## **Department of Computer Science**

inue I , Academic Buil	Huntsville, Texas ding One P.O. Box 2090 • 77341-2090	(936) 294-3846 Fax • (936) 294-4312				
COURSE SYLLABUS COSC 4314.01 (CRN: 80776) Data Mining 3 Credit hours Fall 2017						
Class	room: AB1 #206					
Meeting time:						
Lectures:	MW 08:00am – 09:20am					
Instructor: Name: Office: Office hours: Advising hours: Phone: Email:	Hyuk Cho AB1 #216J MTW 10:00–11:30am (Other times can be availa MTW 11:30–12:00pm (Other times can be availa (936) 294-1535 hyukcho@shsu.edu	ble by appointment.) ble by appointment.)				
URL:	http://www.shsu.edu/~hxc005/					
Teaching Assistant: Name: N/A	Office: N/A Phone: N/A Email: N/A	Office hours: N/A				
Course description: Prerequisite:	This course provides an introduction to the newly- mining is concerned with the automatic extraction of from large amounts of data in practical real wo fundamental concepts, data preparation and featu algorithms (including but not limited to associate prediction), and applications and evaluation of data in <b>COSC 3318</b> and <b>COSC 3319</b>	emerging field of data mining. Data of novel information and knowledge orld problems. Topics will include re selection, standard data mining tion, classification, clustering, and nining techniques.				
Statement of Need:	The explosive increase of data and fast-evolving scien new techniques and methodologies for the automatic knowledge from large amounts of data. Data minin directly addresses this demand and has application Computing Accreditation Commission of the Accred Technology (CAC/ABET) identifies data mining as cr undergraduate Computer Science program is currer addition of this course is vital for the maintenance program.	ce and engineering problems demand e extraction of novel information and ng is the newly-emerging field that on to many real-life domains. The editation Board for Engineering and ritical to continued accreditation. The ntly accredited by CAC/ABET. The e of the current accreditation of the				
Course objectives:	The objective of this course is to expose students to f for discovering patterns from data. First, students will and gain insight into a dataset. Then, students will di algorithms, technologies, and methodologies. Addition students will have hands-on exercises with data mining name a few. Therefore, students will gain both theor on Data Mining. Particularly, top 10 data mining (and related topics will be covered for this semester.	Fundamental concepts and algorithms Il learn data-related topics and issues scuss details of various Data Mining onally, throughout the whole course, ng tools such as Weka and Orange, to retical knowledge and practical skills and machine learning) algorithms and				
Instructor's Objectives:	<b>OBJECTIVES</b> of this course are for students: (1) TO GAIN FACTUAL KNOWLEDGE (terminological) and (2) TO LEARN FUNDAMENTAL PRINCIPAL THEORIES.	ogy, classifications, methods, etc.) L <b>S, GENERALIZATIONS, OR</b>				

## **References (OPTIONAL):** NOTE: (1)-(4) are usual textbooks, while (5)-(7) are hands-on books. (1) Introduction to Data Mining by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Addition Wesley, 2005. (2) Data Mining: Concepts and Techniques (2<sup>nd</sup> Ed.) by Jiawei Han and Micheline Kamber, Morgan Kaufmann Publishers (Elsevier), 2006. (3) Data Mining: Practical Machine Learning Tools and Techniques (3<sup>rd</sup> Ed.) by Ian H. Witten, Frank Eibe, and Mark A. Hall, Morgan Kaufmann Publishers (Elsevier), 2011. (4) Data Mining and Analysis: Fundamental Concepts and Algorithms by Mohammed J. Zaki and Wagner Meira, JR., Cambridge University Press, 2014. (5) Learning scikit-learn: machine Learning in Python by Raul Garreta and Guillermo Moncecchi, Packt Publishing, 2013. (6) Learning Data Mining with Python by Robert Layton, Packt Publishing, 2015. (7) Python Machine Learning by Sebastian Raschka, Packt Publishing, 2015. **Supporting Materials:** All supporting materials will be posted on Blackboard.

Weekly schedule: The follow announced

The following schedule is preliminary and subject to change. Any updates/changes will be announced at least one week in advance. All the related materials (e.g., lecture slides, summary, other resources, programming exercise questions and assignments, etc) will be timely available.

