MIS 496A "SPECIAL TOPICS IN DATA ANALYTICS" - Spring 2016

Hsinchun Chen, Professor, Department of MIS

Instructor: Hsinchun Chen, Ph.D., Professor, Management Information Systems Dept, Eller College of Management, University of Arizona
Time/Classroom: M/W 2:00PM-3:15PM MCCL 122
Instructor's Office Hours: M/W 10:00-11:00AM or by appointment
Office/Phone: MCCL 430X, (520) 621-4153
Email/Web site: hchen@eller.arizona.edu; https://ai.arizona.edu/about/director (email is the best way to reach me!)
Class Web site: http://ai.eller.arizona.edu/hchen/mi496A/ (VERY IMPORTANT!)
Teaching Assistants (TAs):
- Weifeng Li, weifengli@email.arizona.edu, MIS Ph.D. student (office: MCCL 430)
- Sagar Samtani, sagars@email.arizona.edu, MIS Ph.D. student (office: MCCL 430)

TA Office Hours: TA hours will be announced via email.

CLASS MATERIAL (Optional)

- *Data Mining with Weka*, Ian H. Witten and Eibe Frank (also with a 5-week MOOC course). http://www.cs.waikato.ac.nz/ml/weka/
- *Visual Insights: A Practical Guide to Making Sense of Data*, MIT Press, Katy Börner & David E. Polley, 2014 (also with a 7-week MOOC course). http://info.ils.indiana.edu/~katy/S637/
- *An Introduction to Statistical Learning, with Applications in R,* Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, Springer, 2013. http://www.StatLearning.com/
- Additional readings and handouts will be distributed in class and made available through the class web site.

COURSE OBJECTIVES

Business intelligence and analytics and the related field of big data analytics have become increasingly important in both the academic and the business communities over the past two decades. The *IBM Tech Trends Report* identified business analytics as one of the four major technology trends in the 2010s and beyond. A report by the McKinsey Global Institute predicted that by 2018, the United States alone will face a shortage of 140,000 to 190,000 people with deep data analytical skills, as well as a shortfall of 1.5 million data-savvy managers with the know-how to analyze big data to make effective decisions. Big data and data science have begun to transform different facets of the society, from e-commerce and global logistics, to smart health and cyber security.

This undergraduate senior level course (elective) will cover the important concepts and techniques relating to data analytics, including: statistical foundation, data mining methods, data visualization, and web mining techniques that are applicable to emerging e-commerce, government, health and security applications. The course contains lectures, readings, lab sessions, and hands-on projects. Most business school seniors are welcome. The course will require some basic computing and database background. The course will prepare students to become a data scientist or a data-savvy manager for different businesses.

PREREQUISITE FOR THE COURSE

Some programming experience in selected modern computing languages (e.g., Java, C, C++, Python) and DBMS (SQL).

COURSE TOPICS

Topic 1: Introduction
- Business intelligence and analytics
- Data, text and web mining overview

Topic 2: Data Mining
- Statistical analysis: regression, discriminant analysis, principal component analysis
- Symbolic learning: decision trees, random forest
- Neural networks and soft computing: perceptron, self-organizing maps, genetic algorithms
- Statistical machine learning and graph models: Support Vector Machine, social network analysis
- Evaluation and validation: hold-out sampling, cross-validation

Topic 3: Text Mining
- Digital library and search engines
- Information retrieval: vector space model, bag of words, text segmentation
- Information extraction: entity extraction, relation extraction, topic extraction
- Sentiment analysis: lexicon-based, machine learning based
- Data/information visualization: visualizing data

Topic 4: Web Mining
- Web 1.0, 2.0, 3.0
- Search engines: ranking, search logs, search algorithms
- Deep web and dark web
- Social media and crowdsourcing systems: wisdom of the crowd, sentiment analysis
- Cloud computing and big data analytics: Hadoop, MapReduce, Mahout, Spark
- Internet of Things: mobile sensors, mobile security

GRADING POLICY

Project proposal	5%
Midterm exam	30%
Review paper	15%
Research project	40%
Class attendance and participation	10%
TOTAL	100%

MIDTERM EXAM (30%)

The midterm exam will be closed book, closed notes and in the short-essay format (8-10 questions). The questions will be based mostly on classroom lectures. There will be NO Final Exam for this class. Academic integrality will be strictly enforced. Consequence for cheating will be severe.

