Stevens Institute of Technology
School of Business
Syllabus
BIA656 Advanced Data Analytics and Machine Learning

Semester: Fall 2019	Day of Week/Time Th. 12.30-3PM,
Instructor name and contact information Germán Creamer, Babbio 637 gcreamer@stevens.edu	Office Hours: Th., 10.50 AM – 12.20 PM 3 PM – 4 PM Class Website: Canvas

Overview
The significant amount of corporate information available requires a systematic and analytical approach to select the most important information and anticipate major events. Machine learning algorithms facilitate this process understanding, modeling and forecasting the behavior of major corporate variables.

This course introduces statistical and graphical (machine learning) models used for inference and prediction. The emphasis of the course is in the learning capability of the algorithms and their application to several business areas.

Prerequisites: Basic course in probability and statistics at the level of MGT 620 or BIA 652 Multivariate data analytics.

Course Objectives
Students will:

- Learn the fundamental concepts of statistical learning algorithms.
- Explore existent and new applications of statistical learning methods to business problems, and to generic classification problems.
- Learn to solve analytical problems in groups and effectively communicate its results.

Relationship of Course to Rest of Curriculum
Students will have the opportunity to explore the main concepts of statistical learning that will be used in the applied modules of this program.

Learning Goals
By the end of this course, the students will be able to:

1. Understand the foundations of statistical learning algorithms
2. Apply statistical models and analytical methods to several business domains using a statistical language.
3. Recognize the value and also the limits of statistical learning algorithms to solve business problems.

Additional learning objectives include the development of:

1. Written and oral communications skills: students are required to communicate properly during the class discussions and project class presentations. Homeworks and project report should be presented “as if” they were submitted to a senior manager of a major corporation.

2. Solve a major analytical problem using large and heterogeneous datasets in a group project and communicate its results in a professional way.

Pedagogy
The class will combine class presentations, discussions, exercises and case analysis to motivate students and train them in the appropriate use of statistical and econometric techniques.

Required Texts
Foster Provost and Tom Fawcett, Data Science for Business, O'Reilly, 2013. (to get a discount on oreilly.com use this code: AUTHD) (PF)

Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, An Introduction to Statistical Learning with Applications in R, Springer, 2013 (link) (ISLR)

Trevor Hastie, Robert Tibshirani and Jerome Friedman, The Elements of Statistical Learning. Springer-Verlag, 2nd. Ed., New York, 2009 (link) (ESL).

Case Pilgrim Bank A (602104), Harvard Business School. You must register in the HBS website according to instructors' recommendations.

G. Creamer and Y. Freund, Automated Trading with Boosting and Expert Weighting Quantitative Finance, 4 (10): 401–420 (link)

Assignments
The course will have a main project and several assignments/cases of data analysis and several labs. The assignments must be submitted electronically through the course website before the beginning of the class of the assigned day. Each student must submit his/her own report. E-mail submissions will not be accepted.

For all the programming homeworks, you should submit two UNCOMPRESSED files: a report and a Python program organized by questions. Please do not copy and paste large parts of the Python program as part of the solutions. Create your tables with the output of your program and EXPLAIN the results. You can also submit a report as a Jupyter notebook saved as an HTML or pdf file. However, you must still submit the Python code file as an independent file.

If you have generated a Jupyter notebook and you want to download the HTML and Python file, you can go to: File --> Download as:

generate a Python file and an HTML or pdf and a python file. These are the two files that you have to submit.

If you want to improve your homework, you can resubmit it until the deadline.

Do not send sections of your code or ask a complex homework question by email. I cannot debug your program or write a long explanation by email. However, you are welcome to ask any questions about the homework or any other issue related to this class during class, after class or during the office hours.

Project
The project requires that participants build a decision support system (DSS) based on one of the methods explored in this course. Each project must be developed by groups of three students and they should present a project proposal at the middle of the semester.

PhD students should prepare an academic paper that counts as the final project for this course. The paper should be oriented to conferences such as "Innovative Applications of Artificial Intelligence Conference" or “International Conference on Information Systems.” The paper should also be based on a theoretical or applied exploration of one of the methods studied in this course or any other data analysis method approved by the instructor.

