Syracuse University
IST 707 Data Analytics
COURSE SYLLABUS
2019 Spring Semester

_____

Instructor: Frank Marullo   Phone: 315-409-6641 (cell)
Class Room: Hinds 013   Email: fmarullo@syr.edu
Office Hours: By appointment

_____

Faculty Assistance:
Sneha Subramaniam   Email: ssubrama@syr.edu

Prerequisite: IST 687. Exceptions maybe given to students who have acquired skills equivalent to what is taught in IST687

Audience: Graduate Students

Description: General overview in data mining techniques, familiarity with particular real-world applications, challenges involved in applications, and future directions of the field. Optional hands-on experience with available software packages.

Additional Course Description:
This course will introduce data mining methods for extracting knowledge from data. The principles and theories of data mining methods will be discussed and will be related to the issues in applying data mining to problems. Students will also acquire hands-on experience using state-of-the-art software to develop data mining solutions to scientific and business problems. The focus of this course is in understanding of data and how to formulate data mining tasks in order to solve problems using the data.

The topics of the course will include the key tasks of data mining, including data preparation, concept description, association rules, classification, clustering, evaluation and analysis. Through the exploration of the concepts and techniques of data mining and practical exercises, students will develop skills that can be applied to business, science or other organizational problems.

The format of the class meetings will be a combined lecture and lab format, with lectures and class discussions to cover material and lab time to investigate small examples for the topic of the week. There will be weekly readings based on the textbook and on other materials, which will be posted online.

Credits: 3

Learning Objectives:
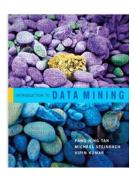**After taking this course, the students will be able to:**
- Document, analyze, and translate data mining needs into technical designs and solutions.

_____

- Apply data mining concepts, algorithms, and evaluation methods to real-world problems.
- Employ data storytelling and dive into the data, find useful patterns, and articulate what patterns have been found, how they are found, and why they are valuable and trustworthy.

Bibliography/ Texts / Supplies – Required:

- Pang-Ning Tan, Michael Steinbach, and Vipin Kumar (2005) Introduction to Data Mining. (ASIN: B011DAYQI2)

Bibliography/ Texts / Supplies– Recommended

- Tom Mitchell (1997) Machine Learning.
  (http://www.cs.cmu.edu/~tom/mlbook.html)
- Brett Lantz (2015) Machine Learning with R (second edition).

Note to students: Given the diversified background of data science students, one textbook might not fit everyone. If you like rigorous algorithm presentation, I would recommend Mitchell's classic book on Machine Learning. If you like more lay-person explanation of machine learning, see if you like Lantz's book better. The current required textbook is a balance between the two views.

Tips for success in this class: Curiosity, critical thinking, math, and programming.
- *Curiosity*: Curious about the data, pay attention to the data details. Don't treat a data set as a blackbox. Don't treat an algorithm as a blackbox. Try see through them.
- *Critical thinking*: Data mining is essentially research. You will learn and practice methods to discover patterns, and also evaluate whether and why the discovered patterns are true and useful.
- *Math:* You will need some math knowledge, such as algebra and probability, to understand how the data mining algorithms work.
- *Programming*: Although GUI tools like Weka and Rapid Miner would allow users with no programming skills to play with data sets, data sets are rarely immediately ready for analysis in these tools. The results from off-the-shelf tools may need

_____

additional transformation to see patterns. Programming skills would help you pre- and post-processing the data. Programming would also help you gain more convenient control over algorithm tuning in your scripts.

Software
R and Weka

Note to students: We will mainly use R but keep Weka as a backup tool for students who do not have enough R skills.

Requirements:
Your final grade is determined by your performance on the items in the table below. An overview of each item is provided in the remainder of this section.

Assessment Item	# of Points
8 Homework assignments (5 pts each)	40
Class Exercises	20
Final Project report	20
Class Participation	20
Total	**100%**

- **Homework assignments:** Assignments must be professionally prepared and submitted electronically to the LMS. All assignments should be submitted in Word files named as "*HW_Num_Lastname_Firstname.doc(x)*", e.g. "HW_1_Smith_John.doc". No PDF please.
- **Participation and Class exercises:** Students are required to participate in class discussions and exercises. These exercises are designed to encourage students to practice their newly learned knowledge. Class exercises that are turned in are required to include the names of all students who worked on the exercises if the assignment was completed as a group assignment.
- **Course project:** The objective of the project is to use the main skills taught in this class to solve a real data mining problem. Groups will be assigned for this project.
  - o Checkpoint 1: project idea proposal and presentation: Your idea proposal should include an overview of the data mining problem, the data set you will use and its availability, and your proposed data mining approach.
  - o Checkpoint 2: project progress presentation: show preliminary results and major challenges.
  - o Checkpoint 3: final project presentation: last round of feedback before submitting the final report.

