FINTECH 540 - Machine Learning for FinTech

General Information

Instructor:	Alessio Brini
Class meetings:	Tuesday 10:15-11:30, Thursday 10:15-11:30
Room meetings:	Teer 115
Office Hours:	By appointment
E-mail:	alessio.brini@duke.edu
Teaching assistant:	Yang Liu
TA E-Mail:	yang.liu@duke.edu

Course Objectives

This course will let you dive into the fast-paced world of **Machine Learning (ML)** through the lens of **Financial Technology (FinTech)**. We will cover ML's past and current trends while focusing on its practical applications in the ever-evolving FinTech landscape.

Our journey will traverse the intricacies of various **classification and forecasting algorithms** - the cornerstones of most machine learning applications - alongside exploring advanced, cutting-edge techniques, such as **generative machine learning** and **reinforcement learning**. The class objective is to transform the students into proficient practitioners or researchers who know the 'what' and the 'why' of applying these tools in diverse financial contexts.

This **hands-on course** provides practical (**coding**) experience and enables the students to comprehend the wide-reaching applications of machine learning across **finance**. The lessons will equip them to articulate concise, well-informed perspectives, empowering them to devise solutions that leverage these tools - skills that will hold them in good stead in their professional journey.

By the end of this course, the students should gain an in-depth understanding of various **machine learning algorithms**. They will be adept at selecting the most suitable one to **tackle financial challenges**. We will take a practical approach, examining various use cases and implementations within finance. Supplementary readings will be provided for each class to facilitate in-depth discussions and enrich your learning experience.

Whether your interest lies in finance, innovation, or financial technology, this course will serve as a sturdy springboard to catapult your understanding and expertise. In today's job market, machine learning prowess is a hot commodity - browse any job site for major banks, insurance companies, or FinTech firms, and you'll see why.

Specific Learning Outcomes

Upon successful completion of the course, students will have gained the ability to:

- Unpack the fundamentals of ML, its unique value proposition, and its distinctive qualities compared to other prevalent statistical techniques.
- Decipher and apply basic ML principles to real-world financial problems.
- Formulate and structure complex problems for effectively applying ML solutions.
- Differentiate between various ML techniques and determine the most suitable ones to address specific financial challenges.
- Assess and articulate the situations where ML adds value, recognizing that not all financial scenarios will benefit.
- Implement ML problem solutions, interpret the results, and translate these findings into actionable business strategies.

Essential Prerequisites

- *Basic understanding of finance*: Familiarity with financial terms, concepts, and systems is essential, given the course's focus on FinTech.
- *Introductory Statistics and Probability*: Students should have a foundational understanding of statistical concepts and methods, which are crucial to understanding machine learning algorithms.
- *Introductory Linear Algebra and Calculus*: These subjects often form the mathematical backbone for machine learning theory.
- *Programming experience*: The course is hands-on and includes coding assignments; students should have prior experience with a programming language, ideally Python, as it is widely used in machine learning.

- *Basics of Machine Learning*: Some exposure to machine learning concepts could be beneficial, although not strictly necessary, as the course provides a comprehensive field view.
- *Dedication and Curiosity*: Given the fast-paced nature of the FinTech industry and machine learning, a keen interest and commitment to stay updated are critical for success in this course.

Grading

Students will participate in a project centered on a practical Machine Learning application for FinTech. The project will involve a presentation in the last two weeks of the class. In addition, there will be graded coding assignments to help students consolidate their newly acquired skills. We will also conduct a midterm and final exam to assess the progress. Those exams will include coding assignments and multiple-choice/open questions.

Students must be prepared to dive into the reading material as part of the course curriculum. Every lecture could involve pop quizzes on assigned readings to ensure the students are absorbing the knowledge effectively.

Assignment	**Percentage**
Homework (4)	30%
Participation	5%
Machine Learning Project	35%
Midterm (coding exercises and multiple choice/open questions)	15%
Final (coding exercises and multiple choice/open questions)	15%

Table 1: Grading Distribution

Final letter grades are assigned based on the following scale (with slight modification as described below):

Before assigning letter grades, the professor may alter the scale by lowering the threshold for a certain grade (e.g., making a B- span 79.5–83 instead of 80–83). Such a change is solely at the instructor's discretion and occurs when the change results in a letter grade more accurately, reflecting the quality of the student's work and effort.

