FINC 360-01
Special Topics in Finance
Machine Learning for Finance
Fall 2019

Professor: Dean Smith
Phone/Voice Mail: 914-523-2180
e-mail: smithdt2@cofc.edu
Office Hours: By appointment
Class: M 4:00-6:45 pm., Beatty 214

NOTE: It is your responsibility to read, understand and abide by all the course information and policies listed below. Failure to do so could result in you failing this course or being withdrawn from this course by your instructor or by the Finance Department. The course syllabus provides a general plan for the course; changes may be made in the instructor’s discretion.

Course Description
This course is designed to give the student an understanding of the use of machine learning (“ML”) techniques in modern financial contexts. Machine learning has revolutionized research and practice for financial markets and products including areas such as investment management, trading and risk management, credit analysis, market analysis, to name just a few. This course will take a very hands-on, practical approach to exploring these applications and use-cases. The emphasis will be on practical application and development of basic skills, and much less on ML theory, although we will discuss the history of ML, the strengths and weaknesses of ML as an analytic method, and current social implications of widespread, rapid adoption of ML in financial settings.

Prerequisites
Calculus -- Math 105 (B or better) or Math 111 or Math 120
Probability/Statistics -- Math 104 (B or better) or Math 250 or DSCI 232
Finance -- FINC 303
Recommended, but not required:
Linear Algebra -- Math 203 or Math 207

Programming Experience -- This course involves using computers to explore machine learning solutions to finance problems. It will require students to write computer code in the Python programming language. Prior experience with Python is not required, but students without any prior programming experience will be expected to devote the necessary time to mastering basic programming skills in connection with the coursework.

Course Overview
Machine Learning has been and continues to be widely deployed in financial settings. In this course we will explore a broad range of ML algorithms, techniques and tools in financial contexts. ML algorithms fall broadly into three classes: Supervised Learning, Unsupervised Learning and Reinforcement Learning. We will explore all three of these broad areas through examples of their application to financial datasets.
The general approach will be to present the mathematical and probabilistic foundations for a particular ML technique, discuss its applicability and advantages or disadvantages for solving specific problems or answering particular types of questions. Once the theoretical motivation has been established, we will then write short- to moderate-length programs in python to demonstrate the application of that ML technique on actual data. In other words, we will alternate back and forth between traditional lecture format and hands-on program development as we explore the various ML algorithms.

In addition to in-class coding work, there will be homework assignments that will require that you extend and further develop the examples introduced in class. These assignments will require significant effort outside of class and will typically be completed over a period of 2-3 weeks. All students have been assigned to a study group of 4-5 students. These groups have been balanced for mathematical background and programming experience. You will be expected to collaborate on homework assignments in much the same way that real-world development teams collaborate in business and research contexts. Some homework assignments will be submitted as a team work product, while for others each student will submit his or own work product.

Your group will also work on a semester long group project which you will present to the class near the end of the semester. Details on this will presented part way through the semester.

There will also be a final examination that you will be expected to complete individually. This final exam will consist of traditional multiple-choice questions, some short essay questions, and one or more programming assignments. The exam will be administered in a take-home format, and you will have at least a week to complete it.

It is important to note that this is entirely new class at CofC. The plan for the class will be flexible and subject to change as the semester progresses. We may add or subtract from the algorithms and ML techniques we explore depending on time and interest. If there are particular financial applications the enrolled students want to explore, we can do so, provided it is pedagogically appropriate. Moreover, this is intended to be a practical course that will provide students with a framework to understand and analyze the application of ML in real-world financial contexts. It is not possible to undertake this work without encountering some math, but there will be a minimum of proofs, or other types of purely mathematical problems students will be expected to solve. Instead, math will be used to explain and motivate discussion of various ML techniques in order to better understand their applicability and limitations.

