in each subject

Instructions for the Paper Setter

The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 12 marks for each question. Section C will have 11 short-answer type questions carrying total of 22 marks, which will cover the entire syllabus uniformly.

Instructions for the Candidates

1. Candidates are required to attempt two questions each from the sections A & B of the question paper and the entire section C.
2. Use of non-programmable scientific calculator is allowed.

MDSM1203T: Data Analytics and Visualisation

Course Objectives: Course Outcome: This Course will help students gain data scientist skillset by learning data science using analytical tools and also enables them to master analytical techniques like data exploration, data visualization and various predictive analytic techniques. On completion of this course, the students will be able to:

- Learn about essential Python libraries such as NumPy, Pandas, Matplotlib, IPython, Jupyter, SciPy, Scikit-learn and Statsmodels.
- Become familiar with data cleaning and preparation, data aggregation.
- Understand and perform Exploratory Data Analysis
- Understand the steps in model fitting and parameters fine-tuning.
- Apply model validation techniques.

Course Contents

SECTION A

Data Analytics: What is data analysis? Libraries like numpy and pandas for data analysis and data science.

Pandas: Data Frames and Series, Data retrieval, Data Ingestion, Data Preparation, Data Wrangling and Data Aggregation, Pandas with functions, Analysis using Groupings, Merging Data Frames, Visualisation with Pandas.

Numpy: Build multidimensional arrays and perform high-level mathematical functions, array slicing and fancy indexing, broadcasting, filtering and analysing data in numpy arrays. Numpy for image storage and analysis.

SECTION B

Data Visualisation: Meaning and Significance, Visualisation for data, findings, and insights.Visualisation with Matplotlib, Visualisation with Seaborn, Tools like Excel for Visualisation.

References:

1. Python – Data Visualisation by Samuel Burns
2. Storytelling with Data: A Data Visualization Guide for Business Professionals by Cole Nussbaumer Knaflic
3. Data Science for Business by Foster Provost and Tom Fawcett:
4. Data Visualization: A Practical Introduction by Kieran Healy
5. Information Dashboard Design: The Effective Visual Communication of Data by Stephen Few
6. Data Visualization with Excel Dashboards and Reports by Dick Kusleika

Pedagogy:

The Instructor is expected to use leading pedagogical approaches in the class room situation, research-based methodology, innovative instructional methods, extensive use of technology in the class room, online modules of MOOCS, and comprehensive assessment practices to strengthen teaching efforts and improve student learning outcomes.

The Instructor of class will engage in a combination of academic reading, analyzing case studies, preparing the weekly assigned readings and exercises, encouraging in class discussions, and live project based learning.

Case/Class Discussion Assignments:

Students will work in groups of up to four to prepare a brief write-up due before the start of each class covering the case study or class material to be discussed in the next session. Questions may include a quantitative analysis of the problem facing the decision-maker in the case.

Class Participation:

Attendance will be taken at each class. Class participation is scored for each student for each class.

Scheme of Examination

- English will be the medium of instruction and examination.
- Written Examinations will be conducted at the end of each Semester as per the Academic Calendar notified in advance
- Each course will carry 100 marks of which 30 marks shall be reserved for the continuous internal assessment and the remaining 70 marks for written examination to be held at the end of each semester.
- The duration of written examination for each paper shall be three hours.
- The minimum marks for passing the examination for each semester shall be 40% in aggregate as well as a minimum of 40% marks in the semester-end examination in each paper.
- A minimum of 75% of classroom attendance is required in each subject

Instructions for the Paper Setter

The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 12 marks for each question. Section C will have 11 short-answer type questions carrying total of 22 marks, which will cover the entire syllabus uniformly.

Instructions for the Candidates

1. Candidates are required to attempt two questions each from the sections A & B of the question paper and the entire section C.
2. Use of non-programmable scientific calculator is allowed.

MDSM1204T: Web Development using Python and Django

Course Objectives: This Course will help students gain proficiency in Web development skills by learning-by-doing approach using using Python and Django and also enables them to master programming constructs like data input/ output, looping and various file management techniques. On completion of this course, the students will be able:

- To develop proficiency in creating applications using the Python Programming Language.
- To understand the various data structures available in Python programming language and apply them in solving computational problems.
- To handle regular expressions, files and exception in Python
- To install and understand basic application structure of Django.
- To design templates, web forms and handle database in Django
- Develop large database applications and other supports such as e-mail in Django
- To be able to do testing and debugging of code written in Python and Django.

