**ECS784P – Data Analytics**

# Sentiment Analysis of r/wallstreetbets During the GameStop Saga

Alexander Murphy (MSc Computing & Information Systems) -190614555
Franklin Antony (MSc Computer Science) - 200626510

**Group size: 2**

**Student Effort Table**

| Student | Effort |
|---|---|
| Alex Murphy | 49% |
| Franklin Antony | 51% |

Table of **Contents**

# 1. Introduction

## 1.1 *Overview*

This paper uses Sentiment Analysis to examine the correlation between GameStop($GME)'s stock price and the massive surge in social media attention that the organisation experienced.

Barriers of entry to investing in the stock market became significantly lower when, in November 2019, Charles Schwab 'slashed commissions on U.S. listed stocks to zero'[1]. As a result, there has been a boom in retail investment, with CNBC reporting in January 2021 that app stores are being 'taken over' by trading platforms such as RobinHood, Revolut and Fidelity[2].

The extreme volatility that $GME experienced in the first part of 2021(low $18.84, high $483) is largely attributed to a collective of individual investors countering hedge fund led shorts on the organisation. Momentum of this counterattack was fuelled by engagement on social media - early successes of the strategy encouraged new traders to buy in, increasing traffic on investment-related forums as these traders moved to access more information and discussion.

'r/wallstreetbets' (WSB) was the epicentre of the movement – an anarchic forum that focuses on risky, speculative investment strategies and related memes. Threads on $GME alone were exceeding 200,000 comments per day at the height of interest and membership grew from 1.5 to 9 million over January and February. Contributors would often frame posts in a highly emotive manner, depicting a kind of class warfare between hedge funds and the individual investor – a reach seam for sentiment analysis!

## 1.2 *Hypothesis*

We believe that related sentiment data being generated on social media contributed to the volatility of $GME, particularly as the value of the stock far exceeded (up to 12x) fundamental analysis estimates, before fluctuating wildly as more interested parties became involved. When compelling stories of market speculation gain traction on social media, a fertile breeding ground for strong narratives, it seems intuitive that the effect of that volatility will be amplified.

## 1.3 *Project Objectives*

- Examine timeline of events to identify a particularly relevant timeframe.
- Decide on appropriate datasets.
- Conduct in-depth exploration of the obtained dataset to gain context.
- Use multiple machine learning models to predict polarity to individual posts.
- Compare frequency of different polarity with movement of stock price, generating an aggregated visual representation of this interaction and use this information to evaluate our hypothesis. Are they correlated?
- Evaluate the output and accuracy of each machine learning model.
- Assess limitations of our analysis and suggest improvements that can be made for future iterations of these methods.

# 2. Literature Review

Investing in stocks and shares is a risky business – it is well accepted in financial circles that the movement of stock prices is uncertain, but to what extent?

The Efficient Market Hypothesis (EMH) states that prices are determined by all available information, with the implication being that prices do not move predictably enough (because of constant new information) to guarantee profits to investors (Cowles, 1933, 1944). Eugene Fama (1970) added nuance by decomposing real world markets

[1]Source:https://www.cnbc.com/2019/10/01/charles-schwab-is-eliminating-online-commissions-for-trading-in-us-stocks-and-etfs.html
[2] Source: https://www.cnbc.com/2021/01/29/robinhood-investment-apps-dominate-app-store-rankings.html

into categories of efficiency, delineated by actual available levels of information available and the 'weak' EMH remains popular to this day.

The common occurrence of 'bubbles', where assets become highly overpriced, seems to challenge the EMH's core principle of investors behaving rationally regarding available information. Teeter and Sandberg (2017) argue that bubbles are social phenomena, with seemingly irrational prices reflecting social trends and narratives.

As a means of identifying links between social media activity and the $GME bubble, SA can provide quantitative and qualitative assessment of the volume and nature of opinion. With 99% of papers on the subject being published after 2004, use of SA's modern web-based incarnation soared 100x between 2005-2016 (Mäntylä, 2020).

While first used mainly as a means of aggregating review data on Amazon et al., attention has turned to extracting public sentiment on social media, particularly Twitter - 3 of the 20 top cited papers of 2020 in data analytics centred on working with data from Twitter (Mäntylä, 2020).

Earlier papers in this field focused on classifying tweets into a simple sentiment/non sentiment classification. As interest in the subject grew, more sophisticated machine learning models were developed to identify polarity – classifying statements as positive, negative or neutral (Pang et al., 2002).

When dealing with sentiment surrounding the stock market, looking at just polarity is suboptimal. The *type* of positive or negative statements is important to consider (Read, 2005). For example, excitement about purchasing shares at a low price has a different real-world impact to happiness about selling shares at a high price, even though both emotions are positive. (Acheampong et al., 2020) investigate Emotion Detection as a finer grained classification approach, comparing the abundance of advancements made in more recent years.

## 3. Timeline Investigation

Due to WSB's subscriber count being in the millions, there is an unwieldy amount of data that can be explored – as mentioned in the introduction, threads on $GME alone were exceeding 200,000 comments *per day* in some cases. To focus the scope of the project down to a manageable scale, we examined the timeline of developments to find events that triggered peak levels of engagement. This is to ensure that we get a selection of sentiment available from the widest possible pool of users, hopefully providing a picture as close to the reality of the situation as possible.

