COMSATS University Islamabad
Department of Computer Science
Course Description Form (CDF)

Course Information

Course Code: **DSC306** Course Title: **Data Mining**
Credit Hours: **3(2,1)** Lecture Hours/Week: **2**
Lab Hours/Week: **3** Pre-Requisites: **None**

Course Objectives

- To discuss the application of preprocessing techniques on any given raw data;
- To discuss data mining algorithms to discover patterns;
- To analyze data mining tasks using advanced datasets from Kaggle, Google and implementation in python using Jupyter, Spider and Julia.

Course Contents

This course provides both theoretical and practical coverage of all data mining topics. The topics include: Data; Classification: Basic Concepts & Techniques; Classification: Alternative Techniques; Association Analysis: Basic Concepts & Algorithms; Association Analysis: Advanced Concepts; Cluster Analysis: Basic Concepts & Algorithms; Cluster Analysis: Additional Issues & Algorithms; Anomaly Detection; Avoiding False Discoveries; and Applications of the Classifiers & other Data Mining Techniques using Python.

Unit wise Major Topics

Unit	Topic	No. of teaching hours
1.	Introduction to Data Mining; Data: Types of Data; Data Quality; Data Preprocessing, and Measures of Similarity & Dissimilarity.	4
2.	Classification: Basic Concepts & Techniques, Decision Trees; Model Evaluation: Model Overfitting; and Evaluating the Performance of a Classifier.	6
3.	Rule-Based Classifier; Nearest-Neighbor Classifiers; Bayesian Classifiers; Artificial Neural Network (ANN), and Support Vector Machine (SVM).	5
4.	Association Analysis: Problem Definition; Frequent Itemset Generation; Rule Generation; FP-Growth Algorithm; Evaluation of Association Patterns; and Sequential Pattern.	6
5.	Cluster Analysis: Basic Concepts; K-means; Agglomerative Hierarchical Clustering; DBSCAN; and Cluster Evaluation	6
6.	Anomaly Detection; Avoiding False Discoveries; and Applications of the Classifiers & other Data Mining Techniques using Python.	3
Total Contact Hours		**30**

Mapping of CLOs and GAs

Sr.#	Unit #	Course Learning Outcomes	Blooms Taxonomy Learning Level	GA
CLO's for Theory				
CLO-1	1	Recognize the fundamental concepts of data mining.	*Understanding*	2

CLO-2	1-3	Apply preprocessing and classification techniques to solve classification problems of moderate complexity.	*Applying*	3,5
CLO-3	4	Apply Association rule mining techniques to extract patterns from a given problem.	*Applying*	3,5
CLO-4	5-6	Apply clustering techniques to solve clustering problems of moderate complexity.	*Applying*	3,5
CLO's for Lab				
CLO-5	1-6	Apply classification and clustering techniques such as decision trees, rule-based classifiers, K-means to real-world datasets.	*Applying*	3-5
CLO-6	1-6	Develop a project based on data mining concepts in a team environment.	*Creating*	3-6

CLO Assessment Mechanism

Assessment Tools	CLO-1	CLO-2	CLO-3	CLO-4	CLO-5	CLO-6
Quizzes	Quiz 1	Quiz 2	Quiz 3	Quiz 4	-	-
Assignments	-	Assignment 1&2	Assignment 3	Assignment 4	Lab Assignments	-
Mid Term Exam	Mid Term Exam	Mid Term Exam	Mid Term Exam	-	Lab Mid Term Exam	-
Final Term Exam	Final Term Exam				-	Lab Project/Lab Final Term Exam

Text and Reference Books

Textbook:

1. Data Mining: Concepts and Techniques, Han, J., Kamber, M. & Pei, J., Morgan Kaufmann Publishers, 2022.

Reference Books:

1. Introduction to Data Mining, Tan, P. N., Steinbach, M. & Kumar, V., Pearson, 2018.