Course Contents

SECTION A

Introduction to Python: Variables, Numeric Data Types: Int& Float, Sequential Types: Str& List, If-Elif-Else Statements, Creating Python Scripts/Files, Definite loops: For loop,Indexing,Slicing

Advanced Data Types in Python : Data Types: Tuples, Dictionaries, Reading and Writing txt files, Writing Iterative Code and Algorithms:

Indefinite loops: while loops, Runtime complexity, Searching and sorting algorithms, Custom functions in Python, Args and Kwargs,Importing files

Introduction to Object Oriented Programming: Classes, Methods, Inheritance, Abstract Classes Function, Module, Numpy, SQL, Concepts of Web Page

SECTION B

Introduction to Back-End Web Development using Django, HTTP protocol, MVC model, Virtual environment, Introduction to Django, Django and Python, Django' s take on MVC: Model, View and Template, DRY programming: Don't Repeat Yourself, How to get and install Django, Getting started with Django, About the 3 Core Files: models.py, urls.py, views.py, Setting up database connections, Managing Users & the Django admin tool, Installing and using 'out of the box' Django features, Django URL Patterns and Views, Designing a good URL scheme, Generic Views, Django Forms, Form classes, Validation, Authentication, Advanced Forms processing techniques, Creating CURD application in Django

References:

1. Web Development with Django by Ben Shaw, Saurabh Badhwar, Andrew Bird
2. Django: Web Development with Pythonby Samuel Dauzon, AidasBendoraitis, Arun Ravindran
3. Mastering Django: Core – The Complete Guide to Django 1.8 LTS*by Nigel George*
4. Django Unleashed*by Andrew Pinkham*

5. The Definitive Guide to Django: Web Development Done Right*by Adrian Holovaty and Jacob Kaplan-Moss*
6. Django 4 By Example: Build powerful and reliable Python web applications from scratchby Antonio Mele (Author), Bob Belderbos (Author)
7. Django for Beginners: Build websites with Python and Django*by William S. Vincent*
8. Beginning Django By Daniel Rubio

Pedagogy:

The Instructor is expected to use leading pedagogical approaches in the class room situation, research-based methodology, innovative instructional methods, extensive use of technology in the class room, online modules of MOOCS, and comprehensive assessment practices to strengthen teaching efforts and improve student learning outcomes.

The Instructor of class will engage in a combination of academic reading, analyzing case studies, preparing the weekly assigned readings and exercises, encouraging in class discussions, and live project based learning.

Case/Class Discussion Assignments:

Students will work in groups of up to four to prepare a brief write-up due before the start of each class covering the case study or class material to be discussed in the next session. Questions may include a quantitative analysis of the problem facing the decision-maker in the case.

Class Participation:

Attendance will be taken at each class. Class participation is scored for each student for each class.

Scheme of Examination

- English will be the medium of instruction and examination.
- Written Examinations will be conducted at the end of each Semester as per the Academic Calendar notified in advance
- Each course will carry 100 marks of which 30 marks shall be reserved for the continuous internal assessment and the remaining 70 marks for written examination to be held at the end of each semester.
- The duration of written examination for each paper shall be three hours.
- The minimum marks for passing the examination for each semester shall be 40% in aggregate as well as a minimum of 40% marks in the semester-end examination in each paper.
- A minimum of 75% of classroom attendance is required in each subject

Instructions for the Paper Setter

The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 12 marks for each question. Section C will have 11 short-answer type questions carrying total of 22 marks, which will cover the entire syllabus uniformly.

Instructions for the Candidates

1. Candidates are required to attempt two questions each from the sections A & B of the question paper and the entire section C.
2. Use of non-programmable scientific calculator is allowed.

MDSM1205P: Software Lab – III

This laboratory course will mainly comprise of exercise based on subject **MDSM1203T**.

MDSM1206P: Software Lab – IV

This laboratory course will mainly comprise of exercise based on subject MDSM2PUP1204T (Web Development using Python and DJango).

MDSM2101T: Soft Skills and Technical Communication

Course Objectives: A soft skills and technical communication course aims to give students the technical competencies and communication strategies they need to succeed in the workplace, especially in technical or professional settings. The goal of the course is for students to improve their professional, interpersonal, and communication skills in order to increase their overall employability and chances for career growth. Students will work on honing their presentation skills, business writing, and email etiquette as well as their oral and written communication abilities. They must be able to effectively communicate knowledge to a variety of audiences, both inside and beyond their specialty. The development of professional networks, making contacts, and maximising possibilities for career growth of students. They should learn how to write resumes, conduct job searches, prepare for interviews, and maintain a professional internet presence.

Course contents

SECTION A

Soft Skills: Soft Skills, Importance of communication, Types of communication – Verbal/ Non-verbal, Barriers of Communication, Interview dress code, controlling nerves, positive visualization, creating an impression, Selling yourself at the interview, Mock interview, Self Esteem & Stress Management, Leadership Skills

Corporate Culture: Importance of etiquette, Professional etiquette, Social etiquette Unit, Corporate Culture, Professional ethics, Mutual respect, Time management, Team Work

CRM- The Management Model

SECTION B

Speaking Skill: Pronunciation, Vocabulary,. Public Speaking a Telephone Etiquettes & Call Handling Skills, Group Discussion, Mock P.I./Role plays, Mock G.D./ Role plays, Personal Interview skills.

