



## Course Syllabus: Computational Methods in Data Mining - CS 340

<b>Division</b>	Computer, Electrical and Mathematical Sciences & Engineering
<b>Course Number</b>	CS 340
<b>Course Title</b>	Computational Methods in Data Mining
<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   Sun Tue

Instructor(s)				
Name	Email	Phone	Office Location	Office Hours
Xiangliang Zhang	Xiangliang.Zhang@kaust.edu.sa	+966128080313	4413, 1, Al-Khawarizmi (bldg. 1)	Tuesday 3:00-5:00PM

Teaching Assistant(s)	
Name	Email
TBD	TBD

Course Information	
<b>Comprehensive Course Description</b>	Focus is on both classical and new emerging advanced techniques in data mining. Topics include computational methods for the most popular types of data, including text, plain and attributed graph, bipartite graph in recommendation system and time series/sequences. Individual research project is required. The project is to solve real data mining problems (in text mining and graph mining), by using what you learned from the course.
<b>Course Description from Program Guide</b>	Focus is on very large-scale data mining. Topics include computational methods in supervised and unsupervised learning, association mining and collaborative filtering. Individual or group applications oriented programming project. 1 unit without project; 3 units requires final project.
<b>Goals and Objectives</b>	Students will master both classical and new emerging techniques in hot data mining areas, such as methods in supervised and unsupervised learning for text, word embedding and topic modeling, collaborative filtering, graph embedding, community detection and other graph analysis, especially deep learning models for solving text, graph problems and recommender systems.
<b>Required Knowledge</b>	Probability and Statistics, Linear Algebra, Data Analytics and/or Machine Learning, Programming skills (C/C++, Java, python, etc, more than Matlab)
<b>Reference Texts</b>	<ol style="list-style-type: none"> <li>1. Data Mining: Concepts and Techniques. Jiawei Han, Micheline Kamber and Jian Pei.</li> <li>2. Introduction to Data Mining. Pang-Ning Tan, Michael Steinbach, and Vipin Kumar.</li> </ol>
<b>Method of evaluation</b>	<b>70.00%</b> - Homework /Assignments <b>30.00%</b> - Course Project(s)

<b>Nature of the assignments</b>	<p>One paper will be assigned for reading under each topic. An in-class quiz will be used to evaluate the understanding of the assigned paper. The final Assignment score will be the average of points earned from all quizzes.</p> <p>There are two projects to complete at CS 340. One is under the topic of text data mining, to be completed by week 9. The other is under the topic of graph data mining, to be completed by week 15. Both projects should target on solving real data mining problems, by using what you learned from the course.</p> <p>Project Requirements:</p> <ul style="list-style-type: none"> <li>-Project reports should be submitted in the format of a research paper.</li> <li>-Project presentation should be given at the class.</li> <li>-Project evaluation will follow: Technical quality (30) + significance (30) + novelty/impact (20) + report/presentation (20)</li> </ul>
<b>Course Policies</b>	Each student must attend every class. A leave of absence must be applied for in advance by sending an email to the instructor.
<b>Additional Information</b>	

### Tentative Course Schedule

*(Time, topic/emphasis & resources)*

Week	Lectures	Topic
1	Sun 08/25/2019 Tue 08/27/2019	Text Mining Introduction, SVD and LDA
2	Sun 09/01/2019 Tue 09/03/2019	Word embedding
3	Sun 09/08/2019 Tue 09/10/2019	Word, Sentence, Document Embedding
4	Sun 09/15/2019 Tue 09/17/2019	Seq2Seq
5	Sun 09/22/2019 Tue 09/24/2019	Attention,Transformer
6	Sun 09/29/2019 Tue 10/01/2019	BERT
7	Sun 10/06/2019 Tue 10/08/2019	Project Presentation
8	Sun 10/13/2019 Tue 10/15/2019	Graph Mining Introduction
9	Sun 10/20/2019 Tue 10/22/2019	PageRank, Graph embedding
10	Sun 10/27/2019 Tue 10/29/2019	Graph Neural network
11	Sun 11/03/2019 Tue 11/05/2019	Graph Project Presentation
12	Sun 11/10/2019 Tue 11/12/2019	Recommendation System
13	Sun 11/17/2019 Tue 11/19/2019	Recommendation System
14	Sun 11/24/2019 Tue 11/26/2019	Time series mining
15	Sun 12/01/2019 Tue 12/03/2019	Time-series Mining
16	Sun 12/08/2019 Tue 12/10/2019	Final

#### Note

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