Quantifying the Emotions Driving the Stock Market

Luis Guillermo Velasco Gonzalez

School of Human Sciences and Technology

Bachelor Information Systems Management

IE University

Supervisor Noa Cruz Díaz, PhD

May, 2021

Declaration

I hereby declare that I am the sole author of the thesis titled "Quantifying the Emotions Driving the Stock Market" and that it has not been previously submitted to any other degree or university.

This thesis is result of my own original work and all derived information has been properly referenced.

Luis Guillermo Velasco

Abstract

This paper explores how the stock market has evolved over recent years and its implication in investment. Accelerated by COVID-19 investing appears to have reached the masses and thus a stock's public perception has to be taken into account when analyzing a security. In this study, public perception was measured from posts and tweets scraped from various social media platforms and used to train a test a Natural Language Processing model. For this the state-of-the-art BERT model was fine-tuned to differentiate between bullish and bearish sentiment, overcoming limitations shown in other research papers on this topic. Moreover, this language model was used to calculate the daily sentiment for Tesla and the S&P 500 and see if there is any correlation between it and the price of the underlying security. At last, this "public perception" was added to a Long-Shot-Term Memory neural network to see if the stock or index price prediction improved; ultimately proving the impact of the retail investor in the stock market and the value of Natural Language Processing in this field.

Table of Contents

1 Introduction **1**

2 Current Market Situation **2**
2.1 The Rise of the Retail Investors 2
2.2 Gamestop Case Study 3
2.3 The Effectos of COVID-19 on Investing 3
2.4 Emotions Against the Market 4

3 Language Model Creation **5**
3.1 Defining BERT 5
3.2 Data Retrieval 6
3.3 Data Labeling 7
3.4 Data Cleaning 8
3.5 Data Separation 9
3.6 Fine-Tuning 9

4 Results **11**
4.1 Results 11
4.2 Comparison Against other Models 12
4.3 Daily Sentiment & Stock Return 13

5 Next Steps **16**
5.1 Future Applications & Improvements 16

6 Conclusion **17**

Chapter 1

Introduction

In 2020 the world experienced one of the worst crashes in stock market history caused by the COVID-19 pandemic. Yet, more shocking than the slashed value of stocks was the V-shaped recovery that took place the following months. Even though COVID cases increased exponentially and countries entered a full lockdown, companies like Tesla saw a 695% increase in their stock price, while indexes such as the S&P 500 reached record-breaking all-time highs. It seemed that the market was disconnected from reality and that traditional investment strategies no longer worked in this post-COVID society. A new force has played a major role in the market and this paper attempts to quantify it.

In this paper, we first explore how the financial markets are rapidly evolving, from the rise of the retail investor and the effects of COVID on the market to the role of online communities, such as wallstreetbets, in today's financial ecosystem. By exploring Gamestop's fascinating 2021 bull run, we find ourselves looking for an explanation beyond the balance sheet. Alluring that one of the major forces driving today's market could be the stock's public perception and online sentiment, surpassing fundamental analyses. Thus by understanding the relationship between a stock's perception and its movement, one could theoretically predict its trajectory.

Forecasting a stock's price based on its public perception scraped from online platforms is not a new concept, yet, as we will see, the current models developed do not take into account major factors such as financial terminology and the medium in which the posts are published. To tackle this, the first public annotated corpus in this field was created as well as implementing a state-of-the-art language model known as BERT. The entire process is documented, from data labeling to fine-tuning, since one of the purposes of this paper is to serve as a guideline for optimizing Natural

Language Processing classification to the financial field. Additionally, a time series Long-Short Term Memory forecasting model was developed to test if there is any improvement by considering online sentiment in addition to classical variables.

Chapter 2

Current Market Situation

2.1 The Rise of the Retail Investors

Over the recent years, there has been an increase in retail investors, which according to Adam Hayes, are defined as a "(...) non-professional investor who buys and sells securities or funds that contain a basket of securities such as mutual funds and exchange-traded funds (ETFs)". Although there are always exceptions, retail investors tend to hold less diversified portfolios and concentrate on "trending" stocks (Hayes, A., 2021). This can cause retail investors to downplay fundamental metrics, such as but not limited to, balance sheets, price-to-earnings ratios, and free cash flow when investing, focusing more on the "public perception" of the asset.

Propelled by the rise of social media platforms, these investors have found channels and communities where they can share their position and strategies. Channels such as Twitter, Reddit, and Youtube offer an opportunity for them to reach a massive audience and niche communities started to form, such as the subreddit wallstreetbets. This clique of high-risk investors is known for sharing their massive losses as symbols of pride and their derogatory language (Boylston, C., 2021). Although before they may not be considered a financial force in the stock market by many, on January 21st, 2021, they were responsible for one of the biggest short squeezes in history.