Listening Skills: Types of Listening, Tips for Effective Listening, Academic Listening- (lecturing), Listening to Talks and Presentations, Listening to Announcements.

Writing Skill: Report writing, CV/Resume writing, Cover Letters, Business letters for general/professional purposes, Email Writing and Forwarding.

References:

1. Personality Development and Soft Skills (Old Edition) by Barun K Mitra
2. Soft Skills - Enhancing Employability: Connecting Campus with Corporate by M S Rao
3. Design Your Career- Soft Skills Of Career by Yogesh P Chopade
4. Soft Skills by Hariharan S and S P Shanmugapriya

Pedagogy:

The Instructor is expected to use leading pedagogical approaches in the class room situation, research-based methodology, innovative instructional methods, extensive use of technology in the class room, online modules of MOOCS, and comprehensive assessment practices to strengthen teaching efforts and improve student learning outcomes.

The Instructor of class will engage in a combination of academic reading, analyzing case studies, preparing the weekly assigned readings and exercises, encouraging in class discussions, and live project based learning.

Case/Class Discussion Assignments:

Students will work in groups of up to four to prepare a brief write-up due before the start of each class covering the case study or class material to be discussed in the next session. Questions may include a quantitative analysis of the problem facing the decision-maker in the case.

Class Participation:

Attendance will be taken at each class. Class participation is scored for each student for each class.

Scheme of Examination

- English will be the medium of instruction and examination.
- Written Examinations will be conducted at the end of each Semester as per the Academic Calendar notified in advance
- Each course will carry 100 marks of which 30 marks shall be reserved for the continuous internal assessment and the remaining 70 marks for written examination to be held at the end of each semester.
- The duration of written examination for each paper shall be three hours.
- The minimum marks for passing the examination for each semester shall be 40% in aggregate as well as a minimum of 40% marks in the semester-end examination in each paper.
- A minimum of 75% of classroom attendance is required in each subject

Instructions for the Paper Setter

The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 12 marks for each question. Section C will have 11 short-answer type questions carrying total of 22 marks, which will cover the entire syllabus uniformly.

Instructions for the Candidates

1. Candidates are required to attempt two questions each from the sections A & B of the question paper and the entire section C.
2. Use of non-programmable scientific calculator is allowed.

DSM2102T: Software Engineering for Data Scientist

Course Objectives: In this course, students will gain a broad understanding of the discipline of software engineering for data scientists and its application to the development of and management of software systems.

Course contents

SECTION A

Introduction to software engineering, requirement of software engineering for Data Scientist, stages of software engineering for Data Scientists, cleaning, use of OO programming, developing packages, version control, scripting, testing, cloud engineering, DevOps, Data analysis.

Software process models, Agile Development model, software requirement engineering, architectural design, real time system design,

Data science workflow, data pipeline, components of data pipeline, deployment models machine learning, data ingestion, pre-processing, training, evaluation, perdition, deployment and monitoring.

Writing robust code, PEP8 Standard, formatting and style of code, modular code, code summary, exception handling, documentation: name, comments, docstrings, readability

SECTION B

Analysing Code performance, time and memory use, time complexity, Optimization of code, vectorization of operations, why to use of Numpy over loops, identification of faster data structures.

Case Study of Data Science Application with Python and Software Engineering practices, Health Care, Banking.

Reference Books

1. Software Engineering for Data Scientist, by Catherine Nelson, O'Reilly Media Inc.
2. Software Engineering, by K. K. Agarwal and Yogesh Singh, New Age International Publication, 2018 edition
3. Perspectives on Data Science for Software Engineering, by Tim Menzies, Laurie Williams, Thomas Zimmermann, Elsevier, 2016
4. Software Engineering: A Practitioner's Approach, by Pressman, 8th Edition, McGraw Hills

Pedagogy:

The Instructor is expected to use leading pedagogical approaches in the class room situation, research-based methodology, innovative instructional methods, extensive use of technology in the class room, online modules of MOOCS, and comprehensive assessment practices to strengthen teaching efforts and improve student learning outcomes.

The Instructor of class will engage in a combination of academic reading, analyzing case studies, preparing the weekly assigned readings and exercises, encouraging in class discussions, and live project based learning.

Case/Class Discussion Assignments:

Students will work in groups of up to four to prepare a brief write-up due before the start of each class covering the case study or class material to be discussed in the next session. Questions may include a quantitative analysis of the problem facing the decision-maker in the case.

Class Participation:

Attendance will be taken at each class. Class participation is scored for each student for each class.

Scheme of Examination

- English will be the medium of instruction and examination.
- Written Examinations will be conducted at the end of each Semester as per the Academic Calendar notified in advance
- Each course will carry 100 marks of which 30 marks shall be reserved for the continuous internal assessment and the remaining 70 marks for written examination to be held at the end of each semester.
- The duration of written examination for each paper shall be three hours.
- The minimum marks for passing the examination for each semester shall be 40% in aggregate as well as a minimum of 40% marks in the semester-end examination in each paper.
- A minimum of 75% of classroom attendance is required in each subject

Instructions for the Paper Setter

The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 12 marks for each question. Section C will have 11 short-answer type questions carrying total of 22 marks, which will cover the entire syllabus uniformly.

