MANIPAL UNIVERSITY JAIPUR

School of Information Technology

Department of Data Science Engineering
Course Hand-out

Data Mining | **DS 2103**| 4 Credits | 3 1 0 4

Session: Jul-Nov 2025

Faculty: Dr. Sumit Srivastava |Dr. Akhilesh Kumar Sharma |Dr. Sudhir Sharma |Mr. Srikanta Pradhan |

Ms. Shatabdi Basu | Class: Department Elective Course

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
[DSE2103.1]. Remember and identify fundamental terms and concepts in data mining.
[DSE2103.2]. Explain the significance and application of data mining methods and visualization techniques
[DSE2103.3]. Apply data mining algorithms to build predictive models.
[DSE2103.4]. Evaluate and compare the performance of different data mining techniques by analysing their effectiveness and efficiency in various data mining tasks
[DSE2103.5]. Design and implement data mining techniques to real-world data sets

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)	Sessional Exam 1	30
	Research Papers on the real time data sets/Quiz/Assignment	20
	NPTEL/SWAYAM course https://onlinecourses.nptel.ac.in/noc21_cs06/preview	10
End Term Exam (Summative)	End Term Exam	40
	Total	100
Attendance (Formative)	A minimum of 75% Attendance is required to be maintained by a student to be qualified for taking up the End Semester examination. The allowance of 25% includes all types of leaves including medical leaves.	
Make up Assignments (Formative)	Students who misses a class will have to report to the teacher about the 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 on this. The attendance for that particular day of absence will be marked blank, so that the student is not accounted for absence. These assignments are limited to a maximum of 15 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 participation by a student will be assessed and marks will be awarded.	

DS2103: Data Mining [L-3 T-1 0 4]

Course Contents

Data preprocessing: Data cleaning, Data transformation, Data reduction, Discretization and generating concept hierarchies, Visualization techniques. Mining Frequent Patterns DataMart, OLTP, OLAP, Transformation and ETL operation, Statistics, Data Mining Goals, KDD Process, Data Mining Techniques, Applications, Issues. **Association Rule Mining**: Associations and Correlations, Basic Concepts and Methods, Market Basket Analysis, Frequent Itemset, Closed Itemset, Association rules, Apriori algorithm, Generating association rules, **Data Mining Algorithms:** Decision Tree, Classification and Prediction, Clustering. **Tree Construction Principle:** Best Split, Splitting Indices, Splitting Criteria, **Decision Tree Construction Algorithms:** CART, ID3, C4.5, Decision Tree Construction with Pre-sorting, Random Forest, Approximate Methods. **Classification and Prediction:** Naïve Bayes, Backpropagation, Support Vector Machine, Lazy Learner, K-Nearest Neighbour. **Clustering:** Cluster Analysis, Partitioning Methods, Hierarchical Clustering, Density based Clustering. **Outlier Detection:** Definition, Outlier Analysis, Outlier Detection Method. Case Study related to advance data mining techniques.

Textbook:

1. Han, J., Pei, J., & Tong, H. (2022). *Data mining: Concepts and Techniques*. Morgan Kaufmann.
2. L. Bing Web Data Mining Springer-Verlag, 2017.

References books:

1. P. Ponniah, Data Warehousing, (2e), Wiley India Pvt. Ltd., 2011
2. Berson and S. J. Smith, Data Warehousing, Data Mining & OLAP, (10e), Tata McGraw – Hill, 2007.

E.   Lecture Plan:

Lec No.	Unit Name	Topics Not more than 250 words	Corresponding CO	Mode of Assessing CO
1	Data preprocessing	Introduction of Data Mining and Course Plan Discussion	DS2103.1	Mid-Term, Case Studies
2	Data preprocessing	Data cleaning, Data transformation,	DS2103.1	Mid-Term, Case Studies
3	Data preprocessing	Data reduction,	DS2103.1	Mid-Term, Case Studies
4	Data preprocessing	Discretization and generating concept hierarchies	DS2103.1	Mid-Term, Case Studies
5	Data preprocessing	Visualization techniques.	DS2103.1	Mid-Term, Case Studies
6	Data preprocessing	Mining Frequent Patterns	DS2103.1	Mid-Term, Case Studies
7	Data preprocessing	DataMart, OLTP, OLAP	DS2103.1	Mid-Term, Case Studies
8	Data preprocessing	Transformation and ETL operation	DS2103.1	Mid-Term, Case Studies
9	Data preprocessing	Statistics, Data Mining Goals	DS2103.1	Mid-Term, Case Studies
10	Data preprocessing	KDD Process	DS2103.1	Mid-Term, Case Studies
11	Association Rule Mining	Associations and Correlations: Basic Concepts and Methods	DS2103.2	Mid-Term, Case Studies
12	Association Rule Mining	Market Basket Analysis	DS2103.2	Mid-Term, Case Studies
13	Association Rule Mining	Frequent Itemset	DS2103.2	Mid-Term, Case Studies
14	Association Rule Mining	Closed Itemset, Association rules	DS2103.2	Mid-Term, Case Studies