Range	Score	Grade
A-range	>97	A+
	93–96	A
	90–92	A-
B-range	87–89	B+
	83–86	B
	80–82	B-
C-range	77–80	C+
	73–76	C
	70–72	C-
F	<70	F

Table 2: Grade Conversion

Attendance and Participation Policy

This course convenes twice a week. Given the sequential nature of the curriculum, each class is crucial to fully grasping the course material. Therefore, missing more than one class during the semester without valid justification could significantly impact your course performance (the final grade will be lowered by one entire letter).

In-person attendance is mandatory to ensure active participation and practical learning. Personal name tents will be provided at the start of each lecture to encourage community building and facilitate interaction. Your commitment to consistent attendance will be vital for your success in this course.

The participation component of your grade is solely at the discretion of the instructor and relies on actively participating in the class as well as promptly replying in case the instructor asks questions about the readings.

Final Exam

The final exam will follow the Duke University schedule available on Duke Hub. It will be on December 16th, from 2 pm to 5 pm in Teer 115. The exam will take place in person, except for online students. No exception will be granted to the in-person student, so the students have to keep in mind that their physical presence is required that day.

Missing a Deadline

While it is understandable that unforeseen circumstances might cause you to miss a deadline occasionally, it's important to note that timeliness is an essential aspect of this course. If you cannot submit your assignment on time, you may submit it afterward. However, be aware that the grading weight of your assignment will diminish proportionately with the delay. Refer to the chart below to understand how the delay impacts the potential grade for your assignment.

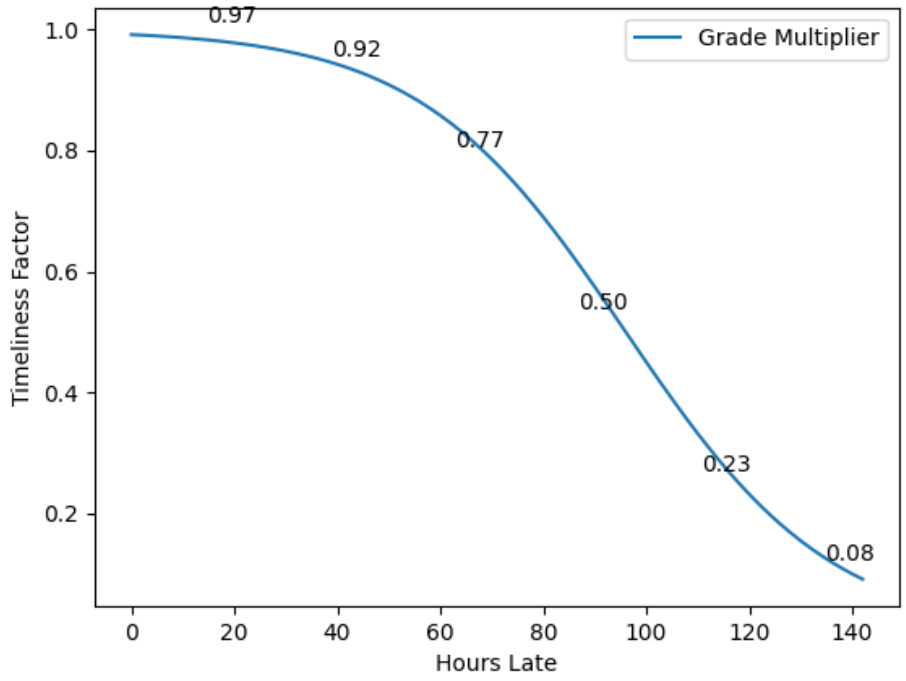

Figure 1: This graph represents the relationship between the penalized grade and the number of hours late. The penalized grade is calculated as the original grade multiplied by a decay factor. The decay factor is determined by the equation $\frac{1}{1+e^{0.05\cdot(hours-96)}}$ where $hours$ is the number of hours late for the assignment. As $hours$ increases, the grade decays exponentially, impacting the overall penalized grade.

Topics and Readings for the Course

Preliminary Information

The material will be uploaded to the dedicated Sakai website. We will also have a dedicated Microsoft Teams Group where students can ask questions anytime regarding the class material, assignments, and team projects.