Week	Торіс	H/W	Note
1	Introduction to Data Mining (DM) (Motivation, Challenges, Tasks, Scope, etc.)	TBA	Quiz
2	Exploring data (Data types, Summary statistics, Measures, etc.)	TBA	Quiz
3	Exploring data (cont.) (Pre-/Post-processing, Data visualization, etc.)	TBA	Quiz/ Exam
4	Supervised Learning (Concepts, Algorithms, etc.)	TBA	Quiz
5	Supervised Learning (Algorithms, Applications, etc.)	TBA	Quiz
6	Supervised Learning (Applications, Summary, etc.)	TBA	Quiz/ Exam
7	Unsupervised Learning (Concepts, Algorithms, etc.)	TBA	Quiz
8	Unsupervised Learning (Algorithms, Applications, etc.)	TBA	Quiz
9	Unsupervised Learning (Applications, Summary, etc.)	TBA	Quiz/ Exam
10	Advanced DM topics (Hands-on programming with real-life applications)	TBA	Quiz
11	Advanced DM topics (cont.) (Hands-on programming with real-life applications)	TBA	Quiz
12	Advanced DM topics (cont.) (Hands-on programming with real-life applications)	TBA	Quiz/ Exam
13	Advanced DM topics (or Special topics) (Hands-on, Recent development, etc.)	TBA	Quiz
14	Advanced DM topics (or Special topics) (cont.) (Hands-on, Recent development, etc.)	TBA	Quiz
15	Final Exam	TBA	Exam

## **Course requirements:**

Attendance:	Attendance of lectures are required and counted to 10% of whole grade. The roster will be circulated at the beginning of each class; thus, please come to the classroom on time and give your signature on the roster. If you are absent continuously for two or more classes, you will be reported to the SAM Center via FirstAlert.				
Reading:	Reading assignment will be announced at the end of each class. Class Participation (e.g., leading discussions, giving presentations, etc.) is encouraged and counted to the grade.				
Assignments:	<ul> <li>Ieading discussions, giving presentations, etc.) is encouraged and counted to the grade.</li> <li>Note that each assignment will be of the following formats: (1) problem solving on paper;</li> <li>(2) programming (in either Python or in Java); and/or (3) hands-on exercise with specific tools. Assignments will be given periodically (i.e., weekly).</li> <li>Unless otherwise mentioned, each assignment is due in class (usually one week after each announcement). Late homework is accepted up to the beginning of the next lecture class with a penalty of 50% off of the specified points. Graded homework and exam papers will be returned back to students usually within one week from the due date.</li> <li>NOTE: In addition to the prerequisites (i.e., COSC 3318 and 3319), mathematical background knowledge (particularly in Statistics, Linear Algebra, and etc.) as well as programming skills (particularly in Python and/or Java) will be greatly beneficial. Furthermore, quick learning of new tools and programming libraries is a big plus.</li> <li>NOTE: Discussion with your colleagues is highly encouraged; however, the (final)</li> </ul>				
Quizzes:	Most of the lecture classes will start (or end) with a brief quiz (5-10 minutes), covering materials from the assignment (or lecture). The nature of the question will be informed for students' preparation. <b>NOTE:</b> Quizzes can replace some of the planned exams (specified as " <b>Ouiz/Exam</b> "), the change will be announced in advance.				
Exams:	At the end of each category (specified in the weekly table), students will take an exam on fundamental concepts and algorithms. Additionally, students will take one comprehensive, in-class exam at the end of the semester. There is no make-up examination available for any students. Students who fail to take any examination fail the class automatically unless excused for good cause by the college Dean				
Presentation:	( <b>OPTIONAL</b> ) If time allows and students would like to get extra points, each student may give a presentation on a specific project topic she/he chooses (during the last three weeks of the semester), which will be peer-evaluated by classmates.				
Special Topics:	( <b>OPTIONAL</b> ) If time allows, special topics will be discussed. Potential topics include recent topics on Data Science, Apache Hadoop, Apache Spark, and/or TensorFlow.				
Grading plan:	Class Participation	10% 10%			
	Assignments	30%			
	Exam1	15%			
	Exam2	15%			
	Exam3	20%			
	Presentation (or Project)		(10%)	(OPTIONAL)	
		100%	(110%)	1	
	Finally, $101AL \ge 90\% \Rightarrow$	"A"			
	$80\% \ll 101AL \ll 89\% \Rightarrow$	B.,			
	$70\% \ll TOTAL < 79\% \Rightarrow$	"C"			
	$60\% \ll \text{TOTAL} < 69\% \Rightarrow$ $\text{TOTAL} < 59\% \Rightarrow$	"D" "F"			
Class participation:	In accordance with University Policy, regular attendance is required and your attendance will be seriously monitored ( <u>http://www.shsu.edu/dept/dean-of-students/absence.html</u> ). So, don't forget to give your signature on the roster. You are responsible for all material covered in classes and labs, regardless of whether you attended or not. It is your responsibility to obtain class materials from fellow classmates if you miss a class.				

