Sentiment Analysis of Stocks from Financial News with tweeter feedback

Khan Shaikhul Hadi

University of Central Florida,Orlando, Florida, USA
shaikhulhadi@knights.ucf.edu

Abstract: Stock value of any company significantly relies on public perception of the company. News headline is one major source of information that influences public perception. Thus sentiment analysis based on headlines is a good way of stock value prediction. With the rise of social media like twitter, public sentiment in social media also influences stock value significantly. In this work, I analyzed difference of public sentiment between headline based sentiment vs tweeter based sentiment and try to formulate weighted sentiment based on both headline and it's relevant tweeter responses. Results shows that, there is significant conflict between headline and how people responses to it in tweeter and there should be significant research on how to combine both to generate more effective sentiment value that we could use to predict stock prices.

1. Introduction

Sentiment analysis has received a great deal of attention from researchers recently and has been used in a variety of fields. The financial industry is particularly interested because it has been demonstrated that news and media have a significant impact on market volatility [1]. Stock market is particularly dominated by public sentiment. Favorable news typically has a positive effect on markets and tends to boost optimism [2]. As a result, textual information processing has developed into an effective instrument for forecasting market dynamics. Predicting stock price using sentiment analysis based on what's on the news is an active research area for financial research field [3], [4]. Due to extreme and increasing influence of social media like tweeter, there is also increase number of research happening to predict stock price from sentiment analysis of tweeter and other social media [5], [6], [7]. But Those are separate research and there is an open question remains, *" What if news based prediction predict price change in one direction and social media based prediction predict opposite direction ?"*. In this work, I will explore this question. In this work, I first implement a sentiment analysis based on financial news headline which could be used as machine learning model input to predict stock price. Then I search related tweet for each headline and calculate overall sentiment on tweeter that may results from the news. Finally, I tried to calculate an effective sentiment value based on headline based sentiment value and relevant tweeter based sentiment value.

2. Overview of Proposed Approach/System

Figure 1 shows our overall algorithm to incorporate tweeter feedback with financial headline to analyze overall sentiment. First, it will collect a fixed number of recent financial news headline and process to increase it's readability. Then perform sentiment analysis using natural language processing tool based on news headline. To include social media feedback i.e. twitter, first we will implement sentiment analysis based on financial news headlines. For each headline, I isolate relevant keywords that will be used for relevant tweeter search. I propose to separate keywords in two different section, (primary keywords and secondary keywords) both consist of nouns. Primary keywords consist of most relevant nouns from the headline like company name, person name etc. On the other hand, secondary keywords will contain rest of the relevant keywords. If headline does not contain enough primary key word, secondary keyword should be included in primary keyword.

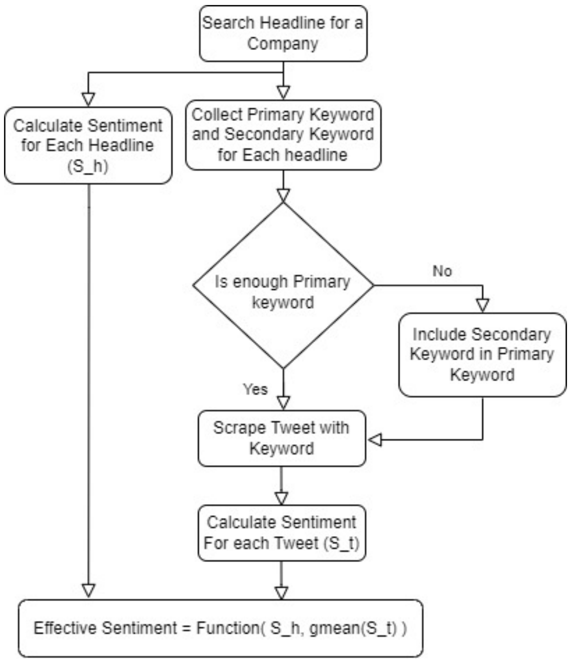


Fig. 1: Algorithm to generate sentiment value from financial headline with tweeter feedback

Based on keywords, I search the tweeter to find relevant tweets and shares. Then collect the tweets and performs sentiment analysis on each tweet. For each headline, I found multiple tweets where each have a sentiment value. So, I take geometric mean of all tweet's sentiment value. I use Function $f()$ to combine sentiment value from headline , S_h with geometric mean sentiment value from tweeter for each headline, S_t. I call it effective sentiment, S_e.

