Name: James Pretti

College: Crown   **Phone:** (510) 325-5000   **Email:** jpretti@ucsc.edu

Mailing Address:

James Pretti
741 Solana Dr.
Lafayette, CA 94549

Project Title: The Money Maker
Project Category
    Research: Economics and Computer Science

Brief Abstract:
    This project will explore the possibilities of stock market prediction using a neural network and machine learning. The goal of this project will be to successfully predict market trends based on a variety of social and economic data.

Total Budget: $582.28

Funds requested/received from other sources*: $282.28

Amount requested from Student Project Fund: $300.00

THE MONEY MAKER

I. Project Summary

The goal of this project is to build a program to monitor and predict the future price of equities in the US Stock Market. The beginning stages of this project will use advanced web scraping techniques to gather both economic and social data of various companies currently traded on the U.S. Stock Exchange. This data will be sorted and analysed using an economic and social scoring system to evaluate optimal trading strategies. Eventually this project will implement a neural network and use machine learning to attempt to find correlations between social and financial data. By collecting and analysing both social and financial information pertaining to specific companies, this project will potentially discover new methods for market price prediction.

II. Methods

a. Data Collection

The first step to this project will be to collect data for one specific company. Before expanding data collection to more companies, it is important to solidify the data collecting strategies and techniques. There are two main types of data that will be collected for each individual company. The first type, economic data, will include financial information such as, current stock price, price per earning ratio, working capital ratio, earnings per share, debt-equity ratio, and return on equity. [1] This current and historical data will be collected using Alphavantage API and stored on our servers on a daily basis.

The second type of data to be collected is social data. This data will be scraped daily from a variety of sources such as instagram, twitter, reddit, facebook, and a variety of market related sites. The type of data to be collected will be basic at first, such as, likes, mentions, hashtags. In the beginning, this project will focus purely on the volume of these potential indicators. Later versions of this project will work toward assessing the sentiment of this data.

b. Data Analysis

The second step of this project will be to analyse the data that we have collected. We will do this manually at first by creating an economic and social ranking system. A company with favorable economic data will be assigned a high economic score. A company with less than favorable economic data will be assigned a low economic score. Similarly, companies who receive a lot of social activity online will receive a higher social score versus companies with a smaller amount of social activity. Later versions of this project will adjust social scores based on sentiment. These two scores will be analysed with the hope of finding correlations between the two indicators. Because it may be difficult for a human to recognize important correlations, the next phase of this project will implement artificial intelligence in order to do this work for us.

c. Neural Network implementation

It will be nearly impossible to know exactly which factors will indicate a market move; therefore, a Neural Network will be implemented and trained to help find correlations between the data. Amazon Web Services Deep Learning AMIs will be used to build custom environments and workflows with TensorFlow, PyTorch and a variety of other tools. The social and economic scoring system will be adjusted as needed. Once the correlations have been discovered, the neural network can then be used to estimate specific undervalued and overvalued equities based on the correlated indicators.

d. Execution

Once the project has successfully accomplished the above, additional code will be written to implement the project in a sandbox using quantopian. Here, the program will be given the opportunity to prove its validity without human interference by buying and selling stocks in a fake market environment. This will allow the project to prove its effectiveness before allowing it to make trades with real money. Because the algorithm will have been written based on past data, it will be important to be aware of any over-fitting that may have occurred and adjust the analysis and neural network as needed.

III.   Expected Results

It is expected that using our unique ranking system, the analyses of the data collected will provide insight into future market trends. In order to prove the success of this project, the program will need to successfully predict market prices using both historical data for backtesting and current data in a market sandbox. If the program can consistently predict market prices above 65% of the time using back testing, then the program should be considered successful and used in realtime to predict actual market prices. The expectation is to not only successfully predict market prices in the immediate future, but also for a longer period of time (a week or month out). It is our hope that this program will be used to identify accurate predictions before the general public and take advantage of this by buying and selling stocks accordingly.

IV.   Timeline

Month	Project Goal	Related Objective	Activity	Expected Date	Person
1	Collect/Store Relevant Data	Data Storage	Setup Cloud	Month 1	James Pretti
1		Collect Market Data	Sign up for service to gather and store market data in real time. (Alphavantage)	Month 1	James Pretti
1		Collect Social Data	Web scrape all relevant social sources and store data collected	Month 1 / Ongoing	James Pretti
2	Analyse Stored Data	Decide what data is relevant.	Assign scores to various data representing each stock's ranking within that data pool.	Month 2 / Ongoing	James Pretti
			Look for correlations using a human	Month 2 / Ongoing	James Pretti
		Run Neural Network	Use NN to look for correlations	Month 3	James Pretti
			Use NN to predict future market prices	Month 3	James Pretti
3	Determine Accuracy / Effectiveness	Analyse NN	Compare NN predicted prices with actual data and determine accuracy of predictions	Month 4	James Pretti
			Determine which data points are not influential in market price determination	Month 4 / Ongoing	James Pretti
4	Begin Trial	Use the program to predict market prices	Make a few trades using predicted data from program	Month 5 / Ongoing	James Pretti
		Analyse Program Predictions	Check accuracy and validity of program	Month 5 / Ongoing	James Pretti

V. Budget

Costs

Part	Supplier	Unit Price	Quantity	Total Cost
Cloud Storage / Server				
General Purpose (SSD) Storage	AWS	$0.168 per GB-month	1 terabyte-month	$336.00
Data Processing	AWS	$0.016 per GB	1 terabyte-month	$32.00
SageMaker Processing	AWS	$0.05 per hour	1,440 hours	$72.00
Model Training	AWS	$0.157 per hour	200 hours	$31.40
Model Deployment	AWS	$0.077 per hour	1,440 hours	$110.88
			Total:	*$582.28*

While this project will initially last 4 months, the budget above only accounts for 2 months worth of costs. This is because Amazon offers free AWS for the first 2 months. Additionally, the budget above is an overestimate of costs. Because it is impossible to know the exact amount of data that will be stored, an overestimate of 1 terabyte was made. The amount of data will begin at 0 gigabytes and increase throughout the project, therefore the cost of storage will also initially be less expensive than the above estimate. The general purpose storage provided by AWS will house all of the collected economic and social data. This is the hardest item to estimate as the amount of data needed is currently unknown. The data processing will also be handled by AWS. Again, an estimate for this figure is given because we cannot accurately predict how much data will be processed at this time. The Neural Network will be developed, processed and trained using AWS's Model Training. Finally, the project and neural network will also be deployed using AWS.

VI. Relevant Experience

I have spent the last few years pursuing a degree in Computer Science at UC Santa Cruz. During this time I have focussed on web scraping techniques and designing and training various Neural Networks for small projects requiring prediction.

Before studying at UC Santa Cruz, I worked for 4 years at Capital Planning Advisors in Walnut Creek, CA. One of the many benefits of working at this wealth management firm was the education I received on the economy and U.S. stock market.

I have also spent the last 8 years trading in the stock market using various equity and option strategies.

I would like to combine my past experiences and use my knowledge in computer science to help improve my performance in the stock market.

VII.   References

1) Wilkins, G. (2020, January 29). 6 Basic Financial Ratios and What They Reveal. Retrieved from https://www.investopedia.com/financial-edge/0910/6-basic-financial-ratios-and-what-they-tell-you.aspx

2) Hudgeon, D., & Nichol, R. (2020). Machine learning for business: using Amazon SageMaker and Jupyter. Retrieved from https://aws.amazon.com/sagemaker/pricing/