MANIPAL UNIVERSITY JAIPUR
School of Information Security and Data Science
DEPARTMENT OF DATA SCIENCE AND ENGINEERING
Course Hand-out

Data Mining Lab | DSE2130 | Credits | 0 0 2 1

Session: July – Nov 24 | Faculty: Mr. Srikanta Pradhan, Dr. Sudhir Sharma, Ms. Shatabdi Basu, Mr. Ashish Pandey, Dr. Rekha Chaturvedi, Dr. Neeraj Kumar Verma, Ms.Shweta Gangrade

A. **Introduction**: To build competence/aptitude of the student group familiarity with the course, and be able to apply Data Mining to a variety of tasks .

B. **Course Outcomes:** At the end of the course, students will be able to:

CO	Statement	Cognitive Level
[DSE2130.1]	To implement data structure and different operations in data types like vectors, lists, matrices, arrays.	Remembering
[DSE2130.2]	To implement and demonstrate through charts and graphs using for data analysis and PCA.	understanding
[DSE2130.3]	To implement and demonstrate time series, Precision, Recall, TP rate, FP rate and F-measure and Apriori algorithm	applying
[DSE2130.4]	To implement and demonstrate the classification and regression concepts.	analyzing
[DSE2130.5]	To implement and demonstrate the performance and evaluation of different data analysis algorithms.	analyzing

C. **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 modelling 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 team work: 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

[PSO.1]. Understand the role of mathematics, statistics, and AI techniques in the field of data science & engineering.

[PSO.2]. Apply the acquired knowledge and expertise to perform data analytics tasks for multidimensional data sets.

[PSO.3]. Develop effective and scalable industrial solutions for real-world socio-economic problems using data analytics tools and techniques.

D. **Assessment Plan**

Criteria	Description	Maximum Marks
Internal Assessment (Summative)	Continuous Assessments	60
Exam (Summative)	Exam (Small Project/Exam)	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 lab 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 2 throughout the entire semester.

E. Course Content

Explore the various datasets, perform data preprocessing tasks and demonstrate performing association **Rule mining on data sets:** Attribute Selection, Handling Missing Values, Discretization, Eliminating Outliers. Market Basket analysis, Apriori Algorithm, FP Growth Algorithm. **Demonstrate performing classification on data sets:** Decision Tree, Naïve Bayesian Classifier, K- Nearest Neighbour Classifier, SVM. **Demonstrate performing clustering of data sets:** K-Means, Hierarchical Clustering, DBSCAN. Performance evaluation using confusion matrix and related case studies.

Textbook

1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, 2nd edition, Pearson education, 2018.

References books

1. Jlawei Han & Micheline Kamber, Data Mining, "Concepts and Techniques", 3rd edition, Morgan Kaufmann Publishers, 2012.

2. Margaret H Dunham, Data Mining Introductory and advanced topics, 6th edition, Pearson Education, 2009

3. Arun K Pujari, Data Mining Techniques, 1st edition, University Press, 2005.

F. Lecture Plan

A.

Class Number	Topics	Session Outcome	Mode of Delivery	Corresponding Course Outcome	Mode of Assessing the Outcome
1.	Implement data structures and different operators Python.	Understanding Basic Data Structures and operations in Python.	Lab	DSE2130.1	Internal Evaluation Assignments External Evaluation
2.	To read and write data files and present different graphs using Python libraries.	Learning the methods of file handling, data manipulation and data visualization using Python.	Lab	DSE2130.2	Internal Evaluation Assignments External Evaluation
3.	Perform data analysis on time series and use metrics such as Precision, Recall, True Positive Rate (TPR), False Positive Rate (FPR), and F-measure.	Proficient in analyzing time series data, applying metrics for data analysis.	Lab	DSE2130.2 DSE2130.3	Internal Evaluation Assignments External Evaluation
4.	To perform statistical operations on the dataset and data analysis. Perform initial preprocessing and data reduction using Principal Component Analysis (PCA).	Understanding of various data preprocessing and analysis methods.	Lab	DSE2130.2 DSE2130.3	Internal Evaluation Assignments External Evaluation

5.	To perform association Rule Mining based data analysis and visualization on a dataset.	Understanding of association rule mining and discovering interesting relationships between variables of dataset.	Lab	DSE2130.2 DSE2130.3	Internal Evaluation Assignments External Evaluation
6.	To perform apriori Rule Mining based data analysis on a dataset.	Understanding and reducing the computational complexity of finding frequent itemsets.	Lab	DSE2130.2 DSE2130.3	Internal Evaluation Assignments External Evaluation
7.	To perform Decision Tree classification technique using Python	Learning and mimic human decision-making by splitting data into subsets and creating a tree-like model of decisions.	Lab	DSE2130.2 DSE2130.4	Internal Evaluation Assignments External Evaluation
8.	To perform K-nearest neighbor classification technique using Python.	Understanding classification by finding the most common class among the K-nearest neighbors of a data point,	Lab, case study	DSE2130.2 DSE2130.4	Internal Evaluation Assignments External Evaluation
9.	To perform Bayes Classification method using Python.	Understanding classification tasks, including the principle of updating probabilities based on new evidence.	Lab, Demons tration	DSE2130.2 DSE2130.4	Internal Evaluation Assignments External Evaluation
10.	To demonstrate K-means clustering method using Python	Learning partition a dataset into K clusters where each data point belongs to the cluster with the nearest centroid.	Lab	DSE2130.2 DSE2130.5	Internal Evaluation Assignments External Evaluation
11.	To perform hierarchical clustering method using Python.	Learn to determine the optimal number of clusters by cutting the dendrogram at different levels and evaluating cluster quality using metrics	Lab, case study	DSE2130.2 DSE2130.5	Internal Evaluation Assignments External Evaluation
12.	To perform random forest clustering method using Python.	Learn to apply Random Forest clustering to complex datasets, derive insights, and make informed decisions based on the clustering analysis	Lab, Demons tration	DSE2130.2 DSE2130.5	Internal Evaluation Assignments External Evaluation

H. Course Articulation Matrix: (Mapping of COs with POs & PSOs)

CO	STATEMENT	CORRELATION WITH PROGRAM OUTCOMES (POs)												CORRELATION WITH PROGRAM SPECIFIC OUTCOMES (PSOs)		
		PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3
[DSE2130.1]	To implement data structure and different operations in data types like vectors, lists, matrices, arrays.	3	1	1		2							3		3	
[DSE2130.2]	To implement and demonstrate through charts and graphs using for data analysis and PCA.	2	1	1		2							3			
[DSE2130.3]	To implement and demonstrate time series, Precision, Recall, TP rate, FP rate and F-measure and Apriori algorithm.	3	3	3	2								1	2		
[DSE2130.4]	To implement and demonstrate the classification and regression concepts.	2	3	3	3		3						1	3	3	3
[DSE2130.5]	To implement and demonstrate the performance and evaluation of data analysis algorithms.	2	3	3	3		3						1	3	3	3

1-Low Correlation; 2- Moderate Correlation; 3- Substantial Correlation