2.2 Gamestop Case Study

Gamestop or GME is an electronic's retailer that mainly focuses on video games, it is the largest video game retailer in the world, but due to the increase of online retailers, it has suffered a decrease in popularity. Due to the lack of trust of investors that GME would remain competitive over time, combined with COVID-19, which halted on-site sales, it was the most shorted stock in the S&P 500. As of January 22nd, GME shares were shorted by over 140%, preparing the ground field for a short squeeze (Chohan, U. W., 2021).

Although mainly formed by retail investors, r/wallstreetbets has over 10 million members up to date. Because of their high-risk mentality, the community saw an opportunity in going against affluent hedge funds and supporting the beaten-down company. By advertising their GME positions and pushing the increase of GME stock purchases, the subreddit was able to increase the price of the stock and force the short sellers to cover their positions, additionally increasing the value of the underlying stock (Lyócsa, Š., 2021). This short-squeeze pushed the price of the stock from 17 USD at the beginning of January to 483 USD by the 28th of January

In a David vs. Goliath parallelism, a group of retail investors went against big financial institutions and came on top. Hedge fund Melvin Capital, which held the largest shorting position at the time, lost 30% of its value vehicle Citro closed all short positions at a 100% loss (Chohan, U. W., 2021). Gamestop's case exemplifies the underestimated effects that retail investors can have in the market, opening possibilities for a new type of analysis.

2.3 The Effectos of COVID-19 on Investing

Not only did COVID bring one of the most volatile years in recent history, but it fundamentally changed the stock market. During confinement gambling considerably increased, as people searched for ways to escape the mundane (Håkansson, A., 2020) and accelerated by the fiscal stimulus, in the case of the US. Although some conformed to traditional online gambling, many found the stock market as their casino floor.

Platforms such as Robinhood offer a zero-cost trading service where users can trade stocks without paying a transaction fee. The gamification of these platforms served to many as an introduction to investment, and its user base has been constantly increasing, especially among the “millennial” demographic. In March 2020, Robinhood saw customer trading volume triple from 2019, even though the market was suffering one of the most rapid recessions in history (Rooney, 2020). COVID just accelerated the democratization of security investing. Investing has gone mainstream, and thus the variables by which these investors make decisions have to be taken into account. As Black (1986) described, a significant portion of investors can be considered as “noise” traders. Individuals whole base their investment solely on recent news and trends, yet they have the force to destabilize financial markets as they chase the latest news; as was seen with GME.

2.4 Emotions Against the Market

Due to the increase in retail investors and the impact that communities such as r/wallstreetbets can have on the financial market, it is no surprise that investment firms are starting to track these communities. Many papers have explored the topic of predicting stock price based on its public perception or sentiment, yet they suffer from some limitations, specifically in the Natural Language Processing (NLP) area.

In their paper, Stock Prediction Using Twitter Sentiment Analysis, Mittal and Goel (2012) explore predicting the Dow Jones Industrial Average (DJIA) based on posts scraped from Twitter. Although they propose a very compelling machine learning architecture for predicting the price of DIJA, their sentiment prediction was mostly based on a word count model. Moreover, Makrehchi, Shah, and Liao (2013) use a similar approach in their paper Stock Prediction Using Event-based Sentiment Analysis.

As we will see this type of NLP model has been proven useful in many situations, yet when it comes to predicting financial sentiment it has its limitations. First, it may not consider the connotation of financial terms as most important the medium from which these posts are scraped from. The proposed tries to take into account these factors through the use of a granular data labeling process and the implementation of a contextual language model called BERT.

Chapter 3

Language Model Creation

3.1 Defining BERT

BERT or Bidirectional Encoder Representations from Transformers is a language model developed and published by Google engineers in 2018. Since its release, it has not only become one of the standard NLP models but even outperforming other models by large margins, as we can see from the figure below (Devlin, J., 2018).

System	MNLI-(m/mm)	QQP	QNLI	SST-2	CoLA	STS-B	MRPC	RTE	**Average**
	392k	363k	108k	67k	8.5k	5.7k	3.5k	2.5k	-
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.9	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	88.1	91.3	45.4	80.0	82.3	56.0	75.2
$\text{BERT}_{\text{BASE}}$	84.6/83.4	71.2	90.1	93.5	52.1	85.8	88.9	66.4	79.6
$\text{BERT}_{\text{LARGE}}$	**86.7/85.9**	**72.1**	**91.1**	**94.9**	**60.5**	**86.5**	**89.3**	**70.1**	**81.9**

Figure 3.1: Comparison between BERT and other language models

BERT was pre-trained on the BooksCorpus, which contains 800 million words, and on Wikipedia, 2,500 million words (Annamoradnejad, I., 2020). Nevertheless, what is impressive about BERT is not the data on it which it was pre-trained, but its bi-directional, transformer-based machine learning architecture. By being a bi-directional model BERT is able to process the entire sequence of words at once, in contrast to traditional directional models where text is processed sequentially (right to left, left to right). This allows the model to understand the context of the word and how it relates to words either left or right of it (Horev, 2018), significantly improving its accuracy.

