COURSE DESCRIPTION FORM: CS-4038: Data Mining

INSTITUTION FAST School of Computing, National University of Computer and Emerging Sciences, Islamabad Campus

PROGRAM(s) TO BE EVALUATED BS-CS: **Fall-2021**

Course Description

This course is about a rapidly emerging area called Data Mining. Due to the explosion of data, Data Mining has become inevitable in many real world application areas hence it has very bright prospects both in academia and industry. Data Mining itself is a highly diversified field of study i.e. it overlaps with a number of other existing subjects such as Databases, Artificial Intelligence, Statistics and Probability, Data mining digs out valuable non-trivial information from large multidimensional apparently unrelated databases (sets). It is currently regarded as the key element of a more general process called Knowledge Discovery that deals with extracting useful knowledge from raw data. The knowledge discovery process includes data selection, cleaning, coding, using different statistical, pattern recognition and machine learning techniques, and reporting and visualization of the generated structures. The course will cover all these issues and will illustrate the whole process by examples of practical applications. The students will use recent Data Mining software.

Course Code	CS-4038
Course Title	Data Mining
Credit Hours	3
Prerequisites by Course(s) and Topics	Database Systems (CS-2005)
Grading Policy	Relative as per University Grading Scheme All marks of evaluations will be finalized within one week after the announcement
Policy about missed assessment items in the course	Retake of missed assessment items (other than midterm/ final exam) will not be held. For a missed midterm/ final exam, an exam retake/ pre-take application along with necessary evidence are required to be submitted to the department secretary. The examination assessment and retake committee decides the exam retake/ pre-take cases.
Course Plagiarism Policy	Plagiarism in project or midterm/ final exam may result in F grade in the course. Plagiarism in an assignment will result in zero marks in the whole assignments category.
Assessment Instruments/ Evaluations with Weights (homework, quizzes, midterms, final, programming assignments, lab work, etc.)	100% Theory Breakdown of Course Work (Total): Midterm 1 20% Quizzes 5 10% Assignments 5 10% Project 1 10% Final 1 50%

Course Coordinator	Dr. Ejaz Ahmed
URL (if any)	Google class room code: **ap54caa**
Current Catalog Description	Data Pre-processing, Feature Selections & Classifications (ID3, C4.5 and CART), Rules based method, K-Nearest Neighbor (KNN), Artificial Neural Network (ANN) Classifiers, Association Mining methods (Apriori, FP Growth), Clustering, K-means, Linear Regression, Data Warehousing & BI, Web Mining
Textbook (or **Laboratory Manual** for Laboratory Courses)	**Data Mining Concepts and Techniques** by Jiawei Han and Micheline Kamber, 7th edition.
Reference Material	**Introduction to Data Mining with Case Studies** by G.K. Gupta **Advances in Knowledge Discovery and Data Mining** by Usama M. Fayyad et al.
Policy about missed assessment items in the course	Retake of missed assessment items (other than midterm/ final exam) will not be held. For a missed midterm/ final exam, an exam retake/ pre-take application along with necessary evidence are required to be submitted to the department secretary. The examination assessment and retake committee decides the exam retake/ pre-take cases.
Course Goals:	**A. Course Learning Outcomes (CLOs)** After completion of the course, the student shall be able to: 1. Demonstrable knowledge about the key concepts of Data Mining 2. Provide extensive hands-on experience in applying the problem solving concepts to real-world applications. 3. Identify novel opportunities for applying Data Mining 4. Assess the viability and feasibility of deploying a Data Mining solution. 5. Pre-processing relational data to produce data mining feature-based record set data 6. Learn role of Data Mining, DWH and BI Applications

