Senior Project Proposal: An Attempt to Build a Successful Trading Model in the American Equities Universe

Advisor: Professor Stephen Slade

Introduction

The question of finding a way to make money in the markets has been around ever since there were markets, largely due to huge financial incentives. For decades, the average joe would dream of big financial success and try their hand in stock-picking. Some of them go on to be wildly successful, and we hear stories of these amazing day traders.

But were they just lucky? If there were just 100 people out of the roughly 80 million traders interacting in the market daily who invest in wildly volatile portfolios, we’d expect to hear maybe 50 stories of traders who did an amazing trade and got insanely rich. And that’s perhaps enough for the average joe to be biased into thinking there’s a method to this madness, and perhaps that they are smart enough to find the best stocks.

Yet it’s commonly heard that 90-95% or more of day traders fail and lose money; there do seem to be a lot of inherent disadvantages for retail investors. They almost always need to pay brokerage fees, the bid-ask spread, sometimes among other costs whenever they trade. They usually don’t have the infrastructure to crunch anywhere close to the amount of data that professionals can. They also have a lag in reading information, whether that’s the 15-minute delay for stock quotes on google or yahoo finance, or just being slower at digesting real world news. This means that whenever there is a very good trade to be made due to market inefficiencies, a professional probably already got to the trade. The opportunity is gone before the retail investor even has a chance to know it exists.
So here’s the literal million-dollar question. Is it possible to make money trading in the American equities universe as a retail investor?

Some recent literature claim there still is, and show methods of how they got to that conclusion. But this begs the question: why was that paper published in the first place? If there was a money-making opportunity, why didn’t the author simply go make money themselves? One explanation for our plethora of hopeful papers in finance is the author’s incentive in academia to overfit models and give positive, surprising results.

And what about all these technical indicators, or metrics for value investing that retail investors always talk about? Do they work? If they do, why are people talking about them so much? Shouldn’t they keep quiet and make their money?

The consensus among the 3 quantitative finance funds I’ve worked for seem to be that these “indicators” are mostly trash, and making a successful trading model as a retail investor is not possible, or at least extremely hard. But with all these success stories of different individuals and positive results in literature, it’s hard to deny that there’s a chance it’s still possible. Especially given that professional funds probably don’t want smart retail investors to know it’s still possible.

This semester I will explore this question: Is it still possible to make money as a retail investor? My solution will come in the form of machine learning, since it is the core of the quantitative financial world. It is what super successful funds, such as Renaissance Technologies, has been using to profit for decades. It is also sophisticated enough to capture the essence of almost all the strategies retail investors would use when they trade. Either I’ll come out with a successful trading model with machine learning, or, with a sophisticated enough model, show that most people can’t be successful systematically with the methods and resources they are exposed to.
Methods

1. Idea Generation:

I’ve gotten a lot of ideas from my experience in the financial field, and have features I want to try out. But I will also read through the literature, and take any good/allegedly successful features they have to add to my model.

2. Data Collection

The financial data to train my model will come from Google/Yahoo finance, or scraped off the web, or really any source recommended by papers I read or that I find. If I could scrape reports of company financial statements or earnings reports, that would be great too.

3. Setting up the problem

Once I have my ideas, and data, I will create my dataset of features and targets. The targets will be the returns of whatever stock/porfolio is in question on a medium to long time horizon, which can be set by a parameter. (There is a tradeoff – the shorter the time horizon, the less opportunity there is because professionals probably have gotten almost all the edge on trades in shorter time horizons. But with longer time horizons, there is less signal that can be found in a model). This is a regression problem. First, we’ll build a model that is able to predict the returns of individual stocks, and if those are successful, we’ll move onto building optimal portfolios.

4. Building the Model
This will be coded mostly if not all in Python (maybe parts of it, such as building the features, will be in C++ for performance reasons). I will be exploring many different textbook algorithms, such as

- PCA/Diffusion maps (for dimensionality reduction)
- Random forest
- Gradient Boosting Machine
- SVM (with various kernels)
- Neural Networks

And then I will move onto either other algorithms or start tuning hyperparameters. If different algorithms are good at different aspects of the problem, I’ll consider stacking. Depending on how big the models end up being, I can also consider automating the hyperparameter optimization (through algorithms like genetic algorithm).

5. Testing the Model

I’ll split my dataset into training and testing, where I’ll train on some number of trading days and test the accuracy on the next couple days (or whatever time period I set). The training set could also be a certain set of stocks, and then the test set could be different stocks in the same industry. Then I’ll try back-testing the model, building in lags (if I use very short time horizons/predictions), a bid-ask spread, and assuming some fees. I should come out with numbers indicating profit.

Steps and Timeline

1. Literature Review (Now to end of February)
a. Read through papers
b. Review the validity of their results
c. Write summaries of how they worked
d. Note the ideas/features that might work

2. Collecting Data (End of February to beginning of March)
   a. Find a data source and build a data set
   b. Build the needed features

3. Building the model (beginning of March to end of March)
   a. Code out the algorithms using the data
   b. First for individual stocks in technology sector, then general stocks, then we try to
      construct optimal portfolios
   c. Stacking, and hyperparameter optimization

4. Testing the model (end of March – mid-April)
   a. Backtest the model
   b. Note probable improvements and implement them

5. Writing the report (mid-April – early May)
   a. Write a more detailed report to present my methods and results

6. Thinking about future work, cushion time (May)

**Deliverables**

1. A final paper presenting my findings
2. Fully functioning coded trading model, either showing a successful strategy or showing
   that most retail investors probably wouldn’t have the resources to make a successful trading
   strategy