جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Course Syllabus: Data Analytics - CS 220

Division	Computer, Electrical and Mathematical Sciences & Engineering
Course Number	CS 220
Course Title	Data Analytics
Academic Semester	Fall
Academic Year	2018/2019
Semester Start Date	08/26/2018
Semester End Date	12/11/2018
Class Schedule (Days & Time)	10:30 AM - 12:00 PM | Mon Wed

Instructor(s)

Name	Email	Phone	Office Location	Office Hours
Xin Gao	Xin.Gao@kaust.edu.sa	+966128080323	4217, 3, Ibn Sina (bldg. 3)	Time: Sun and Wed 10:30am-12noon. Location: Building 3, Level 4, Room 4217 (sea side). Tel: 012-8080323.

Teaching Assistant(s)

Name	Email
Yu Li; Fatima Smaili	yu.li@kaust.edu.sa; fatimazohra.smaili@kaust.edu.sa.

Course Information

Comprehensive Course Description	**Motivation for Establishing the Course:** This course will provide an introduction to the basics and the principles of artificial intelligence, data mining and machine learning. These areas are well aligned with the research in CEMSE Division and the missions of KAUST. This course also fits well in the fourth paradigm of science: data-intensive scientific discovery. **Course Description:** This course covers the basic concepts and algorithms for artificial intelligence, data mining and machine learning. The main contents of the course are: 1. Artificial intelligence: › Task environment › Performance measure › Problem solving by searching › Uninformed search › Informed search › Constraint satisfaction problems 2. Data mining: › Data and patterns › Summary statistics and visualization › Unsupervised feature selection › Clustering › Supervised feature selection › Individual feature ranking › Feature subset selection 3. Machine learning: › Cross validation › Supervised learning › K-nearest neighbors › Naïve Bayes › Decision trees › Support vector machines › Neural networks
Course Description from Program Guide	The course covers basic concepts and algorithms for artificial intelligence, data mining and machine learning. The main contents are: artificial intelligence (task environment, performance measure, and problem solving by searching), data mining (data and patterns, summary statistics and visualization, unsupervised feature selection, and supervised feature selection), and machine learning (cross validation and supervised learning).
Goals and Objectives	The goal of this course is to introduce the main ideas, fundamental concepts, and key algorithms in the fields of artificial intelligence, data mining, and machine learning. The course will prepare students to adapt to the big data era, to facilitate them to conduct research in data science, and to be able to apply the introduced algorithms in their own research areas.
Required Knowledge	Students who take this course are assumed to be familiar with algorithm runtime analysis (e.g., big O notations), probability theory (e.g., Gaussian distribution and conditional probability), and programming language (e.g., MATLAB or C++).
Reference Texts	There will be no textbook required for this course. The lectures and slides are self-contained. The reference books are “Artificial Intelligence – A Modern Approach” by Stuart Russell and Peter Norvig, the Prentice Hall press, and “Pattern Recognition and Machine Learning” by Christopher M. Bishop, the Oxford University press.
Method of evaluation	**20.00**% - Course Project(s) **10.00**% - Homework /Assignments **20.00**% - Midterm exam **10.00**% - Quiz(zes) **40.00**% - Final exam
Nature of the assignments	There will be written assignments with both theoretical and practical questions.
Course Policies	The final grade is composed of 10% assignment, 20% midterm, 10% in-class quiz, 20% semester-long project, and 40% final exam. The final exam of this course is also one of the three exams for the Ph.D. qualification exam in the CS program.
Additional Information	

Tentative Course Schedule *(Time, topic/emphasis & resources)*		
Week	**Lectures**	**Topic**
1	Mon 08/27/2018 Wed 08/29/2018	Introduction and task environment
2	Mon 09/03/2018 Wed 09/05/2018	Performance measure and problem solving by searching
3	Mon 09/10/2018 Wed 09/12/2018	Uninformed search
4	Mon 09/17/2018 Wed 09/19/2018	Informed search
5	Mon 09/24/2018 Wed 09/26/2018	Local search
6	Mon 10/01/2018 Wed 10/03/2018	Constraint satisfaction problems
7	Mon 10/08/2018 Wed 10/10/2018	Data mining and summary statistics
8	Mon 10/15/2018 Wed 10/17/2018	Data visualization and exploration
9	Mon 10/22/2018 Wed 10/24/2018	Unsupervised feature selection
10	Mon 10/29/2018 Wed 10/31/2018	Individual feature ranking
11	Mon 11/05/2018 Wed 11/07/2018	Feature subset selection
12	Mon 11/12/2018 Wed 11/14/2018	Density estimation and cross-validation
13	Mon 11/19/2018 Wed 11/21/2018	K-nearest neighbor and naive Bayes
14	Mon 11/26/2018 Wed 11/28/2018	Decision trees
15	Mon 12/03/2018 Wed 12/05/2018	Support vector machines
16	Mon 12/10/2018	
17		
18		

Note

The instructor reserves the right to make changes to this syllabus as necessary.