**Course Objectives**

1. Understand how Machine Learning can be used in financial settings, including but not limited to investing, managing market risk, and assessing credit risk.
2. Understand the differences between the broad types of Machine Learning approaches–Supervised, Unsupervised and Reinforcement–and how and when to use each.
3. Learn how to assess the quality and accuracy of specific ML applications
4. Learn about the limitations of ML in financial contexts, and gain an appreciation for the potential for misuse, abuse or other adverse social effects.
**Course Learning Outcomes**

1. Gain an understanding of the process required to analyze financial data using ML techniques.
2. Develop skills in data preparation and exploratory data analysis.
3. Develop basic Python programming skills to enable acquisition of financial data sets and the creation and debugging of code to implement ML algorithms.
4. Develop the ability to read, understand and extend ML programs and applications written by others.

**School of Business Learning Goals**

1. Communication Skills: Students will demonstrate, via written and spoken word, the ability to communicate, effectively present, critique, and defend ideas in a cogent, persuasive manner.
2. Quantitative Fluency: Students will demonstrate competency in logical reasoning and data analysis skills.
3. Global and Civic Responsibility: Students will be able to identify and define social, ethical and economic challenges at local, national and international bank organizations.
4. Intellectual Innovation and Creativity: Students will be able to demonstrate their resourcefulness and originality in addressing extemporaneous problems such as the role of credit underwriting in the recent financial crisis.
5. Synthesis: Students will demonstrate the ability to integrate knowledge from multiple disciplines incorporating learning from both classroom and non-classroom settings in the completion of complex and comprehensive risk management tasks.

**Required Course Materials**

There is no textbook required for this course. We will make extensive use of public domain and open-source resources found online.

There are several *optional* books that you may find very valuable. The first is:


There is an electronic version of this text that I will make available to student via OAKS, but you may find the hard copy useful, as well. This book covers the most popular ML algorithms in great detail, and contains hundreds of code examples.

Another text you may find useful, particularly if you continue to work on ML applications after this class is:

*Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*, by Wes McKinney
We will be using Python extensively in this class, but you will not be expected to become expert in writing Python code. If you wish to proceed with ML, you will need to develop better Python skills on your own, and this book is excellent.

For any student looking to be seriously challenged, this new book explores topics on the cutting edge of the field. Be warned, this book contains difficult material:

*Advances in Financial Machine Learning*, by Marco de Lopez de Prado

**Other written materials** – Not surprisingly, there is a vast wealth or material available online. From time to time, we will make use of papers, news article, blog posts or other material from the internet.

**Online resources** – You will be expected to sign up for access to several free online resources, including the following:
- Piazza.com
- Github.com
- Quandl.com
- Kaggle.com
- Stackoverflow.com

We will use other online resources over the course of the semester, but these are essential.

**Laptop Computers** – We will make extensive in-class use of computers to cover the material in this class. It is essential that you bring a laptop to each class meeting, ideally one that is not more than a few years old. You will be required to install a significant amount of software on your machine, so it is essential you have the authority to install software on the machine (e.g. administrator privileges). You will also be downloading a significant amount of data and storing it locally, even if temporarily. Please make sure you have an absolute minimum of 5 GB of free space on your laptop. More would be highly desirable.

**Student Responsibility**
Because this class meets only once per week, there will be a greater than typical amount of work to be completed between classes. All students are expected to have completed all the assigned readings and exercises before the next class meeting. Specific readings and exercises for the next class meeting will be posted on the class webpage following each class.

Students are also expected to regularly check the class webpage for updates or changes to the
class plan. In addition, we will be using Piazza for communication and interaction with respect to the lectures, assignments, and other matters. Student will be expected to check piazza regularly.

**Attendance**
Students are strongly encouraged to attend all scheduled classes. Class participation with your study group will count materially toward your final grade, and it will not be possible to obtain your best grade if you miss classes. In addition, the course material is cumulative, with each class building upon skills and knowledge obtained in prior classes, and it will be exceedingly difficult to catch up. If you fall behind due to absences, you will have a very hard time with this class.

