Integrating Sentiment Analysis and Technical Data to Predict Changes in the Stock Market

Natural Language Processing - Final Project Proposal

Brandon Kynoch - bk23638
Student in the School of Natural Sciences - Computer Science
The University of Texas at Austin
email: kynochb@utexas.edu

***Abstract*—With the total value of the world's stock market capitalization surpassing $93 trillion, investing in the stock market is an essential part of the global economy. Despite its importance, the stock market is a notoriously complex and dynamic system that can be difficult to understand and predict. One of the biggest challenges faced by investors is the task of analyzing stocks and predicting their future performance accurately.**

The Efficient Market Hypothesis argues that stock prices follow a random walk and cannot be predicted solely on past information [1]. However, a recent survey on methodologies, developments and the future direction of stock market prediction using machine learning by Nusrat Et Al. [3] has demonstrated that with the advent of new technologies many researchers have actually been able to predict stock prices to a certain extent.

In this study I propose a novel approach that utilizes machine learning and natural language processing techniques to analyze stock data and news articles to provide predictive insights into the stock market. I hope to build on the findings from previous researchers as discussed by Nusrat Et Al. and focus primarily on effectively performing and integrating sentiment analysis into the stock market time series data for prediction.

This research has significant implications for the financial industry as it offers a practical solution to one of the most critical and ongoing challenges investors face - predicting stock market trends. The proposed approach would provide investors with valuable insights into the stock market, allowing them to make informed decisions and improve their investment outcomes.

I. Introduction and Related Works

A recent survey on methodologies, developments and the future direction of stock market prediction using machine learning by Nusrat Et Al. [3] has demonstrated that with the advent of new technologies many researchers have actually been able to predict stock prices to a certain extent. This survey was conducted in 2021 and focused on the advancements in the stock market industry due to the advent of new technology.

The review highlighted the importance of machine learning and deep learning in the stock market prediction (SMP) process. The SMP process has become more optimal due to the inclusion of big data analytics, machine learning, and deep learning. The literature review analyzed studies based on a generic framework for SMP. These studies used a variety of different input data types, data pre-processing approaches, and machine learning techniques for the predictions. SVM is a pattern classifcation and regression algorithm which has proven to be the most popular method for stock market prediction in the previous works that were analyzed. Some other techniques that were explored include Naive Bayes (NB), k-Nearest Neighbors (kNN), Artificial Neural Networks (ANN), Decision Trees (DT), Principle Component Analysis (PCA), Factor Analysis (FA), Latent Dirichlet Allocation (LDA), Genetic Algorithms (GA) and Firefly Optimization (FO).

The review concluded that support vector machines (SVM) was the most popular technique use for SMP, but artificial neural networks (ANN) and deep neural networks (DNN) were mostly used for accurate and fast predictions. Almost all studies achieved prediction accuracies in the range of 50% to 90%. Typically these studies used MSE loss or F-Measure for optimization, and noted that regularization was important while training. The review also noted that the inclusion of both market data and textual data from online sources improves the accuracy of stock market predictions. Additionally, the review discussed the challenges and open issues in SMP systems, which are yet to be resolved by researchers. One of these challenges is that although data is becoming more available, it is still very difficult to collect and pre-process such data and remove noise from textual datasets.

II. Scope of Project

The price of a publicly traded stock is affected by a range of factors that can vary depending on the company, the industry it operates in, and the broader economic and political climate. Some of the key factors that can influence stock prices include a company's financial performance, industry trends, changes in interest rates, government policies, and global economic conditions. Furthermore, public sentiment can play a critical role in determining the price of a publicly traded stock. The way a company is perceived by the public, its customers, and its employees can influence investor confidence and ultimately impact the stock's price. Positive public sentiment, such as strong brand reputation or favorable media coverage, can drive up a stock's price. Likewise, a negative sentiment such as scandals or public backlash, can cause it to plummet. With the rise of social media and online news platforms, public sentiment can spread rapidly and impact stock prices within minutes. As such, understanding the impact of public sentiment on stock prices is becoming increasingly important for investors and analysts looking to make informed investment decisions.

While most previous studies focus on using either sentiment analysis or stock data to predict price changes, I intend to combine these approaches for a hybrid analysis that should provide a more holistic interpretation of the stock. Furthermore, all of the studies that I have looked at were conducted before the conception of Transformers. Hence, I would like to explore using Transformers for sentiment analysis to predict stock prices. In my approach I will compose a time series where each element shall be a combination of the stock price with a corresponding sentiment score which is computed using articles published over the same time period. I shall prepare datasets for this architecture by using the Trading View API to collect stock market price data and FinViz.com to find articles related to the stock of interest. Datasets will be created using real stock data and news articles from the recent past and this will be partitioned into multiple smaller samples. Hence it should be easy to generate large datasets for training. FinViz provides the date and time of each article published, meaning that after performing sentiment analysis on each article, I can easily map these sentiment scores into the time series data. This time series data shall be used as the input features to a model such as an LSTM network that will output a prediction of the change in stock price. An overview of the dataset preparation process can be seen in figure 'a'.