Instructions for the Candidates

1. Candidates are required to attempt two questions each from the sections A & B of the question paper and the entire section C.
2. Use of non-programmable scientific calculator is allowed.

MDSM2103T: Business Intelligence

Course Objectives: The objectives of this course are to provide comprehensive and in-depth knowledge of Business Intelligence (BI) principles and techniques by introducing the relationship between managerial and technological perspectives. The course will cover general concepts in the BI field (report authoring, ETL). The focus will be on how the techniques are to be used, and the details of the methodologies will be covered to the extent necessary to understand when and how each technique can be used.

Course Contents

SECTION A

Business Intelligence: Introduction, Meaning, Purpose and Structure of Business Intelligence Systems. Data-driven results through business intelligence, similarities and differences between BI and data analytics. Business intelligence tools and techniques.

BI Analysis using Spreadsheet software, Using Database MySQL for managing data, Understanding SQL and executing queries.

SECTION B

BI Tools like Power BI for visualization to enhance BI capabilities and building dashboards.

Machine Learning and Business Intelligence, Differences between ML and BI, Changing BI using ML in automating tasks, segmenting customers, uncovering hidden patterns, working on unstructured data, Challenges brought by ML in BI

Machine Learning for Predictions, Supervised and Unsupervised, Linear Regression and Classification, Algorithms like K-neighbors, Naïve Bayes, Decision Tress and its Implementations, Accuracy of Models.

References:

1. Business Intelligence Guidebook: From Data Integration to Analytics by Rick Sherman
2. Business Intelligence: A Managerial Perspective on Analytics by Ramesh Sharda, Dursun, Efraim
3. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications by Larissa T. Moss and Shaku Atre
4. 4.The state of AI in Business Intelligence (e-book)
5. The Definitive Guide to DAX: Business Intelligence for Microsoft Power BI, SQL Server Analysis Services, and Excel by Marco Russo and Alberto Ferrari
6. Introduction to Machine Learning with Python, O'reilly
7. Understanding Machine Learning: From Theory To Algorithms by Shai Shalev-Shwartz
8. Learning MySQL by Hugh E. Williams (O'reilly
9. 11. Learning SQL: Generate, Manipulate, and Retrieve Data by Alan Beaulieu

Pedagogy:

The Instructor is expected to use leading pedagogical approaches in the class room situation, research-based methodology, innovative instructional methods, extensive use of technology in the class room, online modules of MOOCS, and comprehensive assessment practices to strengthen teaching efforts and improve student learning outcomes.

The Instructor of class will engage in a combination of academic reading, analyzing case studies, preparing the weekly assigned readings and exercises, encouraging in class discussions, and live project based learning.

Case/Class Discussion Assignments:

Students will work in groups of up to four to prepare a brief write-up due before the start of each class covering the case study or class material to be discussed in the next session. Questions may include a quantitative analysis of the problem facing the decision-maker in the case.

Class Participation:

Attendance will be taken at each class. Class participation is scored for each student for each class.

Scheme of Examination

- English will be the medium of instruction and examination.
- Written Examinations will be conducted at the end of each Semester as per the Academic Calendar notified in advance
- Each course will carry 100 marks of which 30 marks shall be reserved for the continuous internal assessment and the remaining 70 marks for written examination to be held at the end of each semester.
- The duration of written examination for each paper shall be three hours.
- The minimum marks for passing the examination for each semester shall be 40% in aggregate as well as a minimum of 40% marks in the semester-end examination in each paper.
- A minimum of 75% of classroom attendance is required in each subject

Instructions for the Paper Setter

The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 12 marks for each question. Section C will have 11 short-answer type questions carrying total of 22 marks, which will cover the entire syllabus uniformly.

Instructions for the Candidates

1. Candidates are required to attempt two questions each from the sections A & B of the question paper and the entire section C.
2. Use of non-programmable scientific calculator is allowed.

MDSM2105P: Software Lab – V

This laboratory course will mainly comprise of exercise based on subject MDSM2103T.

MDSM2106P: Software Lab – VI

This laboratory course will mainly comprise of exercise based on Elective – I from MDSM2104E1 to MDSM2104E5.

MDSM2104E1: Digital Image Processing

Course Objectives: This course is designed to introduce students with the fundamental concepts and techniques in basic digital image processing and their applications to solve real life problems. The topics covered include Digital Image Fundamentals, Image Transforms, Image Enhancement, Restoration and Compression, Morphological Image Processing, Nonlinear Image Processing, and Image Analysis. Application examples are also included.

