Shyam Lal College (University ofDelhi)
Shahdara- Delhi 110032
www.slc.du.ac.in

Lesson Plan (Discipline Specific Elective - (DSE), August to November 2024)

Name of Teacher	Mr. Parveen Kumar		Department	Computer Science
Course	B.Sc. (Physical Science)		Semester	V
Paper	Data Mining and Knowledge Discovery		Academic Year	2024

Learning Objectives

This course introduces data mining techniques and enables students to apply these techniques on real-life datasets. The course focuses on three main data mining techniques: Classification, Clustering and Association Rule Mining tasks.

Learning Outcomes

On successful completion of this course, the student will be able to:

1. Pre-process the data, and perform cleaning and transformation.
2. Apply suitable classification algorithm to train the classifier and evaluate its performance.
3. Apply appropriate clustering algorithm to cluster data and evaluate clustering quality 4. Use association rule mining algorithms and generate frequent item-sets and association rules

Lesson Plan

Week	Unit	Topics	Reference Book	Chapter
Week 1–2	Unit 1 - Introduction	Need for data mining, Data mining tasks, Applications of data mining, Measures of similarity and dissimilarity, Supervised vs. unsupervised techniques.	1	1.1-1.4, 2.4.2, 2.4.3 (excluding properties)
Week 3–5	Unit 2 - Data collection and preparation	Measurement and data collection issues, Data aggregation, Sampling, Dimensionality reduction, Feature subset selection, Feature creation, Discretization and binarization, Variable transformation.	1	2.1,2.2, 2.3.1, 2.3.2, 2.3.3 (introduction), 2.3.4 (introduction), 2.3.5 (introduction), 2.3.6 (Binarization and Discretization of Continuous attributes), 2.3.
Week 6-10	Unit 3 - Clustering data	Basic concepts of clustering, Partitioning Methods: K-means algorithm, Hierarchical Methods: Agglomerative Hierarchical Clustering, Density-Based Methods: DBSCAN Algorithm, Strengths and weaknesses of different methods, Cluster evaluation.	1	5.2 (5.2.1-upto Data in Euclidean Space, 5.2.5), 5.3 (5.3.1, 5.3.2-Excluding Ward's and Centroid methods, 5.3.6), 5.4, 5.5(5.5.1,5.5.5,5.5.7)
Week 11-13	Unit4 – Classification	Preliminaries, Naive Bayes classifier, Nearest Neighbour classifier, Decision tree, Artificial Neural Network, overfitting, Confusion matrix, Evaluation metrics and Model evaluation	1	3 (up to 3.3.3), 3.4 (introduction) 3.6, 6.3, 6.4, 6.7 (introduction), 6.11(introduction, 6.11.2)

Week 14-16	**Unit5 - Ensemble Methods**	**Need for ensembles, Random Forest, Concept of Bagging and Boosting in ensembles.**	**1**	**6.10 (Excluding 6.10.3)**

References

Text Book:

1. Tan P.N., Steinbach M, Karpatne A. and Kumar V. Introduction to Data Mining, Second edition, Sixth Impression, Pearson, 2023.

Additional References:

1. Han J., Kamber M. and Pei J. Data Mining: Concepts and Techniques, 3rd edition, 2011, Morgan Kaufmann Publishers.

2. Zaki M. J. and Meira J. Jr. Data Mining and Machine Learning: Fundamental Concepts and Algorithms, 2nd edition, Cambridge University Press, 2020.

3. Aggarwal C. C. Data Mining: The Textbook, Springer, 2015 4. Dunham M. Data Mining: Introductory and Advanced Topics, Pearson, 2006.

Assignment and Class Test Schedule for Semester	Assignment to be allocated in week 5-6 and week 9-11. Class test to be held as per schedule during week 12-13

Practical Examination

For practical, datasets may be downloaded from :

1. https://archive.ics.uci.edu/datasets
2. https://www.kaggle.com/datasets?fileType=csv
3. https://data.gov.in/
4. https://ieee-dataport.org/datasets
5. Time Series Datasets (kaggle.com)

Suggested Practical Exercises

1. Apply data cleaning techniques on any dataset (e.g. Chronic Kidney Disease dataset from UCI repository). Techniques may include handling missing values, outliers and inconsistent values. Also, a set of validation rules may be specified for the particular dataset and validation checks performed.

