GRADUATE COURSE SYLLABUS

IST 707 Applied Machine Learning
Tuesday, 2:00 – 4:45 PM, HH 013

Instructor	Joshua Introne	Phone	315.443.8975
Office	HH 326	E-mail	jeintron@syr.edu
Office Hours	By Appointment		

Course Description:

General overview of industry standard machine learning techniques and algorithms. Focus on machine learning model building and optimization, real-world applications, and future directions in the field. Hands-on experience with modern data science packages.

Additional Course Description

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

The topics of the course will include the key tasks of data analytics, including data preparation, concept description, association rule analytics, classification, clustering, evaluation and analysis. Through the exploration of the concepts and techniques of data analytics 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 on-line.

Prerequisite / Co-requisite:

IST 687, OR IST 387 with a minimum grade of B or higher

Audience:

Graduate Students (707)

Credits:

3 Credits

Course Fees and/or Costs

None

Learning Objectives:

After taking this course, students will be able to:

- Document, analyze, and translate data analytics needs into technical designs and solutions.
- Apply data analytics 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.

Required Texts / Supplies:

- Brett Lantz (2019) Machine Learning with R (third edition).

Texts / Supplies – Additional:

- Pang-Ning Tan, Michael Steinbach, and Vipin Kumar (2005) Introduction to Data Analytics. (Free sample chapters available at authors' website http://www-users.cs.umn.edu/~kumar/dmbook/index.php)
- Tom Mitchell (1997) Machine Learning. (http://www.cs.cmu.edu/~tom/mlbook.html)

Software

- Orange Data Mining

Note to students: We will mainly use Orange in class, but students can use Python or R if they are more comfortable with that. Regardless of platform used, all assignments must be submitted as well-formatted PDF reports (see below).

Course Requirements and Expectations:

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	Weight %

Homework assignments	40
Group Project	40
Participation	20
Total	100

- **Homework assignments:** Assignments must be professionally prepared and submitted electronically to Blackboard. All assignment writeups should be submitted as a well-formatted PDF file in the dropbox folder on Blackboard, along with any supporting code (e.g., Jupyter Notebook, Knitted RMarkdown file, or Orange workflow file). Unless otherwise specified, all homeworks should include the following elements:
  - *An introduction stating the purpose of the assignment*
  - *Documentation of methods used*
  - *Findings that are clearly delineated*

  Do not include long printouts of tables or debugging output. Do not include code in the written report unless there is a specific need to do so. I will take points off for mis-spellings / poor grammar / writing.

- **Group Project:** The objective of the project is to work as a team to apply concepts taught in this class to solve a data analytics problem. Groups of 2 / 3 people will self-select. There are three presentation checkpoints for the group project, and each checkpoint is worth a certain number of points.
  - **Checkpoint 1 (5 points)**: Project idea proposal and presentation: Your idea proposal should include an overview of the data mining problem, with specific attention to the *business case* (how will your project benefit stakeholders?), the data set you will use and an initial descriptive analysis, and your proposed data mining approach, and proposed approach, and how you plan to divide up work in the team.
  - **Checkpoint 2 (10 points)**: Project progress presentation: show preliminary results and major challenges. Should have some initial results.
  - **Checkpoint 3 (10 points)**: final project presentation: Last round of feedback before submitting the final report.

  Presentations will be graded based equally on the quality of the presentation (Were all team members present? Was the material communicated effectively?) and the degree to which the material being presented is sufficiently advanced given the stage of completion. The final project report is worth 15 points.

- **Participation**: Participation includes engagement in class discussion, presentations, and lab work. Lab work that is turned in at the end of most classes is evidence of participation. You must be present in class to receive credit for participation.

Grading:

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

Grading Table

Grades	Grade Points /Credit	Total Points
A	**4.000**	93-100
A-	**3.66**	90-92
B+	**3.33**	87-89
B	**3.00**	83-86
B-	**2.66**	80-82
C+	**2.33**	77-79
C	**2.00**	73-76
C-	**1.66**	70-72
F	**0**	0-69

* source: http://www.syr.edu/registrar/students/grades/faq.html

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

Use Graduate syllabus template with the appropriate grading table

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**: *All homework is due by midnight before then next class unless otherwise specific.* To ensure fast return, all assignments should be submitted on time. One-hour grace period is given to accommodate any incidents around deadline. Late policy will be enforced starting from the second hour. You will lose 10 points off your homework grade once it is marked late, and an additional 10 points for each week it is late thereafter. Homework that is more than 3 weeks late will not be accepted.
- **Collaboration**: 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.
- **Attendance**: Attendance *in person* is required unless there are extenuating circumstances and other arrangements are made (e.g., remote

participation may be possible if there is a health reason). I do not take attendance, but lab work is evidence of attendance.

- **Lab work**: Most classes will involve some lab work. I do not grade this lab work, but if it is not done / done poorly you will lose points on your participation grade.