Course Contents

SECTION A

Introduction: Digital Image Fundamentals: Brightness, Adaptation and Discrimination, Light and Electromagnetic Spectrum, Image Sampling and Quantization, Some Basic Relationships between Pixels Types of images.

Spatial Domain Filtering: Some Basic Intensity Transformation Functions, Histogram Equalization, Spatial Correlation and Convolution, Smoothening Spatial Filters: Low pass filters, Order Statistics filters; Sharpening Spatial Filters: Laplacian filter

Filtering in Frequency Domain: The Discrete Fourier Transformation (DFT), Frequency Domain Filtering: Ideal and Butterworth Low pass and High pass filters, DCT Transform (1D, 2D).

SECTION B

Image Restoration: Image Degradation/Restoration Process, Noise models, Noise Restoration Filters

Image Compression: Fundamentals of Image Compression, Huffman Coding, Run Length Coding, JPEG.

Morphological Image Processing: Erosion, Dilation, Opening, Closing, Hit-or-Miss Transformation, Basic Morphological Algorithms.

Image Segmentation: Point, Line and Edge Detection, Thresholding, Region Based Segmentation.

References:

1. Gonzalez, R. C., & Woods, R. E. (2017). Digital Image Processing. 4th edition. Pearson Education.
2. Jain, A. K. (1988). Fundamentals of Digital Image Processing. 1st edition Prentice Hall of India.
3. Castleman, K. R. (1995.). Digital Image Processing. 1st edition. Pearson Education
4. Gonzalez, R. C., Woods, R. E., & Eddins, S. (2004). Digital Image Processing using MATLAB. Pearson Education Inc.
5. Schalkoff, D. (1989). Image Processing and Computer Vision. 1st edition. John Wiley and Sons.

Pedagogy:

The Instructor is expected to use leading pedagogical approaches in the class room situation, research-based methodology, innovative instructional methods, extensive use of technology in the class room, online modules of MOOCS, and comprehensive assessment practices to strengthen teaching efforts and improve student learning outcomes.

The Instructor of class will engage in a combination of academic reading, analyzing case studies, preparing the weekly assigned readings and exercises, encouraging in class discussions, and live project based learning.

Case/Class Discussion Assignments:

Students will work in groups of up to four to prepare a brief write-up due before the start of each class covering the case study or class material to be discussed in the next session. Questions may include a quantitative analysis of the problem facing the decision-maker in the case.

Class Participation:

Attendance will be taken at each class. Class participation is scored for each student for each class.

Scheme of Examination

- English will be the medium of instruction and examination.
- Written Examinations will be conducted at the end of each Semester as per the Academic Calendar notified in advance
- Each course will carry 100 marks of which 30 marks shall be reserved for the continuous internal assessment and the remaining 70 marks for written examination to be held at the end of each semester.
- The duration of written examination for each paper shall be three hours.
- The minimum marks for passing the examination for each semester shall be 40% in aggregate as well as a minimum of 40% marks in the semester-end examination in each paper.
- A minimum of 75% of classroom attendance is required in each subject

Instructions for the Paper Setter

The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 12 marks for each question. Section C will have 11 short-answer type questions carrying total of 22 marks, which will cover the entire syllabus uniformly.

Instructions for the Candidates

1. Candidates are required to attempt two questions each from the sections A & B of the question paper and the entire section C.
2. Use of non-programmable scientific calculator is allowed.

MDSM2104E2: Natural Language Processing

Course Objectives: This course introduces the fundamental concepts and techniques of natural language processing (NLP). Students will gain an in-depth understanding of the computational properties of natural languages and the commonly used algorithms for processing linguistic information.

Course Contents

SECTION A

Natural Language Processing – Introduction, Stemming and Lemmatization, Tokenizing, Web Scraping, Part of Speech(POS) Tagging and Sequence Labeling
Named Entity Recognition (NER), Text classification with RNN, Sentiment Analysis, Ethical consideration for NLP, NLTK toolkit
Text Features and TF-IDF Classification

SECTION B

Applications using NLP: Machine translation and learning, summarization and natural language generation, natural language querying, information extraction.

References:

1. Steven Bird, Ewan Klein, Edward Loper, Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit (O'Reilly 2009, website 2018) *http://www.nltk.org/book/*
2. Dipanjan Sarkar - Text Analytics with Python, A Practitioner's Guide to Natural Language Processing (Apress)
3. Kamath, Uday - Deep learning for NLP and speech recognition, Springer
4. Alexander Clark, Chris Fox, Shalom Lappin -The Handbook of Computational Linguistics and Natural Language Processing (Wiley)
5. Dan Jurafsky, James H. Martin - Speech and Language Processing
6. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (Pearson Prentice Hall)

Pedagogy:

The Instructor is expected to use leading pedagogical approaches in the class room situation, research-based methodology, innovative instructional methods, extensive use of technology in the class room, online modules of MOOCS, and comprehensive assessment practices to strengthen teaching efforts and improve student learning outcomes.
The Instructor of class will engage in a combination of academic reading, analyzing case studies, preparing the weekly assigned readings and exercises, encouraging in class discussions, and live project based learning.

