AMERICAN INTERNATIONAL UNIVERSITY-BANGLADESH
Faculty of Science & Technology
Department of Computer Science
Undergraduate Program

COURSE OUTLINE
Term: Summer 2020-2021

I - Course Code and Title: CSC 4285: Data Warehouse & Data Mining

II - Credit: 3

III - Nature: Elective Course CSSE, SE, CIS, CS, CSE

IV- Prerequisite: CSC4121: Artificial Intelligence & Expert Sys. CSC2107: Introduction to Database

V- Course Description:

The course focuses on the basic data mining algorithms and their pros and cons. The topics covered in this course will help the students performing their research in data mining.

VI – Objectives:

- Introduction to knowledge discovery and data mining in databases and to present basic concepts relevant to real data mining applications, as well as reveal important research issues related to the knowledge discovery and mining applications.
- Fundamental concepts underlying knowledge discovery, data mining and hands-on experience with implementation of some data mining algorithms applied to real world cases.
- Research issues as well as mining strategies and issues relating specific industrial sectors; Systems for data mining.

VII – Topics to be Covered

TOPICS	Specific Objective(s)	Time Frame	Suggested Activities	Teaching Strategy(s)
Mission & Vision of AIUB, Basic Mathematical Concepts	To make students understand about the Mission & Vision of the university, Contents of Discrete Mathematics course required for Later Topics	Week 1	Discussion on Mission & Vision of AIUB, Review of Database Management	Lecture, Group Discussion
Data for Data Mining	Standard Formulation, Types of variable, Data preparation, Missing values	Week 1	Discussion, Group study and perform of exercises.	Lecture, Group study
Classification: Naïve Bayes and Nearest Neighbor	Classification, Naïve bayes classification, Distance measures, K-NN classification	Week 2	Discussion, Group study and perform of exercises using data mining tools	Lecture, Group study
Decision trees for classification	Decision rules and decision trees, TDIDT algorithm, Using entropy for attribute selection	Week 3 **Quiz 1**	Discussion, Group study and perform of exercises using data mining tools	Lecture, Group study
Decision trees for classification	Using frequency table for attribute selection, Gini Index of diversity, Inductive Bias, Gain Ratio, Split information.	Week 3 & Week 4 **Quiz 2**	Discussion, Group study and perform of exercises using data mining tools	Lecture, Group study
Continues Attribute Discretization	Local & global discretization, local discretization to TDIDT, ChiMerge Algorithm for global discretization	Week 4 & Week 5	Discussion, Group study and perform of exercises using data mining tools	Lecture, Group study
Estimating the Predictive Accuracy of a Classifier	Calculating predictive accuracy, separate training and test set, K-fold & N-flod cross validation, Confusion matrix, True and false positive	Week 6 **Quiz 3**	Discussion, Group study and perform of exercises using data mining tools	Lecture, Group study

Midterm Exams (Week 7)				
Measuring Performance of a classifier	True positive, false positive, false negative, true negative, performance measures, ROC graph, Finding best classifier	Week 8	Discussion, Group study and perform of exercises using data mining tools	Lecture PPT Slides Board Work Home Work
Avoiding Overfitting of Decision Trees	Clashes in training set, over fitting rules, pre-pruning, post pruning	Week 8 & week 9 **Quiz 1**	Discussion, Group study and perform of exercises using data mining tools	Lecture PPT Slides Board Work Home Work
Clustering	Introduction, K-means Clustering	Week 9	Discussion, Group study and perform of exercises using data mining tools	Lecture PPT Slides Board Work Home Work
Clustering	Agglomerative hierarchical Clustering, Exercise on Clustering	Week 10 **Quiz 2**	Discussion, Group study and perform of exercises using data mining tools	Lecture PPT Slides Board Work Home Work
Regression	Linear Regression	Week 11	Discussion, perform of exercises using data mining tools	Lecture PPT Slides Board Work Home Work
Association Rule Mining	Apriori Algorithm FP-Growth Algorithm	Week 12	Discussion, Group study and perform of exercises using data mining tools	Lecture PPT Slides Board Work Home Work
Data warehousing		Week 13 **Quiz 3**	Discussion, Group study and perform of exercises	Lecture PPT Slides Board Work Home Work
Final Term Exams (Week 14)				

VII – Evaluation

Assessment Type	Marks	Term
Short Assessment (Average of three)	30%	**Mid Term (40%)**
Assignment	20%	
Term Assessment (Average of two)	50%	
Total	100%	
Short Assessment (Average of three)	30%	**Fina Term (60%)**
Term Assessment (Average of Two)	50%	
Term VIVA	20%	
Total	100%	
Semester Total		**100%**

VIII – Textbook/ Reference Materials

1. Principles of Data Mining – Max Bramer
2. Data Mining Practical Machine Learning Tools and Techniques Second Edition Ian H. Witten, Eibe Frank
3. Data Mining Techniques: For Marketing, Sales, and Customer Support (Michael J. Berry, Gordon Linoff, Wiley)
4. Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems)
5. http://www.cs.waikato.ac.nz/ml/weka/documentation.html