2. Apply data pre-processing techniques such as standardization/normalization, transformation, aggregation, discretization / binarization, sampling etc. on any dataset

3. Apply simple K-means algorithm for clustering any dataset. Compare the performance of clusters by varying the algorithm parameters. For a given set of parameters, plot a line graph depicting MSE obtained after each iteration.

4. Perform partitioning, hierarchical, and density-based clustering algorithms on a downloaded dataset and evaluate the cluster quality by changing the algorithm's parameters.

5. Use Naive bayes, K-nearest, and Decision tree classification algorithms to build classifiers on any two datasets. Pre-process the datasets using techniques specified in Q2. Compare the Accuracy, Precision, Recall and F1 measure reported for each dataset using the abovementioned classifiers under the following situations:
   i. Using Holdout method (Random sampling):
      a) Training set = 80% Test set = 20%
      b) Training set = 66.6% (2/3rd of total), Test set = 33.3%
   ii. Using Cross-Validation:
      a) 10-fold
      b) Fold

6. Use the Decision Tree classification algorithm to construct a classifier on two datasets. Evaluate the classifier's performance by performing ten-fold cross validation. Compare the performance with that of:
    i. Bagging ensemble consisting of 3, 5, 7, 9 Decision tree classifiers
    ii. Adaboost ensemble consisting of 3, 5, 7, 9 Decision tree classifiers

Project: *Students should be promoted to take up one project on using dataset downloaded from any of the websites given above and the dataset verified by the teacher. Preprocessing steps and at least one data mining technique should be shown on the selected dataset. This will allow the students to have a practical knowledge of how to apply the various skills learnt in the subject for a single problem/project.*

Shyam Lal College (University ofDelhi)
Shahdara- Delhi 110032
www.slc.du.ac.in

Lesson Plan (Generic Elective), August to November 2024)

Name of Teacher	Mr. Parveen Kumar		Department	Computer Science
Course	B.Sc. (Physical Science)		Semester	THREE
Paper	Database Management Systems		Academic Year	2024

Learning Objectives

The course introduces the concepts of database management systems to students, focusing on basics such as the importance and significance of a database, data model, schema creation and normalization.

Learning Outcomes

On successful completion of this course, the student will be able to:

1. Describe the features of database management systems.
2. Differentiate between database systems and file systems.
3. Model an application's data requirements using conceptual modelling tools like ER diagrams and design database schemas based on the conceptual model.
4. Write queries in relational algebra / SQL.
5. Normalize a given database schema.

Lesson Plan

Week	Unit	Topics	Reference Book	Chapter
Week 1–2	Unit 1 – Introduction to Database	Purpose of database system, Characteristics of database approach, data models, database management system, database system architecture, three-schema architecture	1	1.1 - 1.3, 1.4 - 1.5, 1.6, 1.8
Week 3–4	Unit 1 – Introduction to Database	Components of DBMS, data independence, and file system approach vs database system approach.	1	2.1 - 2.2, 2.3.1, 2.4 -2.4.1, 2.6 2.1 - 2.2, 2.3.1, 2.4 -2.4.1, 2.6
Week 5-6	Unit 2 – Entity Relationship (ER) Modeling	Conceptual data modeling - motivation, entities, entity types, attributes, relationships, relationship types, constraints on relationship, Entity Relationship diagram notation.	1	3.1-3.7, 3.9.1
Week 7-8	Unit 3 – Relational Data Model	Update anomalies, Relsational Data Model - Concept of relations, schema-instance distinction, keys, relational integrity constraints, referential integrity and foreign keys, relational algebra operators and queries.	1	5, 8.1 - 8.3.2, 8.4 (except 8.4.3) 8.5
Week 9-12	Unit 4 - Structured Query Language (SQL)	Querying in SQL, DDL to create database and tables, table constraints, update database-update behaviours, DML, aggregation functions group by and having clauses, retrieve data from the database, generate and query views. Access and manipulate databases using ODBC. Basic Database administration SQL commands.	1	6.1-6.4, 7 - 7.1.8 (except 7.1.4), 7.3.1-7.3.2, 7.4
			2	2*(Pg. 48), 3*(Listing 3.4)
			3	11*(Pg. 356-357), 18*(Pg. 532-552)
Week 12-14	Unit 5 -Database Design	Mapping an Entity Relationship (ER) model to relational database, functional dependencies and Normal forms, 1NF, 2NF, 3NF and BCNF decompositions and desirable properties of them.	1	9.1, 14.1 - 14.5 (up to page 488), Chapter 15 15.1.1 (only Armstrong Axioms without proof and the Closure of X under F to find

				the primary key)
Week 15	**Unit 6 - Data Storage and Indexes**	**Need of file indexes, file organizations, index structures, single- and multi-level indexing, concurrent execution of transactions, ACID properties**	**1**	**16.5, 17.1.1, 17.2 (up to page 613), 20.1.1- 20.1.3, 20.3**