Note to students: Given the diversified background of data science students, one textbook might not fit everyone. If you like rigorous algorithm presentation, I recommend Mitchell's classic book on Machine Learning. Lantz's book offer a somewhat less technical treatment, and can be accessed electronically through the library with your SU id (or on Blackboard)

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 black box. Don't treat an algorithm as a black box. Try see through them.
- Critical thinking: Data analytics 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 analytics algorithms work.

Typical Course Schedule

*Unless otherwise specified, homeworks are due in the dropbox on blackboard by midnight (12AM on the date marked) before the next class, or else will be marked *late*. All readings are from the Lantz textbook, 3rd edition.*

Week	Date	Topic	Readings	Assigned	Due
1	8/30	Introduction to Data Analytics	Pages 1-23	HW1	
2	9/6	Data Exploration	Pages 44-62	HW2	HW1
3	9/13	Association Rules	Pages 262-285	HW3	HW2
4	9/20	Clustering	Pages 288-311	HW4	HW3
5	9/27	Classification: decision tree	Pages 126-135 Pages 148-151 Pages 155-157 Pages 135-143	HW5	HW4
6	10/4	Classification: model evaluation	Pages 313-345		HW5; Pick group (2/3)
7	10/11	Group Project: Idea Presentation			
8	10/18	Classification: naïve Bayes	Pages 89-97	HW6	
9	10/25	Classification algorithm: kNN, SVMs, random forest	Pages 66-76 Pages 241-249	HW7	HW6
10	11/1	Text Mining and NLP	TBA	HW8	HW7

11	11/8	Group Project: Preliminary Results			Draft report
12	11/15	Deep Learning 1	Pages 217-229	HW9	HW8
Thanksgiving Break 11/21 – 11/25					
13	11/29	Deep Learning 2			HW9
14	12/6	Group Project: Final Presentation			
	12/16				Group Project Report Due

SYRACUSE UNIVERSITY STUDENT POLICIES & SERVICES

iSchool Values

Excellence; Discovery & Innovation; Integrity; Diversity & Inclusion; Global Citizenship and Engagement

Syracuse University Policies

Syracuse University has a variety of other policies designed to guarantee that students live and study in a community respectful of their needs **and those of fellow students. Some of the most important of these concern:**

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. Faculty should use "ESPR" and "MSPR" in Orange Success to alert the Office of the Registrar and the Office of Financial Aid. A grade of NA is posted to any student for whom the Never Attended flag is raised in Orange SUccess. More information regarding Orange SUccess can be found here, at http://orangesuccess.syr.edu/getting-started-2/.

Students should also review the University's religious observance policy and make the required arrangements at the beginning of each semester.

Diversity and Disability

(ensuring that students are aware of their rights and responsibilities in a diverse, inclusive, accessible, bias-free campus community) can be found here, at: https://www.syracuse.edu/life/accessibilitydiversity/.

Religious Observances Notification and Policy

(steps to follow to request accommodations for the observance of religious holidays) can be found here, at: http://supolicies.syr.edu/studs/religious_observance.htm

Orange SUccess

(tools to access a variety of SU resources, including ways to communicate with advisors and faculty members) can be found here, at: http://orangesuccess.syr.edu/getting-started-2/

Disability-Related Accommodations

Syracuse University values diversity and inclusion; we are committed to a climate of mutual respect and full participation. There may be aspects of the instruction or design of this course that result in barriers to your inclusion and full participation in this course. I invite any student to meet with me to discuss strategies and/or accommodations (academic adjustments) that may

be essential to your success and to collaborate with the Center for Disability Resources (CDR) in this process.

If you would like to discuss disability-accommodations or register with CDR, please visit Center for Disability Resources. Please call (315) 443-4498 or email disabilityresources@syr.edu for more detailed information.

CDR is responsible for coordinating disability-related academic accommodations and will work with the student to develop an access plan. Since academic accommodations may require early planning and generally are not provided retroactively, please contact CDR as soon as possible to begin this process. https://disabilityresources.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.

Course Evaluations

At the end of the term, the iSchool will ask you to share course feedback through EvaluationKIT [https://coursefeedback.syr.edu]. Log in to EvaluationKIT using your NetID and password. Please take the time to share your feedback about this course and your experience in it; all ratings and comments are completely anonymous. The iSchool carefully reviews your feedback. Our instructors use this feedback to fine tune course delivery and instruction; our professors of record use this feedback to fine tune course content and assignments. All feedback is factored into iSchool 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, [http://blackboard.syr.edu] use your Syracuse University NetID & Password. This specific course will appear in your course list.

To search for answers to your Blackboard questions, visit the Answers self-help knowledge [https://answers.syr.edu/display/blackboard01/Blackboard]. If you have problems logging in or need assistance with Blackboard, contact the ITS Service Center at: help@syr.edu or 315.443.2677. The Syracuse University Blackboard support team will assist you.