AUTUMN SEMESTER 2024

DATA MINING AND DATA WAREHOUSING CS30013
Course Handout

1. **Course Code: CS30013**
2. **Course title: DATA MINING AND DATA WAREHOUSING**
3. **Course Faculty: Dr. Chittaranjan Pradhan, Associate Professor**
4. **Contact Details: 9861397865, F-108, Block-A, Campus-15**
   Availability: Tue (5PM-6PM), Wed (5PM-6PM), Thu (5PM-6PM)
5. **Contact hours per week : 3**
6. **Credit : 3**
7. **Time Table:**

	DMDW_CS4	DMDW_CS13
Tue	B-202 (3:00 - 4:00)	B-301 (4:00-5.00)
Wed	B-202 (1:00 - 2:00)	B-301 (11:00 -1 2:00)
Fri	B-202 (11:00 -1 2:00)	B-301 (1:00 - 2:00)

8. **Course Objective(s):**
   - To understand the basic principles, concepts, applications of data mining and data mining tools.
   - To know the kinds of patterns discovered by association rule mining algorithms
   - To understand various classification and prediction algorithms
   - To be able to apply data mining techniques for solving Societal issues.

9. **Course (learning) outcomes:** At the end of the course, the students will be able to:

	Course Outcomes
CO1	Compile the basic principles, concepts, applications of data mining and familiar with mathematical foundations of data mining tools.
CO2	Interpret the fundamental concepts, benefits, operations associated with data warehousing and develop data warehousing models.
CO3	Evaluate the kinds of patterns discovered by association rule mining algorithms.
CO4	Analyze various classification and prediction algorithms for model designing.
CO5	Apply various clustering algorithms to solve the real problems.
CO6	Demonstrate of various Data Mining techniques for solving Societal issues.

10. **Text books**

T1. J. Han and M. Kamber, "Data Mining: Concepts and Techniques", 4th Edition, Morgan Kaufman, 2015.

11. Reference books

R1. M. H. Dunham, "Data Mining: Introductory and Advanced Topics", Pearson Education, 2006.
R2. I H. Witten and E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques," Morgan Kaufmann, 2000.
R3. D. Hand, H. Mannila and P. Smyth. Principles of Data Mining. Prentice-Hall. 2001.

Day wise lesson plan

Module, Hrs	Topics/Coverage	Days
Introduction to data mining systems: (9 hrs)	Introduction to Data Mining, brief introduction to evolution of database system technology, data mining on kind of data types of data needs mining.	1
	Knowledge discovery process, data mining techniques,	2
	Major issues and applications of data mining, data objects and attribute types.	3
	statistical description of data: Descriptive data summarization, measuring central tendency, dispersion of data and data visualization	4
	Data preprocessing: Data Cleaning: Handling Noisy data, and Missing Values,	5
	data integration: entity identification problem, redundancy, correlation analysis (Pearson's product moment coefficient and chi-square test),	6-7
	data reduction, data transformation, Normalization,	8
	basic of sampling,data discretization and concept hierarchy generation	9
Data Warehousing And Online Analytical Processing: (6 hrs.)	Basic concepts, data warehousing architecture,	10
	Multidimensional data model: Data cubes, stars, Snowflakes and Fact Constellations Schemas	11-12
	data warehouse schema for decision support, building a data warehouse.	13-14
	Concept hierarchies, characteristics of OLAP systems, OLTP vs OLAP	15
Frequent Pattern Analysis: (6 hrs.)	Mining frequent patterns, market basket analysis: The Apriori algorithm: Finding frequent item-sets using candidate generation, generating association rules from frequent item-sets.	16-18

	Improving the efficiently of Apriori, mining frequent item-sets without candidate generation, measuring the quality of rules, association mining to correlation analysis.	19-21
	M i d s e m e s t e r e x a m i n a t i o n	
Classification And Prediction: (10 hrs)	Issues regarding classification and prediction, classification by decision tree introduction (ID3 algorithm),	22-23
	NaiveBayesian classification, rule based classification,	24-25
	classification by back propagation,	26-27
	support vector machines,	28
	lazy learners (K-NN) , other classification methods (Genetic Algorithms, Rough set Approach, Fuzzy Set Approaches)	29-30
	Prediction: Simple linear regression, accuracy and error measures, evaluating the accuracy of a classifier or predictor.	31-32
Clustering (7hrs)	Types of data in cluster analysis, data similarity and dissimilarity measures.	33-34
	A categorization of major clustering methods, partitioning methods: K-Means, K-Medoids.	35-36
	Hierarchical methods: Agglomerative and Divisive hierarchical clustering,	37-38
	Density based methods, Grid-based methods, Model-based clustering methods, Clustering high dimensional data.	39-40
	Total	40

12. Assessment components:

Exam		Marks
	End Semester	50
Internal	Mid Semester	20
	Activity	30
	Total	**100**

13. Activity Calendar

Activity	Type	Focus	Tentative Dates	Marks	CO
1	Assignment/Quiz	Critical Thinking	30-07-24	5	CO1
2	Group Activity & Presentation	Creation, Reflections	20-08-24	5	CO2
3	Assignment/Quiz	Quiz	6-09-24	5	CO3
4	Assignment/Quiz	Critical Thinking	18-10-24	5	CO4

| 5 | Assignment/Quiz | Critical Thinking | 06-11-24 | 5 | CO5 |
| 6 | Assignment/Quiz | Quiz | 14-11-24 | 5 | CO6 |

14. Links to e-resources (NPTEL, YouTube, Swayam, Virtual lab etc)
https://nptel.ac.in/courses/106105182

Data Mining and Data Warehousing:

PO	1	2	3	4	5	6	7	8	9	10	11	12	PSO1	PSO2	PSO3
CO1	√											√			
CO2	√			√									√		
CO3													√	√	
CO4		√	√	√		√					√				√
CO5		√	√			√							√	√	√
CO6			√	√									√		