- **Electronic Devices:** Please keep all electronic devices (e.g., laptops and phones) put away (or in your backpack) and silenced during class. You are not allowed to use any electronic device during the lecture and/or discussion, unless otherwise permitted in advance if necessary for your learning or for your special need. Focusing on lecture and participating discussion are the best way to use time and learn topics of each class. Additionally, if you are texting in class and you will be asked to leave class. I take this very seriously.
- **Rules of conduct:** Students will refrain from behavior in the classroom that intentionally or unintentionally disrupts the learning process and, thus, impedes the mission of the university. Please turn off or mute your cellular phone and/or pager before class begins. Students are prohibited from eating in class, using tobacco products, making offensive remarks, reading newspapers, sleeping, talking among each other at inappropriate times, wearing inappropriate clothing, or engaging in any other form of distraction. Inappropriate behavior in the classroom shall result in a, minimally, a directive to leave class or being reported to the Dean of Students for disciplinary action in accordance with university policy.
- Academic dishonesty: All students are expected to engage in all academic pursuits in a manner that is above reproach. Students are expected to maintain complete honesty and integrity in the academic experiences both in and out of the classroom. Any student found guilty of dishonesty in any phase of academic work will be subject to disciplinary action. The university and its official representatives may initiate disciplinary proceedings against a student accused of any form of academic dishonesty including, but not limited to, cheating on an examination or other academic work which is to be submitted, plagiarism, collusion and the abuse of resource materials.

No cheating on an examination or assignments is allowed. A score of zero will be given to the student if such a case occurred.

Visitors in the classroom: Occasion visiting of classes by responsible persons is allowed with prior arrangement with the instructor, as long as it does not interfere with the registered members of the class or the educational process.

## Other administrative matters:

**Students with Disabilities Policy:** It is the policy of Sam Houston State University that individuals otherwise qualified shall not be excluded, solely by reason of their disability, from participation in any academic program of the university. Further, they shall not be denied the benefits of these programs nor shall they be subjected to discrimination. Students with disabilities that might affect their academic performance should register with the Office of Services for Students with Disabilities located in the Lee Drain Annex (telephone 936-294-3512, TDD 936-294-3786, and e-mail <u>disability@shsu.edu</u>). They should then make arrangements with their individual instructors so that appropriate strategies can be considered and helpful procedures can be developed to ensure that participation and achievement opportunities are not impaired.

SHSU adheres to all applicable federal, state, and local laws, regulations, and guidelines with respect to providing reasonable accommodations for students with disabilities. If you have a disability that may affect adversely your work in this class, then I encourage you to register with the SHSU Services for Students with Disabilities and to talk with me about how I can best help you. All disclosures of disabilities will be kept strictly confidential. NOTE: No accommodation can be made until you register with the Services for Students with Disabilities. For a complete listing of the university policy, see: https://www.shsu.edu/dotAsset/7ff819c3-39f3-491d-b688-db5a330ced92.pdf

**<u>Religious Holidays:</u>** An institution of higher education shall excuse a student from attending classes or other required activities, including examinations, for the observance of a religious holy day, including travel for that purpose. A student whose absence is excused under this subsection may not be penalized for that absence and shall be allowed to take an examination or complete an assignment from which the student is excused within a reasonable time after the absence.