Function $f()$ is user dependent. For my implement, I used a weight parameter λ to add weighted value of both S_t & S_h to get S_e.[1]. Function $f()$ description is given below:

$$f(x,y) = (1-\lambda)x + \lambda y; 0 \leq \lambda \leq 1 \tag{1}$$

3. Technical Details and Experiment

3.1. Technical Challenges

To analyze headline sentiment with tweeter feedback, I encountered following challenges:

- **Negative number :** Sentiment analysis provides four types of information: positive, negative, neutral and compound. Compound consist of both negative and positive values. Thus it was not possible to calculate geometric mean. To handle this issue, first I have to remap value of compound data,c from $-1 \leq c \leq 1$ to $0 \leq c \leq 1$, calculate geometric mean and reveres the mapping process
- **Impact of like & retweet :** Impact of tweet is not limited to just it's content but how other people react to it via retweet and likes. As a result, considering all tweet with same significance may undermine the influence of some tweets over tweets that does not reach many people. For this work, I skipped this issues [2] and consider all tweets with same importance.

3.2. Technical Detail

First technical decision I have to make is how to collect financial headline in efficient way. I choose to use FINVIZ [8] website . This website has a dedicated section for headlines related to each stock. Thus sending URL request with a company name will return an html file that contains most relevant headlines for that particular company.

I choose to consider nouns as keyword from headline. Each headline consist of many words, dominated by nouns, adjective , verb and adverb. For tweeter search, which words to consider was a judgement call. I observe that, just for adjective and adverb keyword without noun or verb , tweeter search result became polluted for flexible search (search can contain any one of the keyword; "key-word-1" or "key-word-2") or do not provide any result for strict search (search must contain all of the keyword, "key-word-1" and "key-word-2") . First condition arises as people use

[1] My effective sentiment value calculation is not based on any research, just followed the idea I've seen in the class for different machine learning technique. There is not enough reserach on this issue

[2] I could not find any easy way to incorporate this within the time limit of the project. Also, when tweets are posted may also have some kind of influence that should be represented in sentiment value for effective stock price prediction.

adjective and adverb all the time. For the second condition, for a tweet, people generally do not follow grammar or express emotion with their own word. For verb, it is similar. Thus we choose to only select noun as keyword as it is highly unlikely people used them in tweeter randomly.

Among noun, I observe that proper noun provide most relevant search result in tweeter. For this reason, proper noun is used as primary keyword. On the other hand, other nouns is not as efficient but still provide some relevant search result. But many cases headline does not contain enough primary keyword to provide relevant search result. For this reason, with rest of the noun, secondary keyword is created. When I encountered lower number of primary keywords, secondary keywords will be concatenated in primary keywords.

