MIS 464 "DATA ANALYTICS" - Spring 2019

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: T/TH 9:30AM-10:45AM MCCL 126
Instructor's Office Hours: T/TH 2:00-3:00PM 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/mis464/ (VERY IMPORTANT!)
Teaching Assistants (TAs):
- Shuo Yu, shuoyu@email.arizona.edu, Ph.D. student (office: MCCL 430 Cubical #34-35)
- Hongyi Zhu, zhuhy@email.arizona.edu, Ph.D. student (office: MCCL 430 Cubical #36-37)
TA Office Hours: TA hours will be announced via email.

CLASS MATERIAL (Optional)
- *Data Mining: Practical Machine Learning Tools and Techniques*, by Witten, Frank, Hall & Pal, 4th Edition, 2017, Morgan Kaufmann (also with a 5-week MOOC course). See more at: http://www.cs.waikato.ac.nz/ml/weka/
- *Artificial Intelligence: A Modern Approach*, by Russel & Norvig, 3rd Edition, 2000, Prentice Hall
- *Deep Learning*, by Goodfellow, Bengio & Courville, 2016, MIT Press
- 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, AI, deep learning, and web mining techniques that are applicable to emerging e-commerce, government, and 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
Programming experience in selected modern computing languages (e.g., Java, C, C++, Python) and DBMS (SQL).

COURSE TOPICS (selected topics will be covered)

Topic 1: Introduction (the field of MIS & CS)
- From computational design science in MIS to applied data science in CS
- Business intelligence and analytics, opportunities & techniques
- Emerging AI applications, from face recognition to autonomous vehicle
- Data, text and web mining overview: AI, ML, deep learning
- Data mining and web computing tools (by TAs): Weka, Tableau, Hadoop, SPARK

Topic 2: Web Mining (the changing world)
- Web 1.0, 1995-: WWW, search engines, surface web, spidering, graph search, genetic algorithms
- Web 2.0, 2005-: deep web, web services & mesh-ups, social media, crowdsourcing systems, network sciences
- Web 3.0, 2010-: IoTs, mobile & cloud computing, big data analytics, dark web, mobile analytics, cybersecurity
- Web 4.0, 2015-: AI-empowered society, image/face, translation, drones, autonomous vehicles, health, security

Topic 3: Data Mining (the analytics techniques)
- Symbolic learning: decision trees, random forest
- Statistical analysis: regression, principal component analysis, Naïve Bayes
- Statistical machine learning: Support Vector Machines, Hidden Markov Models, Conditional Random Fields
- Neural networks and soft computing: feedforward networks, self-organizing maps, genetic algorithms
- Network analysis: social network analysis, graph models
- Deep learning: Convolutional NN, Recurrent NN, Long Short-Term Memory
- Representation learning: Transfer Learning, Deep Generative Models

Topic 4: Text Mining (handling unstructured text)
- Digital library and search engines
- Information retrieval & extraction: vector space model, entity & topic extraction
- Authorship analysis: lexical, syntactic, structural, and semantic analysis
- Sentiment and affect analysis: lexicon-based, machine learning based
- Information visualization: scientific, text, and web visualization

Topic 5: Emerging Research in Data and Web Mining (major conferences, groups, opportunities)
- Emerging research in major data and web mining conferences: ACM KDD, IEEE ICDM, WWW, ACM SIGIR, ACM CHI, AAAI, IJCAI, ICML, NIPS, ICLR
- Key journals: MISQ, ISR, IEEE TKDE, JAMIA, JBI, JASIST
- Emerging research in major academic institutions: Stanford, Berkeley, CMU, MIT
- Emerging research in major industry research labs: Google, Facebook, Amazon, Baidu, Microsoft
- Emerging data and web mining applications: health, security, e-commerce, AV, drones, robotics

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 integrity will be strictly enforced. Consequence for cheating will be severe.

REVIEW PAPER PRESENTATION AND PROPOSAL (20%)
Each student will be required to form a two-person team. Each team will 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 team (both students) 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 team will be required to propose and execute a data-driven research project in data analytics for applications of interest to the students. 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 team (both students) 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, Tableau, Weka, etc. Selected guest speakers will present in the class

COURSE OUTLINE (tentative)

DATE	TOPIC	CONTENT/NOTES
Jan 10	Syllabus & registration	Class roster, syllabus
Jan 15 (T)	MIS, CS, design science	Readings, discussions
Jan 17	Big data, applications	Readings, discussions
Jan 22 (T)	BI, data analytics	Readings, discussions
Jan 24	AI, deep learning	Readings, discussions
	PROPOSAL DUE (REVIEW & RESEARCH, 5%)	
Jan 29 (T)	**<u>Web Computing & Mining</u>**	Overview, discussions
Jan 31	Cloud, Hadoop, SPARK	TA session
Feb 5 (T)	Web 1.0, Surface Web	Overview, discussions
Feb 7	Search engine, graph search	Readings, lecture
Feb 12 (T)	Web 2.0, Social Web	Overview, discussions
Feb 14	Deep web, social media, SNA	Readings, lecture
Feb 19 (T)	Web 3.0, Mobile Web, IoT, dark web	Overview, discussions
Feb 21	Web 4.0, AI Web	Overview, discussions
Feb 26 (T)	**<u>Data Mining</u>**	Overview, discussions
Feb 28	Symbolic learning, AI, decision trees	ID3, RF
Mar 4-8	SPRING RECESS	NO CLASS
Mar 12 (T)	**MIDTERM EXAM (30%)**	
Mar 14	Statistical analysis, regression, Bayes	Overview, discussions
Mar 19 (T)	DM tools, Weka, Tableau	TA session
Mar 21	Statistical ML, SVM, CRF	Readings, lecture
Mar 26 (T)	Neural networks, Backprop	Readings, lecture
Mar 28	Deep learning, review	Readings, lecture
Apr 2 (T)	**REVIEW PAPER PRESENTATION (15%)**	
Apr 4	**REVIEW PAPER PRESENTATION**	
Apr 9 (T)	Deep learning, CNN	Readings, lecture
Apr 11	**<u>Text Mining</u>**	Overview, discussions
Apr 16 (T)	IR/IE, Sentiment analysis	Readings, lecture
Apr 18	Information Visualization	Readings, lecture
Apr 23 (T)	**RESEARCH PROJECT PRESENTATION (30%)**	
Apr 25	**RESEARCH PROJECT PRESENTATION**	
Apr 30 (T)	**RRESEARCH PROJECT PRESENTATION**	
May 3-9	FINAL EXAM WEEK	NO EXAM FOR MIS 464
May 9	**FINAL PROJECT PAPER DUE (10%)**	