15	Association Rule Mining	Apriori algorithm	DS2103.2	Mid-Term, Case Studies
16	Association Rule Mining	Generating association rules	DS2103.2	Mid-Term, Case Studies
17	Data Mining Algorithms	Decision Tree	DS2103.2	Mid-Term, Case Studies
18	Data Mining Algorithms	Classification and Prediction, Clustering	DS2103.2, DS2103.3	Mid-Term, Case Studies
19	Data Mining Algorithms	Tree Construction Principle: Best Split	DS2103.2, DS2103.3	Mid-Term, Case Studies
20	Data Mining Algorithms	Splitting Indices, Splitting Criteria	DS2103.2, DS2103.3	Mid-Term, Case Studies
21	Data Mining Algorithms	Decision Tree Construction Algorithms: CART	DS2103.2, DS2103.3	Mid-Term, Case Studies
22	Data Mining Algorithms	Decision Tree Construction Algorithms: CART	DS2103.2, DS2103.3	Mid-Term, Case Studies
23	Data Mining Algorithms	Decision Tree Construction Algorithms: ID3	DS2103.2, DS2103.3	Mid-Term, Case Studies
24	Data Mining Algorithms	Decision Tree Construction Algorithms: C4.5	DS2103.2, DS2103.3	Mid-Term, Case Studies
25	Data Mining Algorithms	Decision Tree Construction with Pre-sorting	DS2103.2, DS2103.3	Mid-Term, Case Studies
26	Data Mining Algorithms	Random Forest, Approximate Methods	DS2103.2, DS2103.3	Mid-Term, Case Studies
27	Classification and Prediction	Naïve Bayes	DS2103.2, DS2103.3	Case Studies & Coding Assignment, End Term
28	Classification and Prediction	Naïve Bayes	DS2103.2, DS2103.3	Case Studies & Coding Assignment, End Term
29	Classification and Prediction	Backpropagation	DS2103.2, DS2103.3	Case Studies & Coding Assignment, End Term
30	Classification and Prediction	Backpropagation	DS2103.2, DS2103.3	Case Studies & Coding Assignment, End Term
31	Classification and Prediction	Support Vector Machine	DS2103.2, DS2103.3	Case Studies & Coding Assignment, End Term
32	Classification and Prediction	Support Vector Machine	DS2103.2, DS2103.3, DS2103.4	Case Studies & Coding Assignment, End Term
33	Classification and Prediction	Lazy Learner	DS2103.2, DS2103.3, DS2103.4	Case Studies & Coding Assignment, End Term
34	Classification and Prediction	Lazy Learner	DS2103.2, DS2103.3, DS2103.4	Case Studies & Coding Assignment, End Term
35	Classification and Prediction	K-Nearest Neighbour	DS2103.2, DS2103.3, DS2103.4	Case Studies & Coding Assignment, End Term
36	Classification and Prediction	K-Nearest Neighbour	DS2103.2, DS2103.3, DS2103.4	Case Studies & Coding Assignment, End Term
37	Clustering	Cluster Analysis, Partitioning Methods	DS2103.2, DS2103.3, DS2103.4	Case Studies & Coding Assignment, End Term
38	Clustering	Cluster Analysis, Partitioning Methods	DS2103.2, DS2103.3, DS2103.4	Case Studies & Coding Assignment, End Term
39	Clustering	Hierarchical Clustering	DS2103.2, DS2103.3, DS2103.4	Case Studies & Coding Assignment, End Term
40	Clustering	Hierarchical Clustering	DS2103.2, DS2103.3, DS2103.4	Case Studies & Coding Assignment, End Term
41	Clustering	Density based Clustering	DS2103.2, DS2103.3, DS2103.4	Case Studies & Coding Assignment, End Term
42	Outlier Detection	Density based Clustering	DS2103.5	Case Studies & Coding Assignment, End Term
43	Outlier Detection	Definition, Outlier Analysis	DS2103.5	Case Studies & Coding Assignment, End Term
44	Outlier Detection	Outlier Detection Method	DS2103.5	Case Studies & Coding Assignment, End Term
45	Outlier Detection	Case Study related to advance data mining techniques.	DS2103.1DS2103.2, DS2103.3, DS2103.4, DS2103.5	Case Studies & Coding Assignment, End Term
46	Outlier Detection	Revision Classes-1	DS2103.1DS2103.2, DS2103.3, DS2103.4, DS2103.5	Case Studies & Coding Assignment, End Term
47	Outlier Detection	Revision Classes-2	DS2103.1DS2103.2, DS2103.3, DS2103.4, DS2103.5	Case Studies & Coding Assignment, End Term
48	Outlier Detection	Revision Classes-3	DS2103.1DS2103.2, DS2103.3, DS2103.4, DS2103.5	Case Studies & Coding Assignment, End Term

F. Course Articulation Matrix: (Mapping of COs with POs)

CO	STATEMENT	CORRELATION WITH PROGRAM OUTCOMES												CORRELATION WITH PROGRAM SPECIFIC OUTCOMES		
		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
[DSE2103.1].	Remember and identify fundamental terms and concepts in data mining	2	2	2	2	2	1	1	1					1	1	
[DSE2103.2].	Explain the significance and application of data mining methods and visualization techniques	2	2	2	2	1	1	1	1					1	1	
[DSE2103.3].	Apply data mining algorithms to build predictive models	2	2	2	2	1	1		1		1	1		1	1	1
[DSE2103.4].	Evaluate and compare the performance of different data mining techniques by analysing their effectiveness in various data mining tasks					1	1	1	1	1		1	1	2	2	2
[DSE2103.5].	Design and implement data mining techniques to real-world data sets		2	1	1	2	2	1	1	2			1	2	2	2

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

G. Course Outcome Attainment Level Matrix:

CO	STATEMENT	CORRELATION WITH PROGRAM OUTCOMES												CORRELATION WITH PROGRAM SPECIFIC OUTCOMES		
		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
[DSE2103.6].	Remember and identify fundamental terms and concepts in data mining															
[DSE2103.7].	Explain the significance and application of data mining methods and visualization techniques															
[DSE2103.8].	Apply data mining algorithms to build predictive models															
[DSE2103.9].	Evaluate and compare the performance of different data mining techniques by analysing their effectiveness in various data mining tasks															
[DSE2103.10].	Design and implement data mining techniques to real-world data sets															

0-No Attainment; 1- Low Attainment; 2- Moderate Attainment; 3- Substantial Attainment