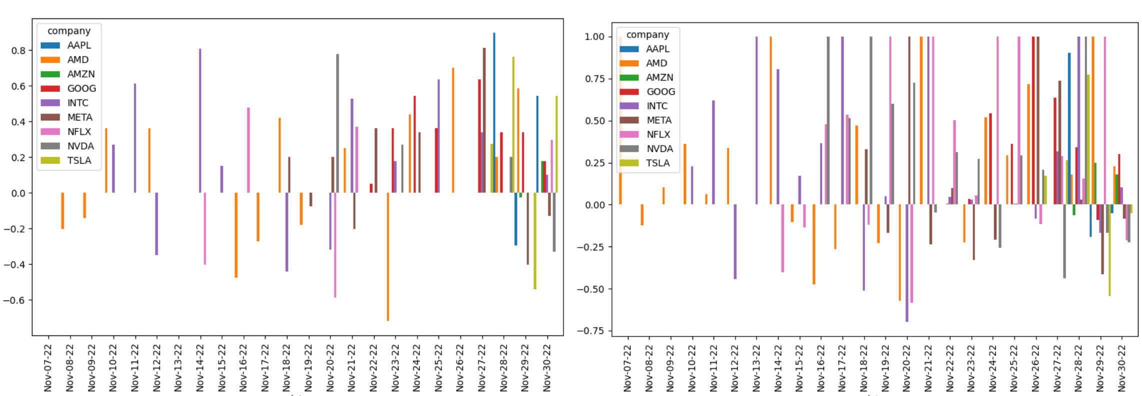
3.3. Experiment

Whole program is written in python. From finviz [8] website, I collected relevant headlines for a set of company using *BeautifulSoup* python library. List of the companies are: Tesla Inc.(TSLA), Amazon.com Inc. (AMZN), Alphabet Inc. (GOOG), Apple Inc. (AAPL), Meta Platforms Inc. (META), Netflix Inc. (NFLX), Advanced Micro Devices Inc.(AMD), Intel Corporation (INTC) & NVIDIA Corporation (NVDA). Upon collecting the headlines, a *pandas* data frame is create for analyzing. Using python library *re*, unnecessary number and symbols were removed and using *word_tokenize & pos_tag* from *nltk* library, keyword are selected. To measure sentiment from a sentence, *nltk* library is used. *nltk* library could analyze a sentence to generate sentiment value that consist of positive sentiment, negative sentiment, neutral sentiment and compound sentiment that consider impact of previous three factor. To plot result, compound sentiment is used.

To search and collect tweet based on keyword, *sntwitter* library from *snscrape* python module is used. To calculate sentiment of the tweet, above mentioned method is used. To visualize the result, I used *pyplot* [3]

3.4. Disclosure of Resources

To calculate sentiment from only headlines, I mainly used article *'Sentiment Analysis of Stocks from Financial News using Python'* [9] as reference. There are also other online blogs and resources that I have used as reference to complete this process [10], [11]. To collect tweet based on search result, blog *'Web Scraping with Python – How to Scrape Data from Twitter using Tweepy and Snscrape'*[12] was my primary resource. There are other blogs also provided significant insights [13], [14], [15].

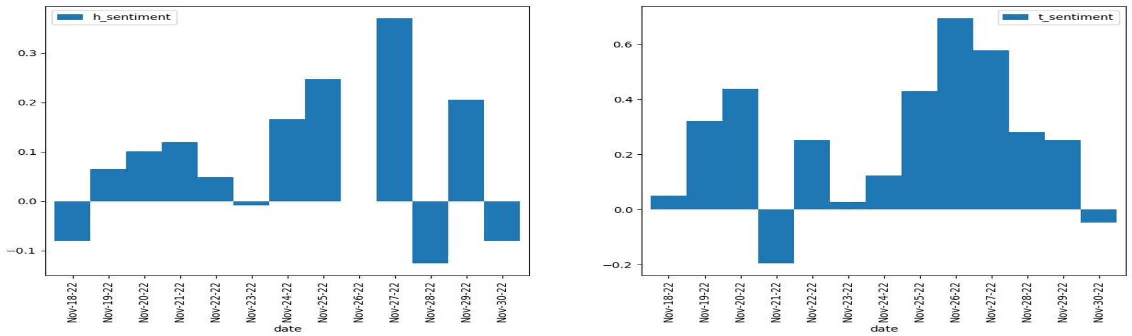


(a) News Headline Sentiment for multiple company   (b) Tweeter Sentiment for multiple company

Fig. 2: Sentiment Analysis of multiple tech company based on financial news headline and tweeter search

[3]Source Code : https://drive.google.com/drive/folders/19J80qm39wHAzKqNN3wzzvj-bAzEBT0Bs?usp=share_link