Grades

Assignment	Grade %
Assignments/cases/labs	11%
Team project	25%
Participation	4%
Midterm	30%
Final	30%
Total Grade	100%

The final grade will be determined according to the following scale:

A	100%	to	94%
A-	< 94%	to	90%
B+	< 90%	to	87%
B	< 87%	to	84%
B-	< 84%	to	80%
C+	< 80%	to	77%
C	< 77%	to	74%
C-	< 74%	to	70%
D+	< 70%	to	67%
D	< 67%	to	60%
F	< 60%	to	0%

Software
Python is the main software packages that will be used. If you are not proficient in Python, you should participate in the Python bootcamp offered by the school at the beginning of the semester.

Class policy
No late homework will be accepted.

Re-grades: If you dispute the grade received for an assignment, you must submit, in writing, your detailed and clearly stated argument for what you believe is incorrect and why. This must be submitted by the beginning of the next class after the assignment was returned. Requests for re-grade after the beginning of class will not be accepted. A written response will be provided by the next class indicating your final score. Be aware that requests of re-grade of a specific problem can result in a regrade of the entire assignment. This re-grade and written response is final; no additional re-grades or debate for that assignment.

Ethics and Cooperation: You are allowed to discuss lecture and textbook materials, and how to approach assignments.

You cannot share ideas in any written form: code, pseudocode or solutions. You cannot submit someone else's work found through internet or any other source, or a modification of that work, with or without that person's knowledge, regardless of the circumstances under which it was obtained, copied, or modified. Of course, no cooperation is allowed during exams.

The following statement is printed in the Stevens Graduate Catalog and applies to all students taking Stevens courses, on and off campus.

"Academic Improprieties

The term academic impropriety is meant to include, but is not limited to, cheating on homework, during in-class or take home examinations and plagiarism. The Institute has adopted a procedure to deal with such actions. An instructor of a graduate course may elect to formally charge a student with committing an academic impropriety to the Dean of Graduate Academics or to adjudicate the issue personally."

Consequences of academic impropriety are severe, ranging from receiving an "F" in a course, to a warning from the Dean of the Graduate School, which becomes a part of the permanent student record, to expulsion.

Reference: https://www.stevens.edu/provost/graduate-academics/handbook/academic-standing.html#PDG

Consistent with the above statements, all homework exercises, tests and exams that are designated as individual assignments MUST contain the following signed statement before they can be accepted for grading.

_____

I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination. I further pledge that I have not copied any material from a book, article, the Internet or any other source except where I have expressly cited the source.

Signature _____ Date: _____

Please note that assignments in this class may be submitted to www.turnitin.com, a web-based anti-plagiarism system, for an evaluation of their originality.

_____

Syllabus

Weeks	Topic	Readings	Assignment
8/29	Introduction to data science & Python	PF, 1 and 2	
9/5	Predictive modeling	PF, 3	
9/12	Linear regression	ISLR, 3	
9/19	Decision trees: From correlation to supervised segmentation	PF, 3 ISLR, 8.1	
9/26	Classification, Linear models	PF 4; ISLR, 4	
10/3	Support vector machines	PF, 4, ISLR 9	
10/10	Model performance analysis	PF, 5, 7 and 8	
10/17	Mean variance decomposition Combining models: Ensemble methods Algorithmic trading	ISLR, 8.2 ADTrees, Bagging, Random Forests Creamer and Freund https://papers.ssrn.com/sol3/papers2.cfm?abstract_id=937847	Project proposal (10/18)
10/24	Midterm		
10/31	Unsupervised learning: clustering	ISLR, 10, PF 6	
11/7	Bayesian probability	ESL, 6.6.3, 8.3, 8.6	
11/14	Markov models Hidden Markov models	ESL, 8.3, 8.6 and 17	
11/21	Poster presentations Application to marketing: Targeting consumer	ESL, 17 Rabiner's HMM tutorial {I,II,III} Case Pilgrim Bank	
12/4	Final exam		Final project report (11/27)

ISLR: Introduction to Statistical Learning with Applications in R (link)
ESL: Hastie, Tibshirani and Friedman, The Elements of Statistical Learning. 2009
PF: Provost and Fawcett, Data Science for Business