In order to achieve this "contextual" understanding of the sentence, BERT uses a masked language model (MLM) where 15% of the words of the input sequence are covered with a [MASK] token (Vaswani, A., 2017). This forces the model to predict the masked words based on the un-masked words surrounding them. Moreover, the model goes through a Next Sentence Prediction (NSP) stage, where the model receives two sentences and predicts whether the second sentence is a subsequent sentence of the first one. Both, MLM and NSP, are trained together to minimize the model's loss and give the model flexibility to perform other NLP tasks such as:

1. Classification
2. Question Answering
3. Name Entity Recognition

For this paper, we will focus on a classification task, which is achieved by adding a layer to the core model through fine-tuning, allowing us to predict if a text sequence is bullish or bearish.

3.2 Data Retrieval

A model is only as good as its data. To train the model over five thousand tweets and two thousand Reddit posts from r/wallstreetbets were retrieved, that mentioned either Tesla or S&P500. To do so, two python libraries were required, *twint* [1] which connected with Twitter's API, and *psaw* [2] which connected to Reddit's API.

Moreover, to target comments that mentioned these entities in a "financial" context the tweets/posts were retrieved using stock tickers. For example, in the New York Stock Exchange (NYSE) the company Tesla is commonly referred to as $TSLA,

[1] github link: https://github.com/twintproject/twint
[2] psaw documentation: https://pypi.org/project/psaw/

while the S&P500 is referred to as $SPY. By focusing our data retrieval on these tickers the data scraped is aligned to the scope of the project.

The process of data retrieval went through various iterations since a closer look at the first batch revealed that there exists a considerable amount of spam in these social media platforms. Spam, referring to posts that only contain hashtags and/or come from automated-bot accounts, carrying no intrinsic value. Thus, additional filters were set in place; at least two likes per post/tweet and a minimum length of 5 characters. These simple filters proved extremely successful at filtering out spam.

3.3 Data Labeling

The retrieved data were manually tagged to assure the quality of the training data used in the model, as well as consistency. Negative or bearish entries were tagged -1, while positive ones, bullish, were tagged 1. When training the model the label for bearish entries had to be changed from -1 to 0 as the previous label was causing problems during the training phase.

Additionally, a certain understanding of the fundamentals of the stock market is required, since there are various terms that are specialized to the field. For example, the scraped tweet: "Just made some quick $ on $TSLA puts. Good start to the day!" has a positive connotation, and popular sentiment analysis models, trained on classical positive and negative tweets, would classify it as positive. But, inside our context the word "put" refers to a market derivative that offers the holder the option to sell an underlying asset (commonly a stock) at a specified price before a date, thus the purchase of a put option connotes a negative sentiment about the future value of the stock (Burghardt, 2008). So it would represent a bearish sentiment and consequently be tagged as negative.

Having a basic understanding of the field is vital when tagging data, yet this is not only true for the financial aspect, but also for the social media one. In particular when dealing with niche communities, such as the case with r/wallstreetbets, subreddit that focuses on high-risk investments that borderlines gambling, and continual teasing. Like in every other community, their own lingo starts to develop. A quick walkaround r/wallstreetbets and one will encounter terms such as, “diamond hands”, “paper hands”, “to the moon”. These terms cannot be found in any academic papers, yet they reference financial positions that can reach hundreds of thousands of dollars.

Due to time constraints and lack of resources, only 1,100 posts/tweets were labeled, 550 bullish and 550 bearish. As with any machine learning model, a greater set of data would improve the accuracy. Nevertheless, as proven in the following sections, high accuracy was achieved because of a great data quality, which inevitably comes from the manually tagged data. For future development services such as Amazon’s Mechanical Turk could be used. Where crowd-sourcing is leveraged to access a population of thousands of online workers to complete various tasks, such as data labeling (Ross, J., 2010).

3.4 Data Cleaning

In order to highlight attributes and omit undesired ones, the data must go through a cleaning process. When preprocessing data for traditional NLP, such as Naive Bayes, it is common for stop-words to be removed. These stop-words are commonly occurring words that are perceived as adding little to no value to the text sequence, words such as: *the, its, an, for, not, nor,* etc. Nevertheless, since BERT is a contextual model it benefits from the use of stop-words since these words can provide the context of the user’s intent (Qiao, Y., 2019).