_____

      o <u>Final project report</u>: The final project report should describe the data mining problem, its significance and broader impact, the data mining approaches, results, and interpretation of the discovered patterns.

Grading:

For this class, an "A" would mean the student has the capability to independently solve a simple data mining task. Below is a common formula for number-to-letter grade conversion.

Grade	Points	Grade	Points	Grade	Points	Grade	Points
		B+	87-89	C+	77-79	F	0-69
A	93-100	B	83-86	C	73-76		
A-	90-92	B-	80-82	C-	70-72		

Grades of D and D- may not be assigned to graduate students.

University Attendance Policy

Attendance in classes is expected in all courses at Syracuse University. Students are expected to arrive on campus in time to attend the first meeting of all classes for which they are registered. Students who do not attend classes starting with the first scheduled meeting may be academically withdrawn as not making progress toward degree by failure to attend. Instructors set course-specific policies for absences from scheduled class meetings in their syllabi. It is a federal requirement that students who do not attend or cease to attend a class to be reported at the time of determination by the faculty. Students should also review the university's religious observance policy and make the required arrangements at the beginning of each semester.

Course Specific Policies on attendance, late work, make up work, examinations if outside normal class time, etc.:

- **Registration**: Students must register prior to the first class or may be restricted from registering. If you are registered but not present at the first class, you run the risk of being administratively deregistered from this course so that your seat can be given to a student on the wait list.
- **Late Policy for Assignments**: A common error analysis file will be released when assignments are turned one week after submission. To ensure fast return, all assignments should be submitted on time. Late submissions will be penalized 20% for any part of the first 24 hours, 50% for any part of the second 24 hours, and 100% thereafter. You are free to discuss the assignments with your classmates, but you must write up the report all by yourself. Plagiarism cases will be reported to the university.
- **Communications:** This course will use Blackboard as the main communication platforms for class exercises and notifications. Students are required to check their Blackboard accounts on a regular basis.

_____

University Policies:

Students should review the University's policies regarding Disability-Related Accommodation; Diversity and Disability; the Religious Observances Notification and Policy; and Orange SUccess, which can be accessed via the Office of the Provost's website at: http://provost.syr.edu/.

Academic Integrity Policy

Syracuse University's Academic Integrity Policy reflects the high value that we, as a university community, place on honesty in academic work. The policy defines our expectations for academic honesty and holds students accountable for the integrity of all work they submit. Students should understand that it is their responsibility to learn about course-specific expectations, as well as about university-wide academic integrity expectations. The policy governs appropriate citation and use of sources, the integrity of work submitted in exams and assignments, and the veracity of signatures on attendance sheets and other verification of participation in class activities. The policy also prohibits students from submitting the same work in more than one class without receiving written authorization in advance from both instructors. Under the policy, students found in violation are subject to grade sanctions determined by the course instructor and non-grade sanctions determined by the School or College where the course is offered as described in the Violation and Sanction Classification Rubric. SU students are required to read an online summary of the University's academic integrity expectations and provide an electronic signature agreeing to abide by them twice a year during pre-term check-in on MySlice. For more information about the policy, see http://academicintegrity.syr.edu. Respect Intellectual Property Rights and cite all sources in your work. Any valid citation style may be used. The following link may be used for further information regarding appropriate citation styles: *http://researchguides.library.syr.edu/citation*.

Disability-Related Accommodations

Syracuse University values diversity and inclusion; we are committed to a climate of mutual respect and full participation. If you believe that you need accommodations for a disability, please contact the Office of Disability Services (ODS), disabilityservices.syr.edu, located at 804 University Avenue, room 309, or call 315.443.4498 for an appointment to discuss your needs and the process for requesting accommodations. ODS is responsible for coordinating disability-related accommodations and will issue "Accommodation Authorization Letters" to students as appropriate. Since accommodations may require early planning and generally are not provided retroactively, please contact ODS as soon as possible. Our goal at the iSchool is to create learning environments that are useable, equitable, inclusive and welcoming. If there are aspects of the instruction or design of this course that result in barriers to your inclusion or accurate assessment or achievement, please meet with me

_____

to discuss additional strategies beyond official accommodations that may be helpful to your success.