Case/Class Discussion Assignments:

Students will work in groups of up to four to prepare a brief write-up due before the start of each class covering the case study or class material to be discussed in the next session. Questions may include a quantitative analysis of the problem facing the decision-maker in the case.

Class Participation:

Attendance will be taken at each class. Class participation is scored for each student for each class.

Scheme of Examination

- English will be the medium of instruction and examination.
- Written Examinations will be conducted at the end of each Semester as per the Academic Calendar notified in advance
- Each course will carry 100 marks of which 30 marks shall be reserved for the continuous internal assessment and the remaining 70 marks for written examination to be held at the end of each semester.
- The duration of written examination for each paper shall be three hours.
- The minimum marks for passing the examination for each semester shall be 40% in aggregate as well as a minimum of 40% marks in the semester-end examination in each paper.
- A minimum of 75% of classroom attendance is required in each subject

Instructions for the Paper Setter

The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 12 marks for each question. Section C will have 11 short-answer type questions carrying total of 22 marks, which will cover the entire syllabus uniformly.

Instructions for the Candidates

1. Candidates are required to attempt two questions each from the sections A & B of the question paper and the entire section C.
2. Use of non-programmable scientific calculator is allowed.

MDSM2104E3: Neural Networks and Fuzzy Logic

Course Objectives: This course is designed to explore the fundamental theory and concepts of computational intelligence methods, in particular neural networks, fuzzy systems, genetic algorithms and their applications in the area of machine intelligence. Upon completion of this course students will be able to:

- To understand the fundamental theory and concepts of neural networks, neuro-modeling, several neural network paradigms and its applications.
- To understand the concepts of fuzzy sets, knowledge representation using fuzzy rules, approximate reasoning, fuzzy inference systems, and fuzzy logic control and other machine intelligence applications of fuzzy logic
- To understand the basics of an evolutionary computing paradigm known as genetic algorithms and its application to engineering optimization problems.

Course Contents

SECTION A

Introduction to Neural Networks: Biological Neurons, Artificial Neurons – various models, Artificial Neural Networks – various structures, Learning Strategies, Applications.

SECTION B

Introduction to Fuzzy sets: Fuzzy relation, Approximate reasoning, Rules; Fuzzy control design parameters: Rule base, data base; Choice of fuzzification procedure; Choice of defuzzification procedure; Nonlinear fuzzy control; Adaptive fuzzy control

References:

1. A Brief Introduction to Neural Networks, David Kriesel, 2005
2. Introduction to Fuzzy Logic using MATLAB, S. N. Sivanandam, and others, 2007, Springer
3. Neural Networks: A Comprehensive Study By:SimonHyken. Macmillan Colledge Publishing, Inc.1996
4. Fuzzy Control and Fuzzy System. By: Witold Pedrycz.Research Studies Press Ltd.2ndd edition 1996
5. Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering, Nikola K. Kasabov, 1998, MIT Press.
6. S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice- Hall India, 2nd Edition, 1999.
7. J. S. R. Jang, C. T. Sun, and E. Mizutani, Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence, 2nd Edition, Pearson Education, 2005.
8. J. A. Freeman and D. M. Skapura, Neural Networks: Algorithms, Applications, and Programming Techniques, 1st Edition, Pearson Education, 2007.

Pedagogy:

The Instructor is expected to use leading pedagogical approaches in the class room situation, research-based methodology, innovative instructional methods, extensive use of technology in the class room, online

modules of MOOCS, and comprehensive assessment practices to strengthen teaching efforts and improve student learning outcomes.

The Instructor of class will engage in a combination of academic reading, analyzing case studies, preparing the weekly assigned readings and exercises, encouraging in class discussions, and live project based learning.

Case/Class Discussion Assignments:

Students will work in groups of up to four to prepare a brief write-up due before the start of each class covering the case study or class material to be discussed in the next session. Questions may include a quantitative analysis of the problem facing the decision-maker in the case.

Class Participation:

Attendance will be taken at each class. Class participation is scored for each student for each class.

Scheme of Examination

- English will be the medium of instruction and examination.
- Written Examinations will be conducted at the end of each Semester as per the Academic Calendar notified in advance
- Each course will carry 100 marks of which 30 marks shall be reserved for the continuous internal assessment and the remaining 70 marks for written examination to be held at the end of each semester.
- The duration of written examination for each paper shall be three hours.
- The minimum marks for passing the examination for each semester shall be 40% in aggregate as well as a minimum of 40% marks in the semester-end examination in each paper.
- A minimum of 75% of classroom attendance is required in each subject

Instructions for the Paper Setter

The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 12 marks for each question. Section C will have 11 short-answer type questions carrying total of 22 marks, which will cover the entire syllabus uniformly.

