FIN 410-001 Data Mining and Machine Learning in Finance Syllabus

Instructor:
Dantong Yu

Phone:
973-596-3253
Email:
dtyu@njit.edu

Face-to-Face Lecture:
Tuesday and Thursday 1:00pm-2:30pm

Classroom Location:
TIER LECT 2

Office Hour Location:
CAB 2004

Office Hours:
12:00 pm-1:00pm, Thursday

TA Name:
Muntasir Tiash
TA Office Hours:
Thursday 12:00noon-1:00pm Thursday. CAB2018

Course Overview

This course provides an in-depth study of data mining and machine learning, focusing on finance applications. As the finance market becomes increasingly complex and data-driven, finance analysts and fund managers must make better and faster decisions using available data. Data mining & machine learning is an approach that uses powerful tools and techniques to unlock the value inherent in available financial market data. Data mining and machine learning now routinely helps investors uncover hidden patterns and correlations in data and gain insights to improve decision-making in the stock market. The course is practice-oriented and develops the required skills to apply machine learning in the stock market. Students will better understand the techniques for data mining and machine learning as well as gain hands-on knowledge of the contemporary analysis tools of data mining & machine learning. The course will enable students to understand better the major concepts, approaches, and techniques for data mining and machine learning. The included learning material provides adequate technical depth for students to understand how data-driven technologies work. Coverage includes data mining and machine learning processes, methods, and techniques; the role and management of data; tools and metrics; and integration with Big Data.

Purpose of the Course

Big data has become ubiquitous in virtually every area, in particular, finance analysis and market prediction. Data mining and machine-learning modeling are undergoing a shift from data scarcity to data abundance, with a re-thinking from traditional, naïve neural networks, towards deep learning models where the neural hierarchy is more rational, optimized, and informative. These two factors contribute to the stock market activities in a new data era. A new generation of financial analysts and market investors must interact with data to perform risk analysis and make investment decisions. The course builds on a foundation of statistical learning and analytical systems to prepare students who select finance technology concentration to use advanced data mining and machine learning tools and conduct stock market analysis, customer credit data evaluation, and other financial datasets.

Learning Outcomes

Our class emphasizes using data analysis tools to identify potential opportunities and risks and make data-driven financial decisions. The learning outcomes of this course aims at understanding and mastering state-of-the-art machine learning and data mining methods that belong to

ms3235@njit.edu   advanced data analytics capability and applying them for routine financial data analysis tasks.

Required Textbook (Online Opentext Book, No Purchase is Needed)

An Introduction to Statistical Learning with Applications in Python, Thir Edition, Authors: James, G., Witten, D., Hastie, T., Tibshirani, R.

https://hastie.su.domains/ISLP/ISLP_website.pdf

The Python Notebooks are in www.statlearning.com/resources-python

This book is an open-source and open book. No need to purchase (I wish all of our books will be as green as this)!

Resources

Please note that this class is not a language class. It provides some basic help on Python Programming. Because we will deliver some course material in Python, you need to warm up your Python at a moderate level. Getting started earlier is always a good idea! Students have some Python programming background or want to learn Python and open-source software in data science.

1. Python: https://developers.google.com/edu/python/
2. Another well organized (http://bender.astro.sunysb.edu/classes/python-science/)
3. Scikit Learning: Machine Learning in Python (Links to an external site.)
4. IPython https://ipython.org/
5. You can download all the above Python stuff at this site!
6. You can download all companion Python codes here: https://dtyu@bitbucket.org/dtyu/fintech410.git

Prerequisites

This course is not a computer science data mining and machine learning class. You do not have to be a fluent programmer to take this course. You need to understand basic statistics (mean, variance, standard error, F-test, p-value), and you are willing to work with some software packages, such as Python Scikit Learning.

MATH 111:	Calculus I	
MATH 135:	Mathematics for Business	
FIN 310:	Data-Driven Financial Modeling	

Grading Scale

Grade	Points	Percentage
A	4.0	92.00-100.00
B+	3.5	86.00-91.99
B	3.0	80.00-85.99
C+	2.5	75.00-79.99
C	2.0	68.00-74.99
F	N/A	0.00-67.99

Course Schedule

Week1: Introduction (September/05)
- Introduction to the course.
- Statistical Learning and Demonstration of the Powerfulness of Financial Data.
- Briefly Introduction to Python Notebook.