For this model, the only pre-processing the tweets/posts went through was the removal of hyperlinks and mentions. For the moment, we are unable to access the content of hyperlinks, thus it doesn’t bring any semantical context to the text sequence. The same is true for mentions, we don’t have any context on what the mentioned user has on a certain topic. Both hyperlinks and mentions were removed using regular expressions (regex).

3.5 Data Separation

The labeled data is separated into three different datasets: training, validation, and testing. Each serving a different purpose. As described by Brian D. Ripley in his book: Pattern Recognition and Neural Networks:

- **Training Dataset:** This is the data used to train the model and fit the parameters.
- **Validation Dataset:** The data provides an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters.
- **Testing Dataset:** The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset.

To prevent future model biases, all data within the datasets is balanced. This means that each dataset has an equal amount of examples, in our case bearish and bullish posts/tweets. The training dataset contains 800 entries (*400 bullish & 400 bearish*), the validation dataset contains 100 entries (*50 bullish & 50 bearish*), and at last, the test dataset contains 100 entries (*50 bullish & 50 bearish*).

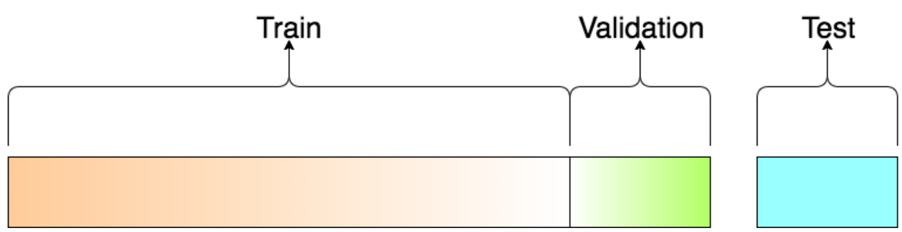

Figure 3.2: Illustration of training, validation, and test datasets

3.6 Fine-Tuning

Two BERT models are available from the open-source provider Hugging Face, BERT-base, and BERT-large. The main difference between these two models is the number of encoders each one contains, 12 for the base model and 24 for the large one. Due to computational limitations, only the BERT-base model was used, which contains 12 layers, 768 hidden ones, and over 109 million parameters (Tenney, I., 2019).

Additionally, all BERT models contain a drop-out layer, which removes random neurons during training to prevent overfitting. It is also important to note, that within BERT-base two sub-models exist, cased and uncased, which refers to if the words are lower-cased or not before tokenizing. For this project, the model 'bert-base-uncased' was used.

The model was run on the training and validation data intending to achieve the highest accuracy on the validation data. As an optimizer Adam was used, which is a gradient-based optimization algorithm that has gained popularity in deep learning (Zhang, Z., 2018). As a learning rate, 3e-5 was used, which represents the "steps" a model takes to achieve a local minimum determined by the loss function.

As shown in the figure below, the model was run through 7 epochs, an epoch is a complete pass through the training dataset (Brownlee, J., 2018). After each epoch, there is an increase in accuracy, which refers to the accuracy of the training dataset, yet, as seen in Table 3.1, after epoch #4 the accuracy of the training dataset continues to increase while the accuracy of the validation dataset starts to decrease, as the model starts to over-fit the data[3].

Table 3.1: Metrics and results for each training epoch.

Epoch	**Loss**	**Accuracy**	**Val. Loss**	**Val. Accuracy**
1	0.6722	0.5831	0.3502	0.7250
2	0.337	0.8750	0.3203	0.7800
3	0.0621	0.9841	0.3895	0.8900
4	0.0193	0.9945	0.4050	0.9150
5	0.0074	0.9968	0.5900	0.8800
6	0.0021	0.9994	0.7783	0.8550
7	0.0061	0.9990	0.5687	0.8850

Overfitting happens when the model starts to "memorize" the training dataset, finding patterns that do not actually apply or translate to new data, thus diminishing the performance of the overall model. Even though as mentioned before, BERT

[3]Code used to fine-tune BERT model can be found in: https://github.com/lgvelasco/tfg_2021

contains a dropout layer, due to the size of our training dataset extra precautions have to be taken.

Chapter 4

Results

4.1 Results

Although the accuracy of the model could be evaluated on the validation data since the model's hyper-parameters were tuned to maximize the validation accuracy, a certain level of bias exists, thus the test data is used for the final evaluation. When running the model on our test data, which has never been seen by the model before we get the following results:

Accuracy: 0.895

F1 Score: 0.893

Confusion Matrix:

Table 4.1: Confusion matrix for predict vs actual labels

	Positive 1	**Negative 0**
Positive (1)	91	1
Negative (0)	12	88

Accuracy measures correct prediction over total predictions, while the F1 score tries to balance precision and recall. Since the number of false positives and false negatives are relatively close, accuracy and the F1 score are quite similar. Reaching almost 90% accuracy with comparatively small training data proves the efficiency of the BERT model, and its capacity to understand the overall context.

