NATIONAL UNIVERSITY
of Computer & Emerging Sciences, Lahore

Department of Data Science

DS3002 – Data Mining
Spring 2024

Instructor Name: Eesha tur Razia Babar **TA Name:** TBA
Email address: razia.babar@lhr.nu.edu.pk **Email address:** TBA....@lhr.nu.edu.pk
Office Location/Number: CS New Building, Faculty offices
Office Hours: Monday: 11 to 1 PM

Course Information

Program: BS (DS) **Credit Hours:** 3 **Course Type:** Core for DS
Class Meeting Time: Section A: Mon & Wed 1-2:20 PM **Class Venue:** NB – 306

Course Updates:

All course material and course announcements will be made on Google classroom.

Course Description/Objectives/Goals:

- To develop the concepts of and the techniques in key data mining tasks.
- To provide hands-on experience with data mining using tools
- To encourage innovative and useful applications of data mining tasks

Course Learning Outcomes (CLOs):

At the end of the course students will be able to:	**Domain**	**BT* Level**
Understand basic concepts of data mining	C	
Apply data mining techniques to extract information from large data sets.	C	
Design and implement the efficient and scalable algorithms	C	
* BT= Bloom's Taxonomy, C=Cognitive domain, P=Psychomotor domain, A= Affective domain **Bloom's taxonomy Levels:** 1. Knowledge, 2. Comprehension, 3. Application, 4. Analysis, 5. Synthesis, 6. Evaluation		

Course Textbook

- Data Mining: Concepts and Techniques. By Jiawei Han and Micheline Kamber.
- Introduction to Data Mining. By P.-N. Tan, M. Steinbach and V. Kumar.

Additional references and books related to the course:

1. Data Mining: Practical Machine Learning Tools and Techniques, Ian H. Witten, Eibe Frank and Mark A. Hall, Third Edition, Morgan Kaufmann Publishers, 2011
2. Introduction to Data Mining, V. Tan et al. Addison-Wesley, 2009.
3. Tools: Weka, Matlab, Python, rapid miner

Tentative Weekly Schedule

COURSE OVERVIEW			
Weeks	Topics	Recommended Readings	Evaluation
Week 1-2	Overview of Data Mining Need and motivation; data mining process; data mining tasks and functionalities, interestingness measures. Overview of Data Preprocessing Data exploration and visualization; basic stats; data cleaning, data reduction	Ch. 1	
Week 3 -4	Mining Frequent Patterns and Associations Basic definitions, market basket analysis, Apriori algorithm, improving Apriori, FP growth, ECLAT, association rules mining	Ch. 6	
	MID 1		
Week 5-6	Classification Basic Concepts, Decision Tree Induction, ID3, C 4.5, CART, Bayes Classification Methods	Ch. 7	
Week 7-8	BBN, KNN, SVM, Random Forest, Model Evaluation, Evaluation methods (confusion matrix, ROC...)	Ch. 7	
Week 9-10	Cluster Analysis Similarity measures, partitioning methods: centroid based methods; K-Means, K-Mode, K Medoid.	Ch. 8	
	MIDTERM EXAM		
Week 11-12	Hierarchical methods, Agglomerative, divisive outlier/anomaly detection. Density based algo: DBSCAN	Ch. 8	
Week (13-14)	Text Mining (revision of concepts from IR) Text Preprocessing, Logistic regression for sentiment analysis, Vector Analysis, Singular value decomposition (SVD) for Latent Semantic Indexing, Word embeddings, Sequence learning models (RNN, GRU, LSTM), Applications in social media (Hate speech recognition), or Applications in sentiment classification		

Week (15)	Principal Component Analysis (PCA), dimensionality reduction, PCA for face recognition systems, PCA for image compression, PCA for analyzing text data, Application in social media behavior towards a celebrity	Ch. 10	
Week (16)	Recommendation system using SVD and Term project		

(Tentative) Grading Criteria:

Assignments 4 or more (**10%)** Quizzes: 4 or more **15 %)** Midterms **(25 %)**
Project **(10%)** Final Exam **40 %)**

Course Policies:

- Quizzes may be announced or surprised.
- All assignments and coursework must be done individually.
- Plagiarism in any work (Labs, Quizzes, Assignments, Midterms, and Final Exams) from any source, Internet, or a Student may result in **F** grade or deduction of absolute marks.
- No Late Submissions
- No Makeup Quizzes.
- 80% attendance is required for appearing in the Final exams.

Passing criteria:

The minimum requirement to pass this course is to obtain at least 50% marks under the application of CS department's grading policies. The grading scheme for this course is **Absolute**.