Week 2: Statistical Learning (September/11)
- Supervised learning vs. unsupervised learning
- Why estimate f(x)?
- Assessing model accuracy

Week 3: Linear Algebra (September/18)

Week 4: Data Pre-Processing (September/25)

Week 5: Linear Regression (October/02)
- Simple linear regression
- Multiple linear regression
- The Marketing Plan
- A comparison with other simple models

Week 6: Classification Models (October/09)

- Logistic regression
- Linear Discriminant Analysis
- A comparison of classification Methods
- Python Notebook for Classification Model

Week 7: Resample Methods and Parameter Tuning in Supervised Models (October/16)

- Cross-Validation
- The Bootstrap
- Training Yahoo stocks with many parameters

Week 8:
- Mid-term Exam and Project 1 Review (**October/24**)
- Financial Statistical Learning Practice

- Machine Learning to Evaluate Your Financial Model

Week 9: Linear Model Selection and Regularization (October/30)

Week 10 Tree-based methods (November/06)
- Decision Trees
- Random Forests
- How to build your portfolios

Week 11: Support Vector Machines (November/13)

- Support Vector Machine with Linear Kernels
- Support Vector Machine with non-linear Kernels

Week 12 Unsupervised Learning (November/20/2023, November/27/2023)
- Clustering algorithm Deal with high dimensional complex data
- Principal Components Analysis

Week 13 Neural Networks and other machine learning models (December/04)
- Single Layer Neural Networks
- Multi-layer perceptron
- Convolutional Neural Networks

Week 14 Neural Networks (December/11)
- Recurrent Neural Networks
- Transformer

Week 15 Review for Final Exam

Course Requirements and Grading

Grading: Your grade will be determined as follows;
   25%-Project
   50%-Midterm exam and Final exam
   20%-Quiz
   5% -Class participation

Projects
There are two Individual Projects.
Here are some reasons why these projects are essential for you, besides that they are assigned.
- They allow you to research a topic relevant to the field of data mining and machine learning. In so doing, you come across many sources of information that may be useful to you later on.
- You practice synthesizing algorithms and data from many public sources to produce business results and knowledge. You should write a report in an original manner that discusses in-depth your analysis results and serves as a well-written and persuasive business document.

- Because the completed assignment has a due date, you will learn to work with timelines and meet predefined deadlines.

Homework and Individual Assignments

There will be two individual homework assignments, as described below. Here are some reasons why the homework assignments are vital for you, besides that they have been assigned.
- They allow you to explore a question you find exciting and intriguing and may help you clarify your future aspirations. This should be a very significant reason.
- You demonstrate to others that you master data mining and machine learning algorithms to perform fundamental data analysis. In that short write-up, selection, and comment, you should strive to make a powerful impression on others, especially as a value-adding member of the class.

Late Work and Make-Up Exams

All assignments are expected when due, as stated in your syllabus

Sharing Information

Students are free to discuss assignments with their colleagues. However, they should not take any written (electronic or otherwise) record away from the discussion. This applies when the project is supposed to be an individual effort or whenever two teams discuss common problems they encounter (inter-group collaboration). After the discussion, it is advisable to engage in at least half-hour of the non-course-related activity before working on the assignment. This will ensure that students can reconstruct by themselves what they learned from the discussion.

Student Conduct

Academic Integrity is the cornerstone of higher education and is central to the ideals of this course and the university. Cheating is strictly prohibited and devalues the degree that you are working on. As a member of the NJIT community, it is your responsibility to protect your educational investment by knowing and following the academic code of integrity policy that is found at: http://www5.njit.edu/policies/sites/policies/files/academic-integrity-code.pdf.
Please note that my professional obligation and responsibility is to report any academic misconduct to the Dean of Students Office. Any student found in violation of the code by cheating, plagiarizing, or misusing any online software will result in disciplinary action. This may include a failing grade of F and suspension or dismissal from the university. If you have any questions about the code of Academic Integrity, please contact the Dean of Students Office at dos@njit.edu.

The student with Disabilities Codes

NJIT adheres to section 504 of the Rehabilitation Act (ADA) of 1990. Appropriate accommodations are provided at no cost to the student. If you have any questions or would like additional information, please contact Dr. Phyllis Bolling, Center for Counseling and Psychological Services (C-CAPS), Campbell Hall, (entry-level), room 205, (973) 596-3420. For further information, visit the Student Disability Services website.