Additionally, based on the confusion matrix, there appear to be slightly more false negatives than positives. Incorrectly classified inputs appear to fall into two categories, contradiction, and unclear subject. The following incorrectly labeled input falls under the first category, "I regret buying a Tesla. Said no Tesla owner ever.

The only regrets I and other owners have is not buying the higher spec version...". At first, the author appears to point out that he/she regrets buying a Tesla, but later clarifies the contrary. This "ironic" language is a well-documented gap in the NLP field (Weitzel, L., 2019) and any model will be affected by it.

An example for the second category, unclear subject, can be observed from the following incorrectly labeled tweet: "Warren Buffett says he can't beat the S&P 500". There are two main subjects on the previous quotes, Warren Buffet and the S&P 500, depending on which subject you focus the post can be read as either bearish or bullish. Bearish for Warren Buffet, since he cannot keep up with the index; and bullish for the S&P 500 since it always surpasses one of the most successful investors.

For future iterations of the model, additional training data and a subject identifier would be required. Additional training data would provide examples of "ironic" language and remove errors related to that area, while a subject identifier would help the model focus on the subject, or stock, that we are interested in. No model can reach 100% accuracy without overfitting, yet there are steps to reduce its error to an efficient minimum.

4.2 Comparison Against other Models

When comparing results of the developed model to the ones from different models, given the same testing data, there is a remarkable difference between their performance. For a deeper examination, the *Flair*[1] language model was chosen as a competitor. The open-source model was created by Zalando Institution and trained on labeled positive and negative text entries scraped from Twitter and IMDB, similar to models used by previous researches. Even though their training dataset contained over 50,000 tagged entries, 50 times more than the data used to train the BERT model, it still underperformed when labeling

Accuracy: 0.735

F1 Score: 0.67

[1] github link: https://github.com/flairNLP/flair

Confusion Matrix:

Table 4.2: Confusion matrix for predict vs actual labels

	Positive 1	**Negative 0**
Positive (1)	92	8
Negative (0)	45	55

With an accuracy of 70% and an F1 score slightly lower, the Flair model appears to have issues handling the unique terminologies of financial posts. Moreover, there is an abnormal difference between false negatives and false positives; false negatives being responsible for over 80% of mislabelled results. This mostly due to a significant difference between the training data and the final objective of the model. For example, the subreddit wallstreetbets is known for its vulgar language. Their borderline insulting posts and comments could easily be confused as negative, when in fact they are showing support to other members of the community. Thus, using general sentiment analysis models to predict if a stock is bearish or bullish has its limitations.

4.3 Daily Sentiment & Stock Return

To see how the stock's public perception fluctuates over time, 500 random tweets and posts were selected from Twitter and Reddit for Tesla and the S&P 500. These posts were labeled 1 for bullish and 0 for bearish. A daily average of these labels was computed, the closer to 1 indicating a stronger bullish sentiment, and the closer to 0 indicating a stronger bearish sentiment.

For Tesla, the average sentiment was about 0.66 with a standard deviation of 0.09. Additionally, the daily sentiment maintained itself above 0.5, with exceptions in the first half of May. This reflects the bullish and growing sentiment that people had throughout the year. This is no surprise as the stock had an unprecedented rally of 700%, in which it became the most valuable car company. When comparing Tesla's daily sentiment to its price, there exists a Pearson correlation coefficient of 0.73. Indicating a high correlation between the two.