B. Program Learning Outcomes:

PLO 1	Computing Knowledge	Apply knowledge of mathematics, natural sciences, computing fundamentals, and a computing specialization to the solution of complex computing problems.
PLO 2	Problem Analysis	Identify, formulate, research literature, and analyze complex computing problems, reaching substantiated conclusions using first principles of mathematics, natural sciences, and computing sciences.
PLO 3	Design/ Develop Solutions	Design solutions for complex computing problems and design systems, components, and processes that meet specified needs with appropriate consideration for public health and safety, cultural,

		societal, and environmental considerations.
PLO 4	Investigation & Experimentation	Conduct investigation of complex computing problems using research based knowledge and research based methods
PLO 5	Modern Tool Usage	Create, select, and apply appropriate techniques, resources and modern computing tools, including prediction and modelling for complex computing problems.
PLO 6	Society Responsibility	Apply reasoning informed by contextual knowledge to assess societal, health, safety, legal, and cultural issues relevant to context of complex computing problems.
PLO 7	Environment and Sustainability	Understand and evaluate sustainability and impact of professional computing work in the solution of complex computing problems
PLO 8	Ethics	Apply ethical principles and commit to professional ethics and responsibilities and norms of computing practice.
PLO 9	Individual and Team Work	Function effectively as an individual, and as a member or leader in diverse teams and in multi-disciplinary settings.
PLO 10	Communication	Communicate effectively on complex computing activities with the computing community and with society at large.
PLO 11	Project Management and Finance	Demonstrate knowledge and understanding of management principles and economic decision making and apply these to one's own work as a member or a team.
PLO 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 changes.

C. Mapping of CLOs on PLOs
(CLO: Course Learning Outcome, PLOs: Program Learning Outcomes

		PLOs											
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
CLOs	1	✔		✔							✔		
	2	✔		✔	✔	✔					✔		
	3	✔		✔	✔	✔					✔		
	4	✔	✔	✔	✔	✔				✔	✔		
	5			✔		✔							
	6									✔			✔

Topics Covered in the Course, with Number of Lectures on Each Topic (assume 15-week instruction and one-hour lectures)

Topics to be covered:

List of Topics	No. of Weeks	Contact Hours	CLO(s)
Introduction to Data Mining	**0.66**	**2**	**1**
Comparison of Data Mining with Databases and Data Warehousing	**1**	**3**	**2**
Data & Data Pre-Processing	**1**	**3**	**3**
Data Classification	**1.66**	**5**	**2, 4**
Overfitting	**0.66**	**2**	**2, 4**
Mapping from conceptual to logical model	**1**	**3**	**2, 4**
ANN	**1**	**3**	**2**
Rule based Models	**1.33**	**4**	**2**
KNN	**1**	**3**	**5**
FP	**1.66**	**5**	**5**
Clustering	**1**	**3**	**4**
Hierarchical Clustering	**1**	**3**	**4**
Linear Regressions	**1**	**3**	**4**
DWH, Mining and BI Applications	**0.66**	**3**	**6**
Total	**15**	**45**	

Laboratory Projects/Experiments Done in the Course

Applications developed using Python, Matlab and Weka

Programming Assignments Done in the Course

Python, Matlab, Weka, C#

Class Time Spent on (in credit hours)

Theory (%)	Problem Analysis (%)	Solution Design (%)	Social and Ethical Issues (%)
50	25	20	5

Oral and Written Communications

Every student is required to submit a final project written reports of typically 20 pages and oral presentations of typically 15 minutes duration.

Tentative course outline and lecture plan

Number of Lectures	Topics	Chapter
3	Information Systems and DBs <Contents Breakdown Structure>	1
3	Introduction to Data Mining , Data Pre-processing & BI systems	
3	Association Rule Mining <Apriori Algorithm>	3-4
3	Association Rule Mining <FP-Growth Algorithm>	6
3	Classification <Introduction, Decision Tress>	5
3	Classification <Decision Trees, K-Nearest Neighbor, Naïve Bayes>	7
3	Classification < Belief Network, Neural Networks>	7
3	Classification <Genetic Algorithm for classification, Performance Measures>	7
3	Clustering <Introduction, K-Means>	8
3	Clustering <DB-Scan, ROCK>	8
3	Clustering <SOM, Cluster Validation>	8
3	Feature Selection <Introduction, Filter methods, Wrapper methods>	9
3	Web Mining <Introduction, Fingerprinting, HITS algorithm>	9
3	DWH, Mining and BI	9
3	Applications of data mining	10