4. Evaluation



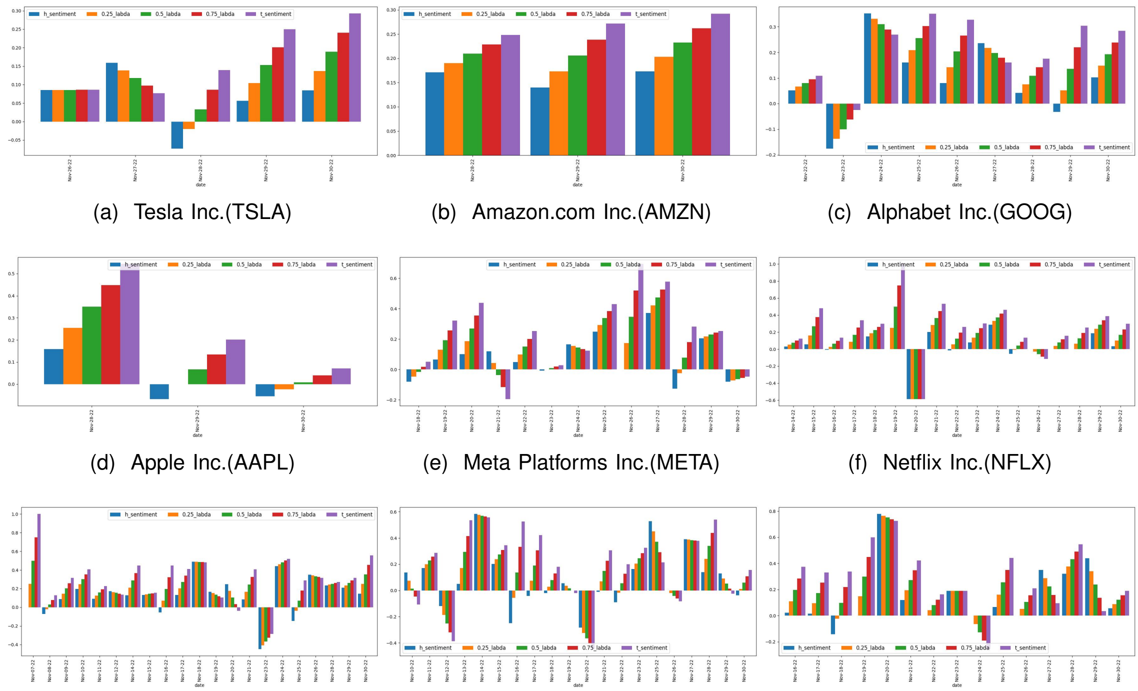
(a) News Headline Sentiment for Meta (b) Tweeter Sentiment for Meta

Fig. 3: Sentiment Analysis of Meta Platforms, Inc. (META) based on financial news headline and tweets

Figure 2 shows final compound value of sentiment analysis for financial news headlines (Fig 2a) and based on tweets (Fig 2b). If we compare this two result, it is obvious that overall sentiment based on just headline is different than overall sentiment based on relevant tweets. I observe frequency of maximum value (i.e. 1) for tweeter based sentiment which indicates that people either way more sensitive to particular type of news that assumed or people express strong emotion when they use social media(i.e. tweeter). Also, Value density in Figure 2b is higher than Figure 2a. It happens because there exist news headlines which have zero or close to zero sentiment compound value (nearly impact less news). But relevant tweet shows that , people may react to those headline in such a way that those tweets have significant sentiment value. I also notice, for some headlines (AMD on Nov-23-22) sentiment value could be very high but people may not reacted similarly.