Instructions for the Candidates

1. Candidates are required to attempt two questions each from the sections A & B of the question paper and the entire section C.
2. Use of non-programmable scientific calculator is allowed.

MDSM2104E4: Big Data Analytics

Course Objectives:

- Understand the Big Data Platform and its Use cases
- Provide an overview of Apache Hadoop
- Provide HDFS Concepts and Interfacing with HDFS
- Understand Map Reduce Jobs
- Provide hands on Hadoop Eco System
- Apply analytics on Structured, Unstructured Data.

Course Contents

SECTION A

Introduction to Virtual Machine, creating and configuring Virtual Machine, Installing Ubuntu Operating System on Virtual Machine, Operating System Concepts: Linux History, Benefits of Linux, Different flavors of Linux, Introducing Ubuntu, Installing Ubuntu: Starting Up, Logging in, Exploring the Desktop, Ubuntu Basics, Understanding Linux Files/Directories: Managing Files: Ubuntu Commands, Running Basic commands, Piping and Filtering Commands, Directory and File handling commands.Users, Groups and Permissions, Root and Other Users, Adding and Deleting Users and Groups, Adding and Changing Passwords, Users and File Permissions.

Introduction to big data and Hadoop: Types of Digital Data, Introduction to Big Data, Big Data Analytics, History of Hadoop, Apache Hadoop, Analysing Data with Unix tools, Analysing Data with Hadoop, Hadoop Streaming, Hadoop Echo System

HDFS(Hadoop Distributed File System): The Design of HDFS, HDFS Concepts, Command Line Interface, Hadoop file system interfaces, Data flow, Data Ingest with Flume and Scoop and Hadoop archives, Hadoop I/O: Compression, Serialization, Avro and File-Based Data structures.

SECTION B

Map Reduce: Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features

Hadoop Eco System: Pig : Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators.

Hive : Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, Querying Data and User Defined Functions.

Hbase :HBasics, Concepts, Clients, Example, Hbase Versus RDBMS.

References:

1. Tom White Hadoop: The Definitive Guide Third Edition, O'reily Media, 2012.
2. Seema Acharya, Subhasini Chellappan, Big Data Analytics Wiley 2015.
3. Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer, 2007.
4. Jay Liebowitz, Big Data and Business Analytics Auerbach Publications, CRC press(2013)
5. Michael Minelli, Michele Chambers, Ambiga Dhiraj, Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, Wiley Publications, 2013.
6. Arvind Sathi, Big Data Analytics: Disruptive Technologies for Changing the Game, MC Press,

2012
7. Paul Zikopoulos, Dirk De Roos, Krishnan Parasuraman, Thomas Deutsch, James Giles, David Corigan, Harness the Power of Big Data The IBM Big Data Platform ,Tata McGraw Hill Publications, 2012.

Pedagogy:

The Instructor is expected to use leading pedagogical approaches in the class room situation, research-based methodology, innovative instructional methods, extensive use of technology in the class room, online modules of MOOCS, and comprehensive assessment practices to strengthen teaching efforts and improve student learning outcomes.

The Instructor of class will engage in a combination of academic reading, analyzing case studies, preparing the weekly assigned readings and exercises, encouraging in class discussions, and live project based learning.

Case/Class Discussion Assignments:

Students will work in groups of up to four to prepare a brief write-up due before the start of each class covering the case study or class material to be discussed in the next session. Questions may include a quantitative analysis of the problem facing the decision-maker in the case.

Class Participation:

Attendance will be taken at each class. Class participation is scored for each student for each class.

Scheme of Examination

- English will be the medium of instruction and examination.
- Written Examinations will be conducted at the end of each Semester as per the Academic Calendar notified in advance
- Each course will carry 100 marks of which 30 marks shall be reserved for the continuous internal assessment and the remaining 70 marks for written examination to be held at the end of each semester.
- The duration of written examination for each paper shall be three hours.
- The minimum marks for passing the examination for each semester shall be 40% in aggregate as well as a minimum of 40% marks in the semester-end examination in each paper.
- A minimum of 75% of classroom attendance is required in each subject

Instructions for the Paper Setter

The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 12 marks for each question. Section C will have 11 short-answer type questions carrying total of 22 marks, which will cover the entire syllabus uniformly.

Instructions for the Candidates

1. Candidates are required to attempt two questions each from the sections A & B of the question paper and the entire section C.
2. Use of non-programmable scientific calculator is allowed.

MDSM2104E5: Data Mining and Warehouse

Course Objective: This Course will help students gain data scientist skillset by learning data mining and warehouse concepts. On completion of this course, the students will be able to understand concepts of knowledge discovery, various applications of data warehouse and analysis techniques.