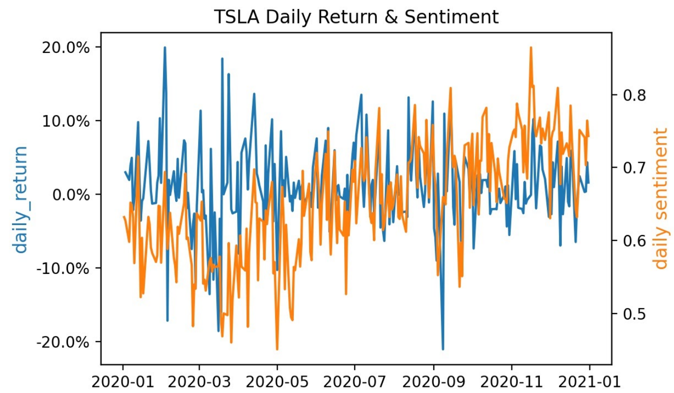

Figure 4.1: Tesla daily sentiment compared against daily return

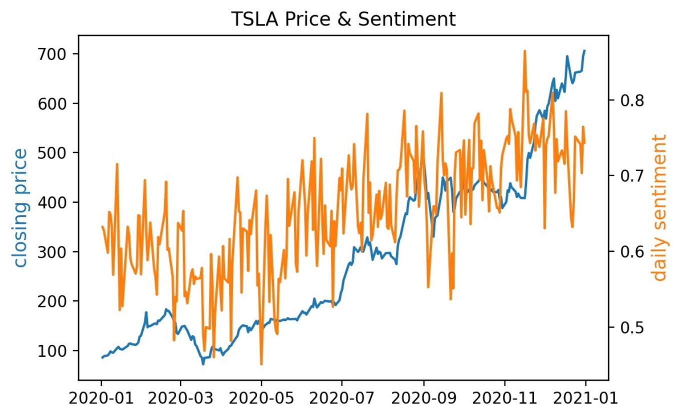

Figure 4.2: Tesla daily sentiment compared against closing price

On the other hand, the S&P 500 had an average sentiment of 0.56 with a standard deviation of 0.05. The stock maintained itself above the 0.5 neutral thresholds even though major dips at the end of February and beginning of March. These daily sentiment dips correspond to one of the major drops in the S&P 500's history, where the index fell over 30% due to COVID-19. Nevertheless, the public perception maintained itself on the bullish side and the index saw a 16% increase in 2020. When comparing the S&P 500 daily sentiment to its price, there exists a Pearson correlation coefficient of 0.4, which does not indicate a strong correlation between the two.

Although Tesla was added to the S&P 500 on December 21st the two financial assets differ in their fundamentals (Arnott, R, 2020). The S&P is well known for its stability and long-term growth, while Tesla for its high volatility, which many would categorize as a high-risk investment. These differences can be seen in the public

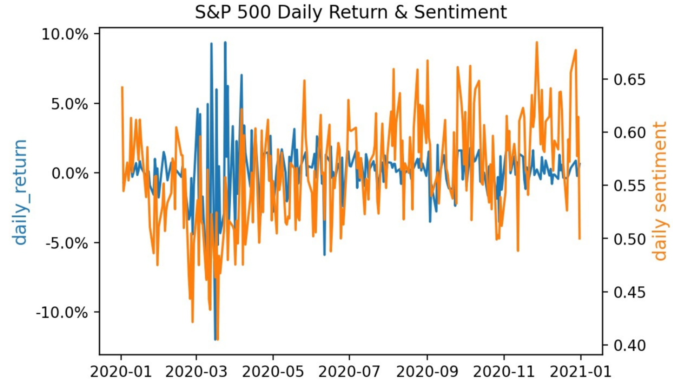

Figure 4.3: S&P 500 daily sentiment compared against daily return

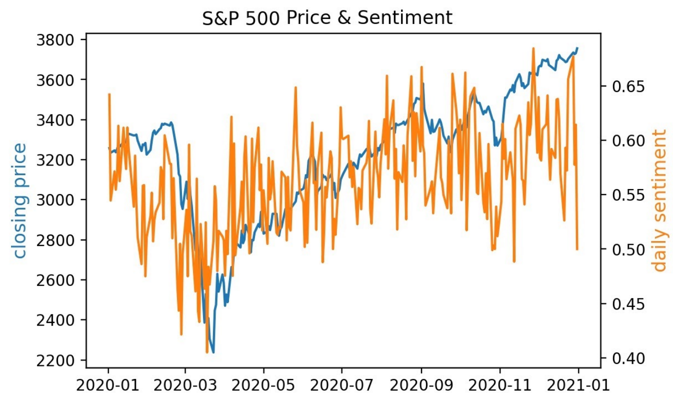

Figure 4.4: S&P 500 sentiment compared against closing price

perception of each asset. Tesla has an 80% higher standard deviation than the S&P 500, which could be described by the stock's high volatility mentioned before. Apart from having a higher standard deviation, Tesla also has a higher average daily sentiment, likely pointing that people that follow Tesla have great confidence in it.

In terms of correlation, stocks that have higher volatility appear to be highly connected to the public perception found in social media platforms. It has been shown that younger demographics tend to take part in riskier investments (Rana, M., H., 2011). Thus, younger digitized generations, which are inclined to favor high-risk investments, serve as constant inputs of financial opinions on social media platforms. A possible explanation why Tesla has a higher correlation to daily sentiment compared to the S&P 500, which is commonly known for its stability and favored by "mature" investors.

It is also important to address the effects that COVID-19 had on the markets.

Both Tesla and the S&P 500 suffered from significant drops, yet both of them reached all-time highs a short time after. Yet, despite this significant variation, the BERT remained accurate considering the limited size of the training data.