To understand better, Figure 3 represents sentiment value for Meta Platforms, Inc. (META). For headline and tweet response, both shows overall same pattern for sentiment value. But when we look into values for independent date, my may encounter discrepancy. On Nov-18-22 and Nov-28-22, only headline based sentiment analysis will show that overall sentiment for META is negative. But tweets related to news headlines on Nov-18-22 and Nov-28-22 have overall positive sentiment. Similarly on Nov-21-22 there were three news headline regarding META which have small but positive sentiment value and it would be logical to expect positive response in tweeter. But, response in tweeter overwhelmingly negative. The headlines are ”Meta Workers Hijacked User Accounts and Charged Bribes, Report Says”,”What does free speech actually mean? Twitter was never censoring speech, despite what Elon Musk and some users have said” and ”The Safest and Riskiest FAANG Stocks to Buy Right Now”. This indicates that either general public interpreted the headlines totally different way or previous history generally have a inductive effect on people that one bad day in news-headline do not impact people perception significantly (i.e. from third headline people could assume META is riskiest stocks to buy based on precious history). As public perception on a particular company have significant impact on it’s stock value, which sentiment value between this two scenario will predict stock price change in future more efficiently remains a open question.

To incorporate both sentiment value as a single parameter, I introduced compounding function $f()$ which will give us an effective sentiment value based on headline based sentiment value and relevant tweet based sentiment value. Optimized and effective property of compounding function is not explored in this work. I used simple weight based approach where both sentiment value is multiplied by a weight factor λ and then return total sum of them. Figure 4 shows change in effective sentiment value of nine top company in USA with different value of λ. Slop between



(a) Tesla Inc.(TSLA) (b) Amazon.com Inc.(AMZN) (c) Alphabet Inc.(GOOG)

(d) Apple Inc.(AAPL) (e) Meta Platforms Inc.(META) (f) Netflix Inc.(NFLX)

(g) Advanced Micro Devices Inc. (AMD) (h) Intel Corporation (INTC) (i) NVIDIA Corporation (NVDA)

Fig. 4: Change of effective sentiment value with λ for different tech. company

different λ value is proportional to discrepancy between sentiment value from only headline and sentiment value from relevant tweets. For example, INTC on Nov-16-22 have totally opposite result for S_h (negative) and S_t (positive) . On the other hand, INTC on Nov-27-22 and Nov-14-22, both values are same. As a result, we could interpreted that sentiment value from on this day's headline is not reliable and we should give less important on this headlines compared to headlines on Nov-14-22 and Nov-27-22 which should get more important when one will try to predict stock price using sentiment value. Right now, we have no knowledge of any theoretical process to incorporate this findings in machine learning model.

5. Technical Limitations

During this works, I encountered multiple limitations that seriously impacted it's result analysis and validation checks. Limitations are given below:

1) **Misleading Headlines :** For news headline based sentiment analysis assumes that headline focuses on the major event or fact that relevant article focuses on thus it is related to the important fact that have impact on human sentiment .Sometimes news have misleading headline that focuses on secondary content of the article rather than conveying main focus of the article [16]. As a result, people could have different interpretation of the news than the headline may suggest. As a result, sentiment analysis may produces false indication. In this work, we ignored this phenomena in headlines.
2) **Missing Relevant Data-set :** There are independent studies performed to analyze public sentiment and social medial (i.e. tweeter) sentiment alongside stock price prediction using those sentiment values. But there are no data set that I know of , that have headline, relevant tweet (or other social media entry) with target stock price change. Thus in this work, I could not test any prediction model to analyze how different value of λ impact prediction

accuracy. To test effectiveness of my effective_sentiment value, I would need training data-set consist for news headline and relevant tweets along with information if stock price increase or decrease in the future.

3) **Inefficient Search on Social Media :** Social media search is not very intuitive and consist of irrelevant sections which was problematic. For example, some keyword search in tweeter also shows relevant advertisement, people account relevant to this. Besides that, there is no method available to connect tweet and the link the shared.