Course Contents

SECTION A

Knowledge Discovery in Databases (KDD), Data Mining Concepts and its Applications, stages and models of Data Mining, Data Warehousing and On-Line Analytical Processing (OLAP) and its operation, Need for Data Warehousing, architecture, Challenges, Application of Data Mining Principles, OLTP Vs Data Warehouse Applications of Data Warehouse. Data warehouse models: Enterprise warehouse, Data mart and virtual warehouse, Extraction, Transformation and loading, Data Cube, Multidimensional data model, Stars, Snowflakes and Fact constellations, Schemas for multidimensional Data models, Dimensions: concepts of Hierarchies, Measures and Their Categorization & computation

Data Pre-processing: Data Pre-processing Concepts, Data Cleaning, Data integration and transformation, Data Reduction, Discretization and concept hierarchy

SECTION B

Association Analysis: Association Analysis: Problem Definition, Frequent Item set Generation, Rule generation. Alternative Methods for Generating Frequent Item sets, FP-Growth Algorithm, Evaluation of Association Patterns.

Classification: Decision Trees Induction, Method for Comparing Classifiers, Rule Based Classifiers, Nearest Neighbour Classifiers, Bayesian Classifiers

Clustering Analysis: Overview, K-Means, Agglomerative Hierarchical 8 Hours Clustering, DBSCAN, Cluster Evaluation, Density-Based Clustering, Graph Based Clustering, Scalable Clustering Algorithms.

Data visualization techniques, pixel oriented, geometric projection, icon based and hierarchical

References:

1. Building the Data Warehouse, W.H.Inmon, John Wiley & Sons.
2. Developing the Data Warehouse, W.H.Inmon, C.Kelly, John Wiley & Sons.
3. Managing the Data Warehouse, W.H.Inmon, C.L.Gassey, John Wiley & Sons.
4. Advances in knowledge discovery & Data Mining, Fayyad, Usama M. et. al.,MIT Press.
5. Pang-Ning Tan, Michael Steinbach, Vipin Kumar: Introduction to Data Mining, Pearson,
6. Jiawei Han, Micheline Kamber, Jian Pei: Data Mining -Concepts and Techniques, 3rd Edition, Morgan Kaufmann Publisher, 2012

Pedagogy:

The Instructor is expected to use leading pedagogical approaches in the class room situation, research-based methodology, innovative instructional methods, extensive use of technology in the class room, online modules of MOOCS, and comprehensive assessment practices to strengthen teaching efforts and improve student learning outcomes.

The Instructor of class will engage in a combination of academic reading, analyzing case studies, preparing the weekly assigned readings and exercises, encouraging in class discussions, and live project based learning.

Case/Class Discussion Assignments:

Students will work in groups of up to four to prepare a brief write-up due before the start of each class covering the case study or class material to be discussed in the next session. Questions may include a quantitative analysis of the problem facing the decision-maker in the case.

Class Participation:

Attendance will be taken at each class. Class participation is scored for each student for each class.

Scheme of Examination

- English will be the medium of instruction and examination.
- Written Examinations will be conducted at the end of each Semester as per the Academic Calendar notified in advance
- Each course will carry 100 marks of which 30 marks shall be reserved for the continuous internal assessment and the remaining 70 marks for written examination to be held at the end of each semester.
- The duration of written examination for each paper shall be three hours.
- The minimum marks for passing the examination for each semester shall be 40% in aggregate as well as a minimum of 40% marks in the semester-end examination in each paper.
- A minimum of 75% of classroom attendance is required in each subject

Instructions for the Paper Setter

The question paper will consist of three sections: A, B & C. Sections A & B will have four questions each from the respective sections of the syllabus carrying 12 marks for each question. Section C will have 11 short-answer type questions carrying total of 22 marks, which will cover the entire syllabus uniformly.

Instructions for the Candidates

1. Candidates are required to attempt two questions each from the sections A & B of the question paper and the entire section C.
2. Use of non-programmable scientific calculator is allowed.

MDSM2203P: Major Project

This laboratory course will mainly comprise of exercise based on subject MDSM2202T.

The evaluation done continuous assessment as well as University Examination will be done on the basis of following components:

1.	Software Demo	:	50% of the total marks allotted
2.	Project Report	:	25% of the total marks allotted
3.	Viva-Voce	:	25% of the total marks allotted

Guidelines for the Major Project:

1. The students are required to undertake a major software development project during the third semester of M. Sc. (Computer Science with Specialization in Data Science) Course along with the regular classes. The project should be done preferably using the programming languages taught in the three semesters of the course.
2. The students will complete systems analysis, design, coding and testing of the software project assigned to them by the project guide. The students are required to complete the minor project in the Department/institute/college given by the concerned teacher of the Department. No outside training/ project work will be allowed.
3. Joint projects may be allowed and joint project reports will also be accepted, with the permission of the teacher concerned. However, the students should highlight their individual contributions in a joint project. The quantum of individual contribution of particular students in joint projects should be such which can be accepted as equivalent to individual major project. The same must also be reflected in joint reports.
4. Each student should submit one project report of his/her project to the concerned guide, as per the format decided by the Department/institute/college.
5. The students are required to give live demo of the software developed by them and there will be viva-voce of the students during the end-semester practical examination.