All the readings in this document are mandatory, and the specified date indicates when they are due. At the beginning or during the class, after the reading is due, the instructor may ask questions or conduct pop quizzes to test the general understanding of the material. It is essential to come prepared and engage with the readings to maximize your learning experience in the class. Some more readings will be added during the course of the semester. It is the student's responsibility always to check the updated version of the syllabus on Sakai.

The specific details of the two exams and the group project will be provided during the semester. As we progress through the class, you will receive comprehensive information about these assessments' format, content, and deadlines. This approach allows us to tailor the evaluations to the topics covered and ensure a fair and meaningful assessment of your learning progress. Stay tuned for further updates on the examination and project guidelines.

Textbooks

The course readings will be sourced from the prescribed textbooks and online articles below. You can locate the specific readings assigned for each week in the course syllabus. This structured approach will facilitate systematic learning and keep you on track throughout the course. We will be using three free online books as a reference for the course:

1. Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville.
2. The Elements of Statistical Learning: Data Mining, Inference, and Prediction by Trevor Hastie, Robert Tibshirani, and Jerome Friedman.
3. Reinforcement Learning: An Introduction by Richard S. Sutton and Andrew G. Barto (Only used for the Reinforcement Learning part).

WEEK 1: INTRODUCTION TO MACHINE LEARNING FOR FINANCE: WHAT IS IT, ITS HISTORY, HOW IS IT REALIZED

Topics:

- What is Machine Learning?
- What are the Machine Learning Paradigms?
- What is the motivation behind Machine Learning in finance?
- Historical context of Machine Learning in Finance
- Current environment of Machine Learning in Finance, i.e. problems being solved for

Readings for August 31st:

To read after the first lecture on Tue 29th of August, and before the second lecture on the 31st of August.

- Deep Learning book – Chapter 1 Introduction
- The Elements of Statistical Learning book - Chapter 1 Introduction
- Will Machine Learning Transform Finance - Charles Elkan, Goldman Sachs' global head of machine learning
- The Pitfalls of Econometric Analysis Lecture 1 – Marcos Lopez De Prado presentation

WEEK 2: PRINCIPLES OF MACHINE LEARNING AND TOOLS FOR THE CLASSES

Topics:

- Probability use in machine learning: sources of uncertainty
- Linear algebra calculations
- Some of the general principles of machine learning
- Python Coding

Readings for September 5th:

- Deep Learning book – Chapter 2 Linear Algebra
- Deep Learning book – Chapter 3 Probability and Information Theory
- Deep Learning book – Chapter 5: from sections 5.1 to 5.4 Machine Learning Basics

DO NOT memorize formulas; we need to understand their concepts.

These readings will provide context for machine learning modeling. Since much of the class delves into probability theories and linear algebra, you must understand the basics. We will also do a lecture dedicated to Python to refresh the coding skills needed throughout the class.

Readings for September 7th:

- Review the Jupyter Notebook *Python Review* to refresh your Python skills. It provides a comprehensive overview of the main building blocks we will need throughout this class, including Python built-in data types, Numpy, and Pandas.
- A rough idea of what is an exploratory data analysis (EDA)
- EDA example on a housing prices dataset

Please come prepared to actively participate during this class, as we will conduct an Exploratory Data Analysis to reinforce the understanding of critical Python skills.

WEEK 3: LINEAR MODELS FOR REGRESSION AND CLASSIFICATION

Topics:

- Example of EDA for a financial application
- Preprocessing method before applying any model
- Introduction to Linear Models: Univariate and Multiple Linear Regression
- Logistic Regression and Multinomial Logistic Regression (Softmax Regression)
- Regularization Techniques: LASSO and ElasticNet

Readings for September 12th:

- Deep Learning book – Chapter 5: section 5.7 and 5.10
- The Elements of Statistical Learning book - Chapter 3: from section 3.1 to 3.2 Linear Methods for Regression, 3.4, 3.5
- Some examples of spurious correlations can be found at this link. These are curious examples that should motivate our in-class discussion regarding the need for detrending variables.
- Detecting stationarity in time series data
- Stationarity for financial returns
- Walasek, R., & Gajda, J. (Year of Publication). Fractional differentiation and its use in machine learning. International Journal of Advances in Engineering Sciences and Applied Mathematics 13.2-3 (2021): 270-277. (Uploaded on Sakai)

Readings for September 14th:

- The Elements of Statistical Learning book - Chapter 4: until section 4.4 (included)

Understanding linear models in a machine learning-focused class provides a strong foundation, interpretability, and real-world applicability, an essential stepping stone to more complex and sophisticated techniques.