This problem can be decomposed into two main parts: Firstly, collecting and performing sentiment analysis from online news articles. Secondly, combining these sentiments with the time series data to provide a holistic prediction. For the sake of this final project, I will focus on tackling the first part of this problem effectively. If time permits, I will also attempt to implement the second part of this architecture.

To perform the sentiment analysis from online articles linked at FinViz.com we first have to download the web pages source code and extract the relevant text from this source code per article. Downloading the linked articles source code is trivial, however, fetching only the relevant english text from the markup source code is more challenging. I plan to use the MarkupLM model by Li Et Al. [2], which is publicly available on Hugging Face, to extract the relevant text from the markup source to be used for sentiment analysis. This sentiment analysis will provide us with the sentiment score as seen in figure 'a'. We of course need to model an objective function by finding the change in stock price after each article was published. To do this, I will use the TradingView API to collect stock price data. Collecting stock price data is also quite trivial — In fact, I have already implemented code to fetch the stock price data from Trading View and prepare this data in Python. As for the actual sentiment analysis approach, here we have some freedom. I will explore different model architectures such as N-Grams, LSTM's, Transformers and DNN to see which techniques will yield the best results.

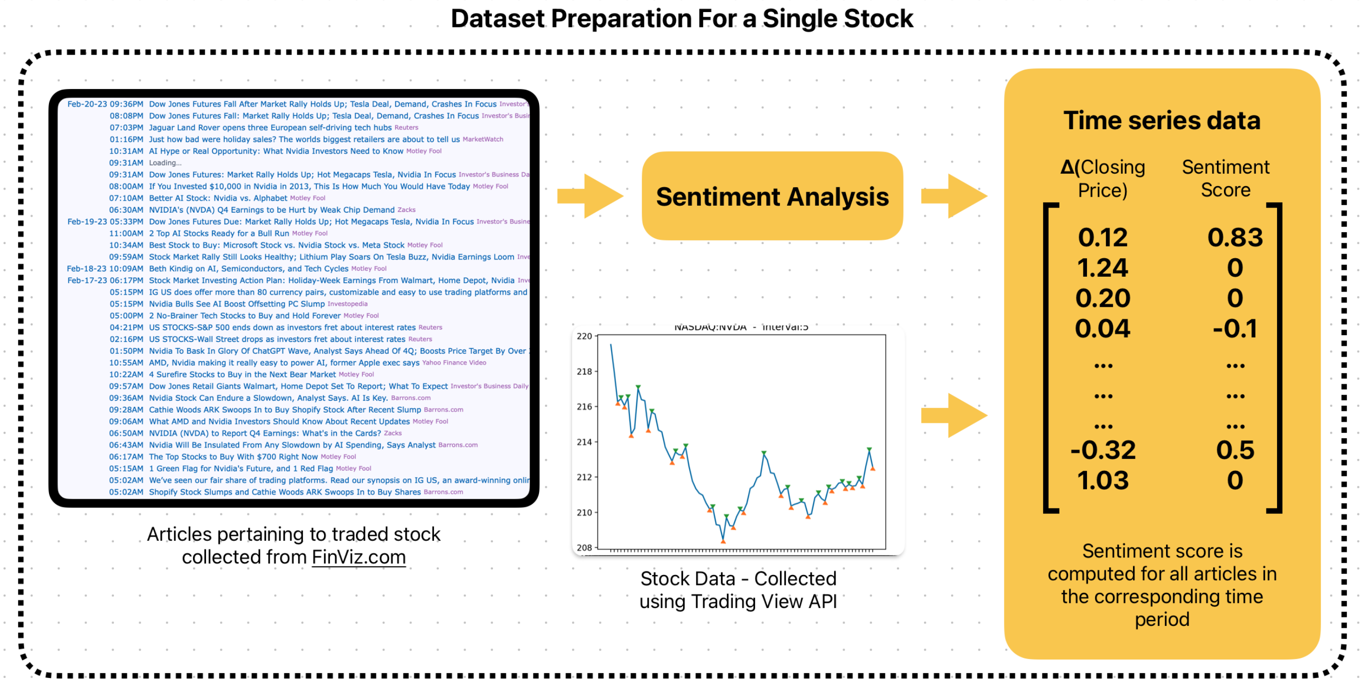
Once we have trained a model to obtain sentiment scores from FinViz.com we can tackle the second part of the problem which is combining this with the stock price time series data and using this combined data to perform even more in-depth and holistic predictions. For this combined approach, I will again be exploring different architectures to determine which will provide the most accurate predictions.

References

[1] Eugene F. Fama. Random walks in stock market prices. *Financial Analysts Journal*, 51(1):75–80, 1995.