Chapter 5

Next Steps

5.1 Future Applications & Improvements

With the ever-increasing influx of data and computing power, it is no surprise that there has been a focus on quantitative trading. Quantitative or algorithmic trading is a method in which an algorithm leverages different variables, such as but not limited to price and volume, to trade securities in an automated manner (Lin, T. C., 2012). Over the recent years, there has been a surge in quantitative trading, it is estimated that around 92% of all transactions in the Forex Market fall under this category (Kissell, R., 2021). Yet, as Gamestop proved earlier this year, the human element cannot be dismissed from the equation.

To exemplify this a very simple recurrent neural network (RNN) was created to predict the price of Tesla in the last 100 days of 2020. A Long short-term memory RNN architecture was used since it has proven itself very competent at solving time-series problems and is very popular among trading algorithms. Historical 2020 data was downloaded from Yahoo Finance for Tesla and this was split into testing, containing the first 200 days, and testing, containing the remaining 100 days. For the base model, shown in Figure 5.1, only the day, adjusted closing price, volume, high, and low variables were used and it relied on the previous 30 days to predict the price. An additional model was created, Figure 5.2, similar to the base model with the exception that this model contained a variable for average daily sentiment, which was calculated in the previous chapter.

Although both models show a similar trend, the one containing the average

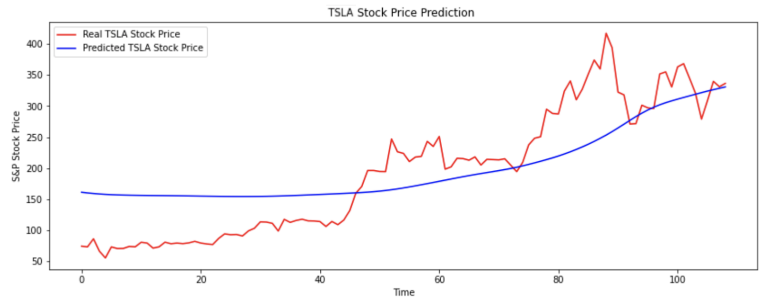

Figure 5.1: Tesla price prediction using RNN

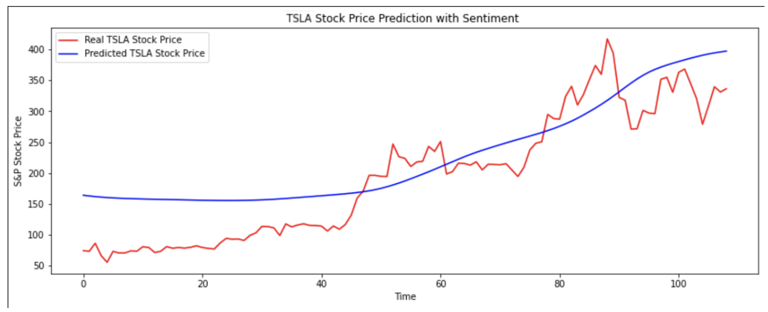

Figure 5.2: Tesla price prediction using RNN & daily sentiment

daily sentiment layer had a 12% higher accuracy compared to the base model. Even though the difference can be considered small there is a measurable benefit in a subjective/human-based layer into a basic stock prediction model. This is a very simplified model and by no means should it be used to predict the future value of Tesla. Yet its purpose is to serve as an example of the impact of adding a simple public sentiment layer that can increase the accuracy of stock forecasting models.

Chapter 6

Conclusion

Trading appears to have reached the masses as the retail investor gains influence in an industry previously dominated by big financial institutions. Zero commission trading platforms like Robinhood combined with online communities created an environment for this type of investor to gain traction and have a measurable impact, which was only accelerated by COVID-19.

Due to this, public perception appears to be one of the major forces driving today's market and it was successfully quantified in this research paper. Through the creation of the first public financial annotated corpus and its application to the fine-tuned state-of-the-art BERT language model, a 90% accuracy was achieved when classifying text into bearish or bullish. As previously seen, this accuracy cannot be achieved through traditional sentiment analysis, which is used in various research papers in this field and requires a deeper understanding of financial terms, as well as the social media platforms, where the posts are retrieved. Making it the first targeted NLP model for financial sentiment recognition.

Using this now trained language model, it was proven that there is a correlation between a stock's sentiment found in social media platforms and its movement. It is no surprise that this correlation is higher for "popular" stocks, such as Tesla, yet even in one of the most stable indexes like the S&P 500, we can see a positive correlation. Pointing that a stock prince time-series forecasting model would benefit from this type of data.

To prove this a Long-Short-Term Memory RNN model was developed and tested. Proving our hypothesis correct, the model that had the public perception as an extra variable showed an increased accuracy. Even though this forecasting model could be considered basic its purpose is not to be used as a final result, but as an example of how the NLP can bring new insights to this field. Moreover, a more comprehensive price prediction model is currently being built based on the findings from this research paper.