REVIEW PAPER PRESENTATION AND PROPOSAL (20%)

Each student will be required to select an emerging data analytics topic of interest and develop a comprehensive review paper for the topic. Secondary literature review will be needed based on recent papers published in press, magazines, conferences, and journals. Each student will be required to present their review in the second half of the semester (10 minutes each). The instructor will suggest selected emerging topics for consideration. A paper review and project proposal will be needed in the third week of the semester.

RESEARCH PROJECT PRESENTATION AND PAPER (40%)

Each student will be required to propose and execute an individual, data-driven research project in data analytics for applications of interest to the student. The instructor will suggest suitable data and algorithms for consideration. The class TAs will also provide assistance in data preparation and analytics using selected open source tools. Each student will present at the end of the semester (15 minutes) and a final research paper (8 pages, IEEE format) will be submitted after all presentation sessions. The instructor will provide details about the final paper format and structure. Students are expected to gain significant hands-on data analytics experience through the project.

LECTURES, ATTENDANCE, AND ACADEMIC INTEGRITY

Students are required to attend all lectures on time and honor academic integrity. Missing classes will result in loss of points or administrative drop by the instructor. Students are required to send excuse notes (via email) to the instructor before missing classes. Students are permitted to bring laptop to classroom for note taking purposes, but not for checking email or web surfing. Professional attitude and strong work ethics are needed for this class. Students are encouraged to consult the instructor for advice and help.

LAB SESSIONS and GUEST SPEAKERS

Selected lab sessions will be provided during the semester on the following topics: web services, cloud computing platforms, Hadoop, Weka, etc. Selected guest speakers will present in the class.

COURSE OUTLINE (tentative)

DATE	TOPIC	CONTENT/NOTES
*Jan 13	Syllabus & registration	class roster, student form (TA)
Jan 18 (M)	Martin Luther King Jr. Holiday	NO CLASS
Jan 20	Class overview and expectation	syllabus
Jan 25 (M)	Data analytics, big data	readings, discussions
Jan 27	Business analytics	readings, discussions
Feb 1 (M)	Emerging applications	handouts, resources
Feb 3	<u>Web computing & mining</u>	overview, web evolution
Feb 8 (M)	Web 1.0, Surface Web	overview
	PROJECT PROPOSAL DUE (REVIEW & RESEARCH, 5%)	
Feb 10	Search engine, deep web	graph search
Feb 15 (M)	Web 2.0, Social Web	readings, overview
Feb 17	Social networking, dark web	readings, lecture
Feb 22 (M)	Web 3.0, Mobile Web	readings, SoLoMo web
Feb 24	IoTs, cybersecurity	readings
Mar 1 (M)	Cloud, Hadoop, Spark	TA lecture
Mar 3	<u>Data mining</u>	overview
Mar 7 (M)	Open source tool, Weka	TA lecture
Mar 8	LAST DAY TO DROP (with "W")	
Mar 9	Classification algorithms	Regression, discriminant analysis
*Mar 14-18	SPRING RECESS	NO CLASS
Mar 21 (M)	**MIDTERM EXAM (30%)**	
Mar 23	Classification algorithms	ID3, SVM
Mar 28 (M)	Classification algorithms	NN
Mar 30	Classification algorithms	HMM, CRF
Apr 4 (M)	**REVIEW PAPER PRESENTATION (15%)**	
Apr 6	**REVIEW PAPER PRESENTATION**	
Apr 11 (M)	Clustering algorithms	K-means, SOM
Apr 13	Clustering algorithms	LDA
Apr 18 (M)	<u>Text mining</u>	overview
Apr 20	Sentiment analysis	opinion mining, social media
Apr 25 (M)	Data visualization	HCI
Apr 27	**RESEARCH PROJECT PRESENTATION (30%)**	
May 2 (M)	**RESEARCH PROJECT PRESENTATION**	
May 4	**RRESEARCH PROJECT PRESENTATION**	
May 6-12	FINAL EXAM WEEK	NO EXAM FOR MIS 496A
May 11	**FINAL PROJECT PAPER DUE (10%)**	