



## Course Syllabus: Data Analytics - CS 220

<b>Division</b>	Computer, Electrical and Mathematical Sciences & Engineering
<b>Course Number</b>	CS 220
<b>Course Title</b>	Data Analytics
<b>Academic Semester</b>	Fall
<b>Academic Year</b>	2019/2020
<b>Semester Start Date</b>	08/25/2019
<b>Semester End Date</b>	12/10/2019
<b>Class Schedule</b> (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
--------------------

<b>Comprehensive Course Description</b>	<p><b>Motivation for Establishing the Course:</b> 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.</p> <p><b>Course Description:</b> This course covers the basic concepts and algorithms for artificial intelligence, data mining and machine learning. The main contents of the course are:</p> <ol style="list-style-type: none"> <li>1. Artificial intelligence: <ul style="list-style-type: none"> <li>› Task environment</li> <li>› Performance measure</li> <li>› Problem solving by searching <ul style="list-style-type: none"> <li>› Uninformed search</li> <li>› Informed search</li> </ul> </li> <li>› Constraint satisfaction problems</li> </ul> </li> <li>2. Data mining: <ul style="list-style-type: none"> <li>› Data and patterns</li> <li>› Summary statistics and visualization</li> <li>› Unsupervised feature selection <ul style="list-style-type: none"> <li>› Clustering</li> </ul> </li> <li>› Supervised feature selection <ul style="list-style-type: none"> <li>› Individual feature ranking</li> <li>› Feature subset selection</li> </ul> </li> </ul> </li> <li>3. Machine learning: <ul style="list-style-type: none"> <li>› Cross validation</li> <li>› Supervised learning <ul style="list-style-type: none"> <li>› K-nearest neighbors</li> <li>› Naïve Bayes</li> <li>› Decision trees</li> <li>› Support vector machines</li> <li>› Neural networks</li> </ul> </li> </ul> </li> </ol>
<b>Course Description from Program Guide</b>	<p>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).</p>
<b>Goals and Objectives</b>	<p>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.</p>
<b>Required Knowledge</b>	<p>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++).</p>
<b>Reference Texts</b>	<p>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.</p>
<b>Method of evaluation</b>	<p>40.00% - Final exam 10.00% - Quiz(zes) 20.00% - Midterm exam 10.00% - Homework /Assignments 20.00% - Course Project(s)</p>
<b>Nature of the assignments</b>	<p>There will be written assignments with both theoretical and practical questions.</p>
<b>Course Policies</b>	<p>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.</p>
<b>Additional Information</b>	

## Tentative Course Schedule

*(Time, topic/emphasis & resources)*

<b>Week</b>	<b>Lectures</b>	<b>Topic</b>
1	Mon 08/26/2019 Wed 08/28/2019	Introduction and task environment
2	Mon 09/02/2019 Wed 09/04/2019	Performance measure and problem solving by searching
3	Mon 09/09/2019 Wed 09/11/2019	Uninformed search
4	Mon 09/16/2019 Wed 09/18/2019	Informed search
5	Mon 09/23/2019 Wed 09/25/2019	Saudi National Day
6	Mon 09/30/2019 Wed 10/02/2019	Constraint satisfaction problems
7	Mon 10/07/2019 Wed 10/09/2019	Data mining and summary statistics
8	Mon 10/14/2019 Wed 10/16/2019	Data visualization and exploration
9	Mon 10/21/2019 Wed 10/23/2019	Unsupervised feature selection
10	Mon 10/28/2019 Wed 10/30/2019	Mid-semester break
11	Mon 11/04/2019 Wed 11/06/2019	Mid-semester break
12	Mon 11/11/2019 Wed 11/13/2019	Mid-semester break
13	Mon 11/18/2019 Wed 11/20/2019	Mid-semester break
14	Mon 11/25/2019 Wed 11/27/2019	Mid-semester break
15	Mon 12/02/2019 Wed 12/04/2019	Mid-semester break
16	Mon 12/09/2019	Exams

### **Note**

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