[2] Junlong Li, Yiheng Xu, Lei Cui, and Furu Wei. Markuplm: Pre-training of text and markup language for visually-rich document understanding. *CoRR*, abs/2110.08518, 2021.

[3] Nusrat Rouf, Majid Bashir Malik, Tasleem Arif, Sparsh Sharma, Saurabh Singh, Satyabrata Aich, and Hee-Cheol Kim. Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions. *Electronics*, 10(21), 2021.


Dataset Preparation For a Single Stock
Sentiment Analysis
NASDAQ:NVDA - Interval:5
220
218
216
214
212
210
208
Time series data
Articles pertaining to traded stock collected from FinViz.com
Stock Data - Collected using Trading View API
Sentiment score is computed for all articles in the corresponding time period

Feb-20-23 09:36PM	Dow Jones Futures Fall After Market Rally Holds Up; Tesla Deal, Demand, Crashes In Focus	Investor's
08:08PM	Dow Jones Futures Fall: Market Rally Holds Up; Tesla Deal, Demand, Crashes In Focus	Investor's Busi
07:03PM	Jaguar Land Rover opens three European self-driving tech hubs	Reuters
01:16PM	Just how bad were holiday sales? The worlds biggest retailers are about to tell us	MarketWatch
10:31AM	AI Hype or Real Opportunity: What Nvidia Investors Need to Know	Motley Fool
09:31AM	Loading...	
09:31AM	Dow Jones Futures: Market Rally Holds Up; Hot Megacaps Tesla, Nvidia In Focus	Investor's Business Da
08:00AM	If You Invested $10,000 in Nvidia in 2013, This Is How Much You Would Have Today	Motley Fool
07:10AM	Better AI Stock: Nvidia vs. Alphabet	Motley Fool
06:30AM	NVIDIA's (NVDA) Q4 Earnings to be Hurt by Weak Chip Demand	Zacks
Feb-19-23 05:33PM	Dow Jones Futures Due: Market Rally Holds Up; Hot Megacaps Tesla, Nvidia In Focus	Investor's Busine
11:00AM	2 Top AI Stocks Ready for a Bull Run	Motley Fool
10:34AM	Best Stock to Buy: Microsoft Stock vs. Nvidia Stock vs. Meta Stock	Motley Fool
09:59AM	Stock Market Rally Still Looks Healthy; Lithium Play Soars On Tesla Buzz, Nvidia Earnings Loom	Inve
Feb-18-23 10:09AM	Beth Kindig on AI, Semiconductors, and Tech Cycles	Motley Fool
Feb-17-23 06:17PM	Stock Market Investing Action Plan: Holiday-Week Earnings From Walmart, Home Depot, Nvidia	Inve
05:15PM	IG US does offer more than 80 currency pairs, customizable and easy to use trading platforms and	
05:15PM	Nvidia Bulls See AI Boost Offsetting PC Slump	Investopedia
05:00PM	2 No-Brainer Tech Stocks to Buy and Hold Forever	Motley Fool
04:21PM	US STOCKS-S&P 500 ends down as investors fret about interest rates	Reuters
02:16PM	US STOCKS-Wall Street drops as investors fret about interest rates	Reuters
01:50PM	Nvidia To Bask In Glory Of ChatGPT Wave, Analyst Says Ahead Of 4Q; Boosts Price Target By Over	
10:55AM	AMD, Nvidia making it really easy to power AI, former Apple exec says	Yahoo Finance Video
10:22AM	4 Surefire Stocks to Buy in the Next Bear Market	Motley Fool
09:57AM	Dow Jones Retail Giants Walmart, Home Depot Set To Report; What To Expect	Investor's Business Daily
09:36AM	Nvidia Stock Can Endure a Slowdown, Analyst Says. AI Is Key.	Barrons.com
09:28AM	Cathie Woods ARK Swoops In to Buy Shopify Stock After Recent Slump	Barrons.com
09:06AM	What AMD and Nvidia Investors Should Know About Recent Updates	Motley Fool
06:50AM	NVIDIA (NVDA) to Report Q4 Earnings: What's in the Cards?	Zacks
06:43AM	Nvidia Will Be Insulated From Any Slowdown by AI Spending, Says Analyst	Barrons.com
06:17AM	The Top Stocks to Buy With $700 Right Now	Motley Fool
05:15AM	1 Green Flag for Nvidia's Future, and 1 Red Flag	Motley Fool
05:02AM	We've seen our fair share of trading platforms. Read our synopsis on IG US, an award-winning onlin	
05:02AM	Shopify Stock Slumps and Cathie Woods ARK Swoops In to Buy Shares	Barrons.com

Δ(Closing Price)	Sentiment Score
0.12	0.83
1.24	0
0.20	0
0.04	-0.1
…	…
…	…
…	…
-0.32	0.5
1.03	0

(a)