FIN 342 ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN INVESTMENT STRATEGIES

Our objective is to use machine learning to identify factors that appear to explain returns on stocks in the S&P MidCap 400 index, and use this analysis to make recommendations on stocks to construct a hypothetical investment fund.

For example, we know that small market capitalization stocks and stocks with a high book value of equity compared to market value (sometimes referred to as "value stocks") have historically earned higher returns than large market capitalization stocks and stocks with a low book value of equity compared to market value (sometimes referred to as "growth stocks"). So one example would be to use machine learning techniques to test the relationship between a company's market capitalization, its book-to-market ratio, and past returns, and make a recommendation on whether we should put a stock in the portfolio on the basis of its value versus growth characteristics.

That is just one example. Over the course we will investigate other factors that rely upon analyst earnings forecasts, industry classification, companies' past investments, and companies' historical profitability. Students will work in groups to perform analysis and ultimately make portfolio recommendations.

Jason Hall, B.Com (Hons) CFA PhD
uqjhall@umich.edu
USA +1 734 926 6989
Australia +61 419 120 348

December 21, 2022

Contents

Tables

Course objective

Our objective is to use machine learning to identify factors that appear to explain returns on stocks in the S&P MidCap 400 index, and use this analysis to make recommendations on stocks to construct a hypothetical investment fund.

For example, we know that small market capitalization stocks and stocks with a high book value of equity compared to market value (sometimes referred to as "value stocks") have historically earned higher returns than large market capitalization stocks and stocks with a low book value of equity compared to market value (sometimes referred to as "growth stocks"). So one example would be to use machine learning techniques to test the relationship between a company's market capitalization, its book-to-market ratio, and past returns, and make a recommendation on whether we should put a stock in the portfolio on the basis of its value versus growth characteristics.

That is just one example. Over the course we will investigate other factors that rely upon analyst earnings forecasts, industry classification, companies' past investments, and companies' historical profitability. Students will work in groups to perform analysis and ultimately make portfolio recommendations.

Basic information

Contact information

Instructor: Jason Hall
Email: uqjhall@umich.edu
Telephone: +1 734 926 6989
Office: R4443

Office hours: I will be hosting regular help sessions aside from our regular classes. These help sessions are optional. You can also call be on 734 926 6989, text me on or email me on uqjhall@umich.edu. It is often more efficient to just call me or text me with a question rather than wait for our next group meeting. I might be able to answer the question in a small amount of time.

Class schedule

Our official class schedule is from 2:30 pm to 3:50 pm on Mondays and Wednesdays in B0570. Please bring your laptop as class will be very hands on. Class time will be a combination of direct instruction and students working together in teams. **The objective of the course is to make stock recommendations in teams. You will be meeting with your teams in and out of class.**

Workload

At University of Michigan each hour of class time is meant to be accompanied by 2 to 3 hours of time spent in independent preparation (for example, reading, analysis, meetings).[1] In other words, each week you should devote 6-9 hours of outside class time to a 3 credit hour course. That's achievable, provided you are not over-committed to courses, clubs, recruiting and other events. Those thing are important. Just be realistic about how many hours there are in the day and consider how much you value quality over quantity.

You will need more time than you think to prepare a stock recommendation. Why? Because you are doing something for the first time and when we do things for the first time we make mistakes, or do things inefficiently, or aren't sure what to believe. You will write some code, tell me that the result says that if you buy stocks run by CEOs who like donuts you beat the benchmark. Then, I'll ask, "Are you sure about that, because it seems an unlikely reason for a stock to generate high returns? Maybe there is something we are missing."

Course material

Reading

I will ask you to read a custom book from Wiley which has two chapters: Multiple regression and Machine Learning. We don't have exams in the course, so this isn't a memorization exercise. Reading these chapters forms part of understanding why we perform particular sets of analysis. You can purchase the material from VitalSource or RedShelf using the links below.

https://www.vitalsource.com/custom/9781394197514

https://www.redshelf.com/book/2217944

Data and code

I will post data and code to the course website, along with other material (slides, teaching notes, videos). We will have students in the class with varying knowledge of finance, statistics and programming. So I will need to cater this material to people with different backgrounds during semester.