**Class Format**
The class format will include classroom lectures and discussions of assigned reading material, group projects/case studies and computer lab sessions devoted to the development of ML applications to actual financial datasets. Both the classroom lectures and the computer lab sessions will be conducted in the assigned classroom.

**Course Grading**
The grading weights are provided below. Homework assignments that are turned in late will be subject to a material penalty.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Contribution to Final Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework Exercises</td>
<td>60%</td>
</tr>
<tr>
<td>Group Project</td>
<td>15%</td>
</tr>
<tr>
<td>Final Exam</td>
<td>25%</td>
</tr>
</tbody>
</table>

Letter grades will be granted based on the numerical scores shown in Exhibit A, below.

**Assignments & Class Participation**
Homework assignments will be given following some but not all class meetings. Typical assignments will include multiple choice and short essay questions. From time to time, case studies or analytical problems will be presented. Homework assignments will generally be due before the start of the next class. Regular participation is an integral part of this class. Each student is expected to contribute by answering questions and presenting data analysis in class. During each session students will be selected to respond to questions. It is expected that each student come to class prepared to participate in class discussion and present his or her analysis. Questions at the end of each chapter will be used as a basis for discussion.

**College of Charleston Honor Code and Academic Integrity**
Lying, cheating, attempted cheating and plagiarism are violations of our Honor Code that, when identified, are investigated. Each incident will be examined to determine the degree of deception involved.

Incidents where the instructor determines the student’s actions are related more to a
misunderstanding will handled by the instructor. A written intervention designed to help prevent
the student from repeating the error will be given to the student. The intervention, submitted by
form and signed both by the instructor and the student, will be forwarded to the Dean of Students
and placed in the student’s file.

Cases of suspected academic dishonesty will be reported directly by the instructor and/or others
having knowledge of the incident to the Dean of Students. A student found responsible by the
Honor Board for academic dishonesty will receive a XF in the course, indicating failure of the
course due to academic dishonesty. This grade will appear on the student’s transcript for two
years after which the student may petition for the X to be expunged. The student may also be
placed on disciplinary probation, suspended (temporary removal) or expelled (permanent
removal) from the College by the Honor Board.

Students should be aware that unauthorized collaboration--working together without permission--
is a form of cheating. Unless the instructor specifies that students can work together on an
assignment, quiz and/or test, no collaboration during the completion of the assignment is
permitted. Other forms of cheating include possessing or using an unauthorized study aid (which
could include accessing information via a cell phone or computer), copying from other students’
exams, fabricating data and giving unauthorized assistance.

Research conducted and/or papers written for other classes cannot be used in whole or in part for
any assignment in this class without obtaining prior permission from the instructor.

Students can find the complete Honor Code and all related processes in the Student Handbook at
http://studentaffairs.cofc.edu/honor-system/studenthandbook/index.php

**Students Eligible for SNAP Accommodation**

Any student in the class who has a documented disability and has been approved to receive
accommodations through the Center for Disability Services / SNAP, please come and discuss
this with me during my office hours.
<table>
<thead>
<tr>
<th>Final Grade</th>
<th>Course Total Score Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>100-93%</td>
</tr>
<tr>
<td>A-</td>
<td>92-90%</td>
</tr>
<tr>
<td>B+</td>
<td>89-87%</td>
</tr>
<tr>
<td>B</td>
<td>86-84%</td>
</tr>
<tr>
<td>B-</td>
<td>83-80%</td>
</tr>
<tr>
<td>C+</td>
<td>79-77%</td>
</tr>
<tr>
<td>C</td>
<td>76-74%</td>
</tr>
<tr>
<td>C-</td>
<td>73-70%</td>
</tr>
<tr>
<td>D+</td>
<td>69-67%</td>
</tr>
<tr>
<td>D</td>
<td>66-64%</td>
</tr>
<tr>
<td>D-</td>
<td>63-60%</td>
</tr>
<tr>
<td>F</td>
<td>Below 60%</td>
</tr>
</tbody>
</table>