References

1. **Elmasri, R., Navathe, B. S. Fundamentals of Database Systems, 7th Edition, Pearson Education, 2015.**
2. **Krogh, J. W. MySQL Connector/Python Revealed: SQL and NoSQL Data Storage Using MySQL for Python Programmers, Apress, 2018.**
3. **Murach J. Murach's MySQL, 3rd edition, Pearson, 2019.**

Additional References

(i) **Ramakrishnan, R., Gehrke J. Database Management Systems, 3rd Edition, McGraw-Hill, 2014.**

(ii) **Silberschatz, A., Korth, H. F., Sudarshan S. Database System Concepts, 7th Edition,McGraw Hill, 2019.**

(iii) **Connolly, T. M., Begg, C. E. Database Systems: A Practical Approach to Design,Implementation, and Management, 6th edition, Pearson, 2019.**

Assignment and Class Test Schedule for Semester	Assignment to be allocated in week 5-6 and week 9-11. Class test to be held as per schedule during week 12-13

Practical Examination

Suggested Practical List

I. Create and use the following student-society database schema for a college toanswer the given (sample) queries using the standalone SQL editor.

STUDENT	**Roll No**	**StudentName**	**Course**	**DOB**
	Char(6)	Varchar(20)	Varchar(10)	Date

SOCIETY	**SocID**	**SocName**	**MentorName**	**TotalSeats**
	Char(6)	Varchar(20)	Varchar(15)	Unsigned int

ENROLLMENT	**Roll No**	**SID**	**DateOfEnrollment**
	Char(6)	Char(6)	Date

Here Rollno (ENROLLMENT) and SID (ENROLLMENT) are foreign keys.

1. Retrieve names of students enrolled in any society.
2. Retrieve all society names.
3. Retrieve students' names starting with the letter ‘A'.
4. Retrieve students' details studying in courses ‘computer science' or ‘chemistry'.

5. Retrieve students' names whose roll no either starts with ‘X' or ‘Z' and ends with ‘9'
6. Find society details with more than N TotalSeats where N is to be input by the user
7. Update society table for the mentor name of a specific society
8. Find society names in which more than five students have enrolled
9. Find the name of the youngest student enrolled in society ‘NSS'
10. Find the name of the most popular society (on the basis of enrolled students)
11. Find the name of two least popular societies (on the basis of enrolled students)
12. Find the students names who are not enrolled in any society.
13. Find the students names enrolled in at least two societies
14. Find society names in which maximum students are enrolled
15. Find names of all students who have enrolled in any society and society names in whichat least one student has enrolled
16. Find names of students who are enrolled in any of the three societies ‘Debating',‘Dancing' and ‘Sashakt'.
17. Find society names such that its mentor has a name with ‘Gupta' in it.
18. Find the society names in which the number of enrolled students is only 10% of itscapacity.
19. Display the vacant seats for each society.
20. Increment Total Seats of each society by 10%
21. Add the enrollment fees paid (‘yes'/'No') field in the enrollment table.
22. Update date of enrollment of society id ‘s1' to ‘2018-01-15', ‘s2' to the current date and ‘s3' to ‘2018-01-02'.
23. Create a view to keep track of society names with the total number of students enrolled init.
24. Find student names enrolled in all the societies.
25. Count the number of societies with more than 5 students enrolled in it.
26. Add column Mobile number in student table with default value ‘9999999999'.
27. Find the total number of students whose age is > 20 years.
28. Find names of students who were born in 2001 and are enrolled in at least one society.
29. Count all societies whose name starts with ‘S' and ends with ‘t' and at least 5 students are enrolled in the society.
30. Display the following information:

Society name	Mentor name	Total Capacity	Total Enrolled	Unfilled Seats