Programs

We will use Python to analyze financial data. You will have to use code provided, and make adjustments to the code to perform analysis. Some of those adjustments will be simply changing inputs (e.g. load a different file) while others will be more complicated (e.g. analyze data in a different manner). But you will not be graded on your ability to write code. Writing

---

[1] The Michigan Ross credit hour policy is located here:
https://www2.bus.umich.edu/sites/default/files/files/Final%20Credit%20Hour%20Policy.pdf
accessed December 23, 2022.

code is an intermediate step towards making an investment decision. You will be assessed according to the quality of your written work and presentations of investment decisions.

Agenda

What to deliver

Your task is, within teams, make recommendations on stocks to form a portfolio from the S&P MidCap 400 using machine learning techniques. Your recommendation will be for a stock within a specified FactSet Economic sector. You will repeat the process **three times**, and present your work either to me, small groups of other students, or to the entire class. Each recommendation will be accompanied by a **5-page written report** and **presentation**. I have deliberately kept the presentation schedule fluid because (1) I want you to have lots of practice (that's why there are three presentations) but (2) I am mindful that having the whole class watch every presentation could use a lot of student time when those students could be working on their own presentations (e.g. 40 presentations x 15 minutes each = 10 hours of presentations).

Teams

I will attempt to build teams in which the members have different skill sets, some with expertise in finance, some with expertise in Python, and some with expertise in accessing data from providers like FactSet, Bloomberg, and Refinitiv.

You will not stay in the same group for all three presentations. Sometimes team members don't work well together, because they are made up of people, not machines. I don't want to risk having people work together the entire semester who don't work well together, I also want to ensure you meet some new people. When students form their own teams, I generally observe teams of students with very similar backgrounds. So, I will form the teams, but with your preferences in mind. In other classes I just randomly assign team members. But in this class I want to combine team members with different skills.

Topics and quantitative techniques

In Table 1 I list the quantitative techniques we will cover. I have not outlined the precise timetable because I do not know the pace at which we will be able to cover these topics. Remember, our objective is to form a stock portfolio, as a conduit to developing quantitative skills with a particular focus on machine learning. The course is not on *all machine learning*, just like we don't have a single class that covers *all accounting,* or *all finance,* or *all marketing.*

Table 1. Topics

Topic	Quantitative technique	Reading
Understanding the S&P MidCap 400	Compilation of cumulative returns	S&P MidCap 400 Fact Sheet iShares Core S&P Mid-Cap ETF Fact Sheet
Defining our performance objective: Alpha versus a passive investment in the benchmark	Ordinary Least Squares (OLS) regression	**Multiple regression** (pp. 6-67)
Stock characteristics that are plausibly indicators of alpha (signals): value, quality, smart money, momentum		
Overfitting and testing results out of sample		
Introduction to machine learning		**Machine Learning** 1. Introduction (pp. 104–105) 2. What is machine learning? (pp. 106–110) 3. Overview of evaluating a machine learning algorithm (pp. 110–115)
Overfitting and penalized regression	Least absolute shrinkage & selection operator (LASSO) regression	**Machine Learning** 4. Supervised machine learning algorithms: Penalized regression (pp. 115–117)
Non-linearity in the relationship between signals and returns	Decision trees: Basic	**Machine Learning** 7. Classification & regression tree (pp. 120–124)
Non-linearity in the relationship between signals and returns, combined with over-fitting	Decision trees: Random forests	**Machine Learning** 8. Ensemble learning & random forest (pp. 124–138)
Non-linearity in the relationship between signals and returns, combined with over-fitting	Decision trees: Gradient boosting	
Sophisticated identification of relationships between signals and returns	Neural networks	**Machine Learning** 13. Neural networks (pp. 153–156) 14. Deep learning nets, and reinforcement & learning (pp. 156–167)
Comparison of machine learning techniques		**Machine Learning** 15. Choosing an appropriate machine-learning algorithm (pp. 167–170)