6. Conclusion

In this work, I analyzed overall sentiment value from financial news headlines, tweets sentiment related to those headlines and compare them to gain insight. My analysis shows that, regular sentiment analysis just from news headline may not show proper sentiment the people express on the social media. I argue that, by comparing this two type of sentiment value, we could weight sentiment value for different news headline based on reliability of those headline sentiment value before using them in any machine learning model to predict stock price. I argue, if sentiment value of news headline find similar sentiment on social media related to that headline, should have higher importance to predict stock price compared to news headline that shows contradictory sentiment on social media with it's own sentiment should have lower importance. There is lack of mathematical expression to interpreter sentiment analysis value that incorporate sentiment value from social medial related to that headline.

References

[1] Rohitha Goonatilake and Susantha Herath. The volatility of the stock market and news. *International Research Journal of Finance and Economics*, 3(11):53–65, 2007.

[2] Marjan Van de Kauter, Diane Breesch, and Véronique Hoste. Fine-grained analysis of explicit and implicit sentiment in financial news articles. *Expert Systems with Applications*, 42(11):4999–5010, 2015.

[3] Saloni Mohan, Sahitya Mullapudi, Sudheer Sammeta, Parag Vijayvergia, and David C Anastasiu. Stock price prediction using news sentiment analysis. In *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*, pages 205–208. IEEE, 2019.

[4] Ayman E Khedr, Nagwa Yaseen, et al. Predicting stock market behavior using data mining technique and news sentiment analysis. *International Journal of Intelligent Systems and Applications*, 9(7):22, 2017.

[5] Venkata Sasank Pagolu, Kamal Nayan Reddy, Ganapati Panda, and Babita Majhi. Sentiment analysis of twitter data for predicting stock market movements. In *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, pages 1345–1350, 2016.

[6] Thien Hai Nguyen, Kiyoaki Shirai, and Julien Velcin. Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, 42(24):9603–9611, 2015.

[7] Thien Hai Nguyen and Kiyoaki Shirai. Topic modeling based sentiment analysis on social media for stock market prediction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1354–1364, 2015.

[8] FINVIZ financial vizualization. https://finviz.com. [Accessed 01-Dec-2022].

[9] Damian Boh. Sentiment Analysis of Stocks from Financial News using Python — medium.datadriveninvestor.com. https://medium.datadriveninvestor.com/sentiment-analysis-of-stocks-from-financial-news-using-python-82ebdcefb638. [Accessed 01-Dec-2022].

[10] Shashank Vemuri. Stock News Sentiment Analysis with Python! — towardsdatascience.com. https://towardsdatascience.com/stock-news-sentiment-analysis-with-python-193d4b4378d4. [Accessed 01-Dec-2022].

[11] [Part 3] Use Your Computer to Make Informed Decisions in Stock Trading — pythoninvest.com. https://pythoninvest.com/long-read/sentiment-analysis-of-financial-news. [Accessed 01-Dec-2022].

[12] Mehul Mohan. Web Scraping with Python – How to Scrape Data from Twitter using Tweepy and Snscrape — freecodecamp.org. https://www.freecodecamp.org/news/python-web-scraping-tutorial/, SEPTEMBER 25, 2020. [Accessed 30-Nov-2022].

[13] Christophe Garon. How to Scrape Tweets Containing Keywords with Python - Christophe Garon — christophe-garon.com. https://christophegaron.com/scrape-tweets-containing-keywords/. [Accessed 01-Dec-2022].

[14] Rashi Desai. How to scrape millions of tweets using snscrape — medium.com. https://medium.com/dataseries/how-to-scrape-millions-of-tweets-using-snscrape-195ee3594721. [Accessed 01-Dec-2022].

[15] Martin Beck. How to Scrape Tweets With snscrape — betterprogramming.pub. https://betterprogramming.pub/how-to-scrape-tweets-with-snscrape-90124ed006af. [Accessed 01-Dec-2022].

[16] Ullrich KH Ecker, Stephan Lewandowsky, Ee Pin Chang, and Rekha Pillai. The effects of subtle misinformation in news headlines. *Journal of experimental psychology: applied*, 20(4):323, 2014.