Assignment on linear predictive model released on September 15th: due to September 23th

WEEK 4: SUPERVISED NONLINEAR MODELS: TREE BASED METHODS

Topics:

- Tree-based methods with applications in Python
  - Regression/Decision trees
  - Bagging methods (Random Forest)
  - Boosting methods (Adaboost and XGboost)
- Tree visualization and interpretation
- Credit card default problem solved by using tree-methods

Reading for September 19th:

- The Elements of Statistical Learning book - Chapter 9: Section 9.2 Tree-based methods

Reading for September 21th:

- The Elements of Statistical Learning book - Chapter 8: Section 8.7 Model Inference and Averaging
- The Elements of Statistical Learning book - Chapter 10: Boosting and Additive Trees

"Machine Learning" spans quite a large segment of approaches; practitioners must be familiar with these various tools. This familiarity will be valuable in the semester projects and future workplace considerations as solutions are considered and developed. For this reason, we start feeling the broad class of tree-based models.

WEEK 5: GUEST SPEAKER ON XGBOOST AND TREE ENSEMBLE

We will have a guest speaker, Matt Harrison, founder of Metasnake, who will give us a presentation of the practical usage of the XGBoost Python library.

Revise the readings from the previous week.

WEEK 6: HOW TO START A ML PROJECT AND MIDTERM EXAM

Project assignment will be released, and we will spend a lecture learning how to start organizing your project. Since the project will deal with time series, we will refresh the time-series analysis for your convenience.

The Midterm exam will be in the second lecture of the 6th week (October 5th). The exam will cover the material seen so far.

Topics:

- Guidance in articulating the project problem
- Detect all the needed steps to do good modeling and interpreting results

- Issues arising during problem formulation and their impact on the validity of the solution, including bias in variable selection.
- ML Project assignment with detailed instructions with presentations on the final weeks of class.

Assignment on tree-based model released on October 6th: due to October 14st

WEEK 7: UNSUPERVISED CLUSTERING METHODS AND INTRODUCTION TO NEURAL NETWORKS

Topics:

- Unsupervised Learning: clustering algorithms
- Understanding the reasoning behind the two different ML paradigms and grasping the differences concerning other supervised learning models.
- We will start introducing neural networks as a primary machine learning tool, which is helpful for various applications we will see until the end of the class.

Reading for October 10th:

- The Elements of Statistical Learning book - Chapter 14: Section 14.3 Cluster Analysis
- Deep Learning book – Chapter 5: Section 5.8.1 Principal Component Analysis

Readings for October 12th:

- Deep Learning book – Chapter 6 Deep Feedforward Networks
- Deep Learning book – Chapter 8 Optimization for Training Deep Models

WEEK 8: INTRODUCTION TO NEURAL NETWORKS (SECOND PART)

Revise readings on neural networks due to the previous week. This week is going to have a single class due to the Fall break.

Assignment on neural network modeling released on October 20th: due to October 28th

WEEK 9: RECURRENT AND CONVOLUTIONAL NEURAL NETWORKS

Topics:

- Introduction of specific architectures of neural networks
- Understanding the different use cases of each architecture
- Financial applications for those specific models.

Readings for October 24th:

- Deep Learning book – Chapter 10 Sequence Modeling: Recurrent and Recursive Nets

Readings for October 26th:

- Deep Learning book – Chapter 9 Convolutional Networks

WEEK 10: FIRST PROJECT CHECKPOINTS AND UNSUPERVISED LEARNING WITH NEURAL NETWORKS

Topics:

- First project checkpoint: We will have a class entirely focused on evaluating progress and solving problems
- Autoencoders for nonlinear unsupervised modeling

Readings for November 2nd:

- Deep Learning book – Chapter 14 Autoencoders

WEEK 11: REINFORCEMENT LEARNING AND SECOND PROJECT CHECKPOINT

Topics:

- Exploring how to apply reinforcement learning to a financial setting

- Reinforcement learning reasoning
- Use cases in finance
- Second project checkpoint on November 7th: We will have a class entirely focused on evaluating progress and solving problems before the submission of the final presentation

Readings for November 7th:

- Reinforcement Learning: An Introduction – Chapter 1 Introduction
- Reinforcement Learning: An Introduction – Chapter 3 Finite Markov Decision Processes

WEEK 12: NATURAL LANGUAGE PROCESSING LARGE LANGUAGE MODELS FOR FINTECH

Topics:

- Machine Learning and sentiment analysis
- Practical use of NLP in the industry – we will have a speaker in our class.
- Sentiment in all areas is complicated to assess quantitatively; we will explore methods for accomplishing this and apply these methods to actual data sets. This class will also review an experiment in sentiment analysis and how the operator controls the depth of the "sentiment" examined.
- Usage and application of LLMs for FinTech

Readings:

- TBD

WEEK 13: LAST PROJECT CHECKPOINT

There will be only one lecture since this is the week of Thanksgiving.

WEEK 14-15: PROJECT PRESENTATIONS

Topics:

- Machine Learning Projects to be presented by the students

FINAL EXAM

As per the Duke Hub date, the final exam will be in person on December 16^{th} from 2 pm to 5 pm in Teer 115.

Statistics Vocabulary - English/Chinese Translation:

The students can find an Excel spreadsheet on Sakai regarding the translation from English to Chinese of common statistics words for their convenience.

AI Policy

ChatGPT and other large language model (LLM) AI tools are being integrated into professional practice within the context of our Machine Learning for Fintech course. While the course emphasizes both programming and algorithmic thinking, it's important to recognize that using AI tools to deal with datasets can significantly impact the results, extending beyond mere programming proficiency.

In this course, we acknowledge the following points regarding the use of AI tools:

- Skillful prompt engineering, which involves creating effective prompts for LLMs, has been demonstrated to enhance productivity by generating relevant text and code.
- However, it is also observed that LLM-generated outputs can be erroneous, necessitating careful review and potential correction. The effectiveness of AI tools can vary between cases, sometimes offering valuable insights and, in other instances, leading to misunderstandings.
- While 'prompt engineering' is not a primary focus or a skill we explicitly teach in this course, understanding the limitations and capabilities of AI tools is essential. Balancing the time invested in developing prompts with core algorithmic and programming skills is a consideration.

- It is recommended to approach LLM-generated outputs with skepticism and prioritize the development of skills to evaluate the quality of AI-generated text and code.
- Given the nature of this course, which focuses on Machine Learning for Fintech, remember that the outcome is not just about the code you write but how you work with your dataset and set up the problem.

To facilitate your engagement with AI tools in an academically responsible manner, the following terms apply:

- If you opt to use AI tools for assignments, you are required to include a 'prompts' file alongside your submission. This file should detail the specific AI tool used, the prompts provided, and the corresponding tool-generated responses.
- Dedicated 'prompt engineering' channels will be available for sharing prompts and AI tool responses.
- Using AI tools without giving appropriate credit will be treated as a violation of academic integrity.

In this course, when it comes to using LLMs, the aim is to leverage those tools as supplementary aids to your learning experience. By being mindful of their strengths and limitations, you can effectively harness these tools to enhance your understanding of machine learning concepts within the context of financial technology.

Duke Reach:

As a student, you may experience various issues that could pose a learning challenge. These might include anxiety, stress, feeling down, problems with attention/focus, or loss of motivation. Duke has many helpful resources that all students will need at some point.

Visit https://studentaffairs.duke.edu/dukereach1/resources-assistance for more information.

Duke OIT

Visit https://oit.duke.edu for IT assistance.

Duke's Community Standard

Duke University is a community dedicated to scholarship, leadership, and service and the principles of honesty, fairness, respect, and accountability. Citizens of this community commit to reflect upon and uphold these principles in all academic and non-academic endeavors and to protect and promote a culture of integrity.

To uphold the Duke Community Standard:

- I will not lie, cheat, or steal in my academic endeavors;
- I will conduct myself honorably in all my endeavors; and
- I will act if the standard is compromised.