Assessment

Your grade will be determined by the overall quality of your written reports and presentations, with weights of 20%, 30% and 50% on recommendations, 1, 2 and 3. The reports and presentations are not entirely new pieces of work. You can use the same data, code, writing, and presentation materials three times, improving the quality of your work each time, and making a new stock recommendation. **Importantly, you will be graded on the quality of your writing and presentation, not on your Python code.**

You should specifically consider the two items below.

☐ The **economic rationale** underpinning your recommendation. You are going to tell the reader and audience to buy a stock on the basis of your quantitative analysis. But few people are going to invest in your fund purely because a computer says that stock returns are correlated with factor X. Why do you believe that future stock returns are likely to be correlated with factor X? If you can't answer that question it leaves open the possibility that your have observed a correlation in past data that will not continue in the future. For example, one can make the argument that stocks with a high book-to-market ratio will continue to earn high returns because investors persistently overpay for growth. Someone might counter with the argument that this will not continue because investors will learn from past mistakes. To which you can respond that investors overweight their own experience in decision-making and are over-confident, so new investors will continue to overpay for growth because these new investors believe that they can identify the quality growth companies, not like those other foolish investors who overpaid for growth that did not materialize. **We do not have to decide whose argument is correct. What matters is that you have a logical argument that underpins the result.**

☐ Your **ability to convey meaning** to the reader with words, tables, charts and overall presentation. This is not about grammar, and I recognize we have many students for whom English is not their first language. Hopefully, the team formation process can help overcome this challenge, and please show your work to other students to help improve the quality of your writing.

I will provide assistance and you can collaborate with other students as much as you like. Successful completion of three reports and presentations means that the you have demonstrated competency in the course and will be awarded at least a passing grade (C- or better). **There is only one grade constraint, according to the Ross grading policy. No more than 60% of students can receive a grade of A- or better.**[2] For example, if there is enrolment of 51 students, the maximum number of students who can be awarded grades of A- or better is 30.

---

[2] https://rossweb.bus.umich.edu/academics/wp-content/uploads/sites/2/2022/05/Fall-2020%E2%80%93Winter-2023-Modified-Grade-Distribution-Guidelines-BBAs.pdf accessed on December 22, 2022.

Administration

Recordings

I will record classes using Zoom. You will find it useful to replay particular sections of the presentation that explain code and finance theory. You will find links to class recordings on the course website. The recordings will be made available to other students in your section. Note, however, that the fact that recordings are not intended to be a substitute for class. Class absences are covered below.

Absence

COVID-19 is a serious illness. Vaccines are highly effective against severe disease caused by COVID-19. But it is still possible for a vaccinated person to become sick and transmit the virus to another person. So if you are not feeling well, stay home and join the class via the Zoom meeting. Staying home won't make any difference to your grade.

Illness is one reason to miss class. But there are other reasons for class absence and it isn't possible to rank those reasons from most to least important. So, I have adopted a simple absence policy. You can miss up to three classes without explanation, provided you attend an outside class help session within the week of class, or the following week.

What if you miss more than three classes? Then, we have a problem. If you miss more than three classes then we have a signal that there are other events that are impeding your ability to complete the course. Sure, you might be a very independent learner. But remember that it's very difficult for me to observe who is making good progress on their own versus someone who is getting further and further behind to the point at which it will be difficult to catch up. From my outside view, those two students can look the same. So if you miss that fourth class we need to meet one-on-one to work out whether you should withdraw from this class, withdraw from another class, drop some other commitment, or get some help in either finance, programming, writing, focusing or something else. **In short, if you miss that fourth ca call me straight away so we can fix the problem.**