Religious Observances Notification and Policy

SU's religious observances policy, found at supolicies.syr.edu/emp_ben/religious_observance.htm , recognizes the diversity of faiths represented in the campus community and protects the rights of students, faculty, and staff to observe religious holy days according to their tradition. Under the policy, students should have an opportunity to make up any examination, study, or work requirements that may be missed due to a religious observance provided they notify their instructors no later than the end of the second week of classes through an online notification form in MySlice listed under **Student Services/Enrollment/My Religious Observances/Add a Notification**.

Educational Use of Student Work

Student work prepared for University courses in any media may be used for educational purposes, if the course syllabus makes clear that such use may occur. You grant permission to have your work used in this manner by registering for, and by continuing to be enrolled in, courses where such use of student work is announced in the course syllabus. I intend to use academic work that you complete this semester in subsequent semesters for educational purposes. Before using your work for that purpose, I will either get your written permission or render the work anonymous by removing all your personal identification.

Course evaluations:

There will be an end of course evaluation for you to complete this term. This evaluation will be conducted online and is entirely anonymous. You will receive a notification from the Syracuse University Office of Institutional Research & Assessment (OIRA) department in your email account with the evaluation website link and your passcode. Please take the time and fill out this evaluation as your feedback and support of this assessment effort is very much appreciated. The school carefully reviews ratings and comments that you submit, and these factor into decisions about course, program and instructor development.

Use of Blackboard

This course involves the use of Syracuse University's Blackboard system as an online tool. The environment is composed of a number of elements that will help you be successful in both your current coursework and your lifelong learning opportunities. To access Blackboard, go to the following URL: http://blackboard.syr.edu   Use your Syracuse University NetID & Password to log into Blackboard. This specific course will appear in your course list.

_____

If you have questions regarding technical aspects of Blackboard, please submit a help ticket to the iSchool dashboard at My.iSchool Dashboard at https://my.ischool.syr.edu Log in with your NetID, select "Submit a Helpdesk Ticket," and select Blackboard as the request type. The iSchool Blackboard support team will assist you.

Syracuse University
IST 707 Data Analytics
COURSE SYLLABUS
2019 Spring Semester

_____

Course Schedule:

Week	Date	Topic	Textbook Readings (Pang et al.)	Item due
1	January 15, 2019	Introduction to Mining	Ch.1	
2	January 22, 2019	Data Exploration	Ch. 2, 3	**HW1** – due before class begins
3	January 29, 2019	Association rules	Ch. 6	**HW2** – due before class begins
4	February 5, 2019	Clustering algorithm: k-Means	Ch. 8.1-8.2	**HW3** – due before class begins
5	February 12, 2019	Clustering: HAC	Ch. 8.3	**HW4** – due before class begins
6	February 19, 2019	Classification algorithm: decision tree **Final Project Presentation**: Assign Final Project Groups	Ch. 4.1-4.3	
7	February 26, 2019	Classification: Model evaluation	Ch. 4.4-4.6	**HW5** – due before class begins
8	March 5, 2019	Classification algorithm: naïve Bayes		

9	March 12, 2019	**SPRING BREAK** **NO CLASSES**		
10	March 19, 2019	Classification algorithm: kNN and SVMs **Final Project Presentation**: Project idea proposal and presentation: Your idea proposal should include an overview of the data mining problem, the data set you will use and its availability, and your proposed data mining approach. (Check point 1)	Ch. 5.3	**HW6** – due before class begins
11	March 26, 2019	Classification: text mining **Final Project Presentation**: show preliminary results and major challenges (Check point 2)	Ch. 5.2, 5.5	**HW7** – due before class begins
12	April 2, 2019	Classification: application review **Final Project Presentation**: Review for all groups (check point 3)		**HW8** – due before class begins
13	April 9, 2019	Final project presentation		Presentation slides
14	April 16, 2019	Final project presentation		Presentation slides
15	April 23, 2019	Final project presentation		Presentation slides

_____