Although the BERT model could benefit from additional training data, the model offered promising results and when added to the stock price forecasting model there is a measurable benefit. Now that a framework has been established the next steps already being taken are increasing the corpus created, improving the BERT language model with this data, and applying it to the price prediction model. It is expected to have an applicable model by the end of the year that will help drive personal investment.

References

Akbik, Alan et al. (2019). 'FLAIR: An easy-to-use framework for state-of-the-art NLP'. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pp. 54–59.

Annamoradnejad, Issa and Gohar Zoghi (2020). 'Colbert: Using bert sentence embedding for humor detection'. In: *arXiv preprint arXiv:2004.12765*.

Black, Fischer (1986). 'Journal of Finance, Volume 41, Issue 3, Papers and Proceedings of the Forty-Fourth Annual Meeting of the America Finance Association, New York, New York, December 20-30, 1985 (Jul., 1986), 529-543.' In: *The Journal of finance* 41.3.

Boylston, Christian et al. (2021). 'WallStreetBets: Positions or Ban'. In: *arXiv preprint arXiv:2101.12110*.

Brownlee, Jason (2018). 'What is the Difference Between a Batch and an Epoch in a Neural Network?' In: *Deep Learning; Machine Learning Mastery: Vermont, VIC, Australia*.

Burghardt, Matthias, Marcel Czink and Ryan Riordan (2008). 'Retail investor sentiment and the stock market'. In: *Available at SSRN 1100038*.

Chohan, Usman W (2021). 'Counter-Hegemonic Finance: The Gamestop Short Squeeze'. In: *Available at SSRN*.

Devlin, Jacob et al. (2018). 'Bert: Pre-training of deep bidirectional transformers for language understanding'. In: *arXiv preprint arXiv:1810.04805*.

Håkansson, Anders et al. (2020). 'Gambling during the COVID-19 crisis–a cause for concern'. In: *Journal of addiction medicine* 14.4, e10.

Hayes, Adam (Apr. 2021). *Retail Investor Definition*. URL: `https://www.investopedia.com/terms/r/retailinvestor.asp`.

Horev, Rani (2018). 'BERT Explained: State of the art language model for NLP'. In: *Towards Data Science, Nov* 10.

Kissell, Robert (2021). *Algorithmic trading methods: applications using advanced statistics, optimization, and machine learning techniques*. Academic Press.

Lin, Tom CW (2012). 'The new investor'. In: *UCLA L. Rev.* 60, p. 678.

Lyócsa, Štefan, Eduard Baumöhl and Tomáš Výrost (2021). 'YOLO trading: Riding with the herd during the GameStop episode'. In:

Makrehchi, Masoud, Sameena Shah and Wenhui Liao (2013). 'Stock prediction using event-based sentiment analysis'. In: *2013 IEEE/WIC/ACM International Joint*

Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). Vol. 1. IEEE, pp. 337–342.

Mittal, Anshul and Arpit Goel (2012). 'Stock prediction using twitter sentiment analysis'. In: *Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf)* 15.

Pagano, Michael S, John Sedunov and Raisa Velthuis (2021). 'How did retail investors respond to the COVID-19 pandemic? The effect of Robinhood brokerage customers on market quality'. In: *Finance Research Letters*, p. 101946.

Qiao, Yifan et al. (2019). 'Understanding the Behaviors of BERT in Ranking'. In: *arXiv preprint arXiv:1904.07531*.

Rana, MH et al. (2011). 'Effects of demographic factors on risky decision-making behaviour'. In: *European Journal of Social Sciences* 25.3, pp. 69–76.

Ripley, Brian D (2007). *Pattern recognition and neural networks*. Cambridge university press.

Ross, Joel et al. (2010). 'Who are the crowdworkers? Shifting demographics in Mechanical Turk'. In: *CHI'10 extended abstracts on Human factors in computing systems*, pp. 2863–2872.

Tenney, Ian, Dipanjan Das and Ellie Pavlick (2019). 'BERT rediscovers the classical NLP pipeline'. In: *arXiv preprint arXiv:1905.05950*.

Tesla, the Largest-Cap Stock Ever to Enter S&P 500: A Buy Signal or a Bubble? (N.d.). URL: https://www.researchaffiliates.com/en_us/publications/articles/819-tesla-the-largest-cap-stock-ever.html.

Vaswani, Ashish et al. (2017). 'Attention is all you need'. In: *arXiv preprint arXiv:1706.03762*.

Weitzel, Leila, Ronaldo Cristiano Prati and Raul Freire Aguiar (2016). 'The comprehension of figurative language: what is the influence of irony and sarcasm on NLP techniques?' In: *Sentiment Analysis and Ontology Engineering*. Springer, pp. 49–74.

Zhang, Zijun (2018). 'Improved adam optimizer for deep neural networks'. In: *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*. IEEE, pp. 1–2.