**MANIPAL UNIVERSITY JAIPUR**



School of Computer Science & Engineering

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE & MACHINE LEARNING**

Course Hand-out

AI3132 Foundations of Data Science Lab [0 0 2 1]

Session: July 2023 – Nov 2023 |

Course Instructors: Ms. Shubh Lakshmi Agrwal, Dr. YP Singh, Dr. Ajay Garg, Mr. Harish Sharma

Class: Core Lab

1. **Introduction:** This lab is offered by Department of Artificial Engineering & Machine Learning for fifth semester students. The core aim of this course is to implement the basic concept of data science and its related algorithms. The substantial aim of this lab is to understand the basic underlying concepts of data science using visualization tool and exploratory data analysis algorithms.
2. **Course Outcomes:** At the end of the course, students will be able to

[AI3132.0]. To prepare clean data by data analytics tool for modelling.

[AI3132.1]. To perform exploratory data analysis on dataset for real world applications.

[AI3132.2]. To predict the class / cluster for the real-world dataset by applying data science modelling algorithms.

[AI3132.3]. To compare and analyse the performance of classification / clustering process applied.

[AI3132.4]. To prepare the data visualization in a meaningful and concise manner for better understanding using various tools.

1. **PROGRAM OUTCOMES AND PROGRAM SPECIFIC OUTCOMES**

**[PO.1] Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

**[PO.2] Problem analysis**: Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

**[PO.3] Design/development of solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

**[PO.4] Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

**[PO.5] Modern tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

**[PO.6] The engineer and society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

**[PO.7] Environment and sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

**[PO.8] Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practices.

**[PO.9] Individual and teamwork**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings**.**

**[PO.10] Communication**: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

**[PO.11] Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one’s own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

**[PO.12] Life-long learning**: Recognize the need for and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

**[PSO1]:** Graduates will be able to examine the applications of Artificial Intelligence and Machine Learning in real-life problems.

**[PSO2]:** Graduates will be able to design and implement intelligent systems for multidisciplinary problems.

1. **Assessment Plan:**

|  |  |  |
| --- | --- | --- |
| **Criteria** | **Description** | **Maximum Marks** |
| Internal Assessment | Continuous Assessment  File (10)  Attendance (5)  Viva (20)  Program execution (10)  Project + report (15) | 60 |
| End Term Exam | End Term Exam Assessment | 40 |
|  | Total | 100 |
| Attendance  (Formative) | A minimum of 75% Attendance is required to be maintained by a student to be qualified for taking the End Semester examination. The allowance of 25% includes all types of leaves including medical leaves. | |
| Make up Assignments  (Formative) | Students who miss a class will have to report to the teacher about their absence. A makeup assignment on the topic taught on the day of absence will be given, which has to be submitted within a week from the date of absence. No extensions will be given to this. The attendance for that day of absence will be marked blank, so that the student is not accounted for absence. These assignments are limited to a maximum of 5 throughout the entire semester. | |
| Homework/ Home Assignment/ Activity Assignment  (Formative) | There are situations where a student may have to work in home, especially before a flipped classroom. Although these works are not graded with marks. However, a student is expected to participate and perform these assignments with full zeal since the activity/ flipped classroom/ project participation by a student will be assessed and marks will be awarded. | |

1. **Syllabus:**

**AI3132: Foundation of Data Science LAB [0 0 2 1]**

Data Analytics & Visualization using various tools such as R, MATLAB, WEKA, RapidMiner. Experiments related to refinement of data and untimely the dataset will be elaborated in lab manual. Implement concepts of data pre-processing, data types, privacy and confidentiality, samples vs. population, comparative statistics, statistical inference, Association, and clustering. Creating Visual Representations, descriptive, inferential statistics, univariate and multivariate analysis. Grouping – Cluster Analysis, distance measures, partitioning, hierarchical, density-based methods. Market Basket Analysis Association Analysis Suggested tools are MS Excel, Power BI, Tableau. MS Excel Pivot Tables and charts, Visualization of Groups, Volumetric Data, Case Studies in Various Perception Techniques.

**References:**

1. Glenn J. Myatt., Making sense of Data: A practical Guide to Exploratory Data Analysis and Data Mining, (2e), Wiley Press, 2006.
2. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, “An Introduction to Statistical Learning with Applications in R”, Springer Science, 2017
3. Tufte. The Visual Display of Quantitative Information, (2e), Graphics Press, 2007.
4. Cole Nussbaumer Knaflic, Storytelling with Data: A Data Visualization Guide for Business Professionals, (1e), John Wiley and Sons, 2015.
5. Brian R. Hunt, Ronald L. Lipsman, Jonathan M. Rosenberg, Kevin R. Coombes, John E. Osborn, Garrett J. Stuck , Guide to MATLAB: For Beginners and Experienced Users,(2e), Cambridge University Press, 2011.
6. **List of Experiments:**
7. Implementation of important libraries of Python for data science.
8. Categorization and implementation of different data formats for data analysis.
9. Find missing values with estimation and their categorization in the various dataset.
10. Outlier estimation and data cleaning using quartile and visualization methods.
11. Implementation of data transformation using various methods.
12. Perform hypothesis testing on dataset (T-test, z-test, Annova, Chi square Test).
13. To execute data normalization and segmentation methods and their different variants.
14. To implement the K-mean clustering method using different datasets.
15. To implement Hierarchical clustering methods using different datasets.
16. To perform Linear Regression on given datasets and estimate the values.
17. To perform Logistic Regression on given datasets and estimate the classes.
18. Data analysis using MS Excel, Pivot Tables and Charts.
19. To demonstrate 10-15 Data visualization using tableau for given dataset.
20. Implementation of a case study project in Data analytics / Data Science including Data Cleaning, Normalization, Hypothesis testing, Data visualization.