Using our own knowledge (having both bought shares and actively engaged with WSB), we felt the most significant development in the context of wider engagement with the GME phenomenon was Elon Musk tweeting 'Gamestonk!' on January 26. As of 9:50 am the next morning the price had increased by 105%[3] and we suspect that Musk's endorsement was the primary trigger for this dramatic rise in engagement and, consequently, the subsequent price volatility over the next week. Because we want to capture the ups and downs of the saga, we selected a time frame a couple of days after Musk's tweet to counter the innate bias towards positive sentiment at that specific juncture.

## 4. Data Retrieval

### *4.1 Reddit Data*

Reddit is an extremely popular social network and the rise of WSB's visibility in mainstream media, given the real-world implications of the GameStop saga, has only increased engagement. Combined with Reddit's straightforward API, relevant and substantial datasets were easy to find. All datasets were imported as CSV files and then extracted into pandas data frames.

- For prediction, we are using **`reddit_wsb.csv`**, a substantial set containing nearly 45,000 posts drawn from the week after Musk's tweet from Kaggle. Attributes are:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **`title`** | **`score`** | **`id`** | **`url`** | **`comms_num`** | **`created`** | **`body`** | **`timestamp`** |

[3] Source: https://markets.businessinsider.com/news/stocks/gamestop-stock-price-elon-musk-gamestonk-tweet-extends-trading-rally-2021-1-1030009065

### 4.2 Labelled Reddit Data

- For training and testing our models, we will be using **`Reddit_wsb_labelled.csv`**, which provides large set of posts with associated polarity. Attributes are:

| | |
|---|---|
| `clean_comment` | `category` |

### 4.3 Historical Stock Price Data

To gather GME's historical stock price data we used the financial research API Tiingo to download the ticker values of the stock from the 28th January 2021 to the 5th February 2021. This is stored in the variable **`historical_prices`** which is then merged with the **`Reddit_wsb_labelled`** dataset using Pandas at the later stages.

## 5. Tools and Libraries

For the purposes of this report, we will be using the Python programming language. Written in C, Python has emerged as the industry standard for conducting data analysis. With its intuitive syntax the language has been widely adopted and a wide range of powerful, relevant libraries have been developed within its framework. Listed below are the important libraries used:

- Pandas: A popular package combining simple-to-use data extraction utilities with powerful analytical tools and structures.
- Numpy: An important reason for Python's ubiquity in the scientific community. Provides support for large multidimensional arrays and the mathematical computing power of more fundamental languages like C and Fortran.
- Matplotlib: A data visualisation module that allows easily digestible plotting of Pandas Data Frames.
- Plotly: Another data visualisation package that utilises JSON to plot interactive graphs online.
- Seaborn: An extension of Matplotlib for advanced visualisation tools, such as combining graphs to produce an aggregate display of the data in the context of this report.
- Scikit-learn: This package provides the machine learning models we will be using to classify posts as positive or negative. The library interacts closely with Numpy.
- Re: This library is used to load the Regular Expressions to clean the text and other pre-processing needed.
- Text2emotion: This package is used for classifying posts into 5 emotional states for our data exploration: Happiness, Sadness, Surprise, Anger and Fear.

## 6. Data Exploration

Before modelling can begin, it was important to examine and visualise the data to provide context for our analysis and support our claims. All code is documented in the appendices.

### 6.1 Engagement by day of the week

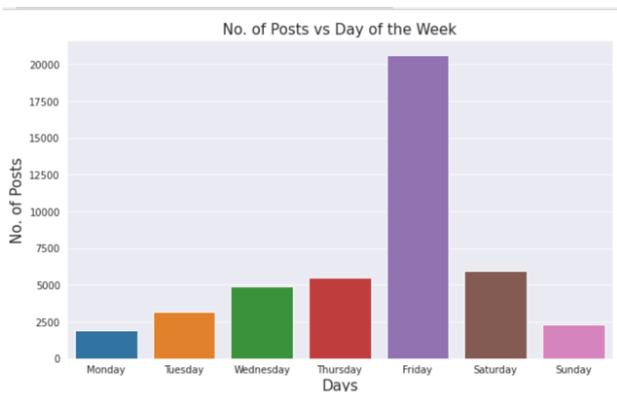


*Figure 1: Number of posts, by weekday*

Figure x shows an extraordinary leap in engagement on Fridays. This on the outset would appear odd – the stock has experienced white knuckle volatility on all weekdays. One would expect the levels of engagement to reflect that with a more even distribution across weekdays. One explanation could be that there is a greater level of positivity (and intoxication!) on Friday that encourages more engagement. Another interesting theory is related to options (i.e., calls and puts) expiring on the third Friday of every month. This could be a result of using bot farms to influence investors and protect assets, although this assertion is little more than a conspiracy theory and should be verified in further work.

### *6.2 Commonly used words*

This frequency analysis chart shows just how dominant the subject of $GME is in the WSB community. This one word had over twice the frequency of the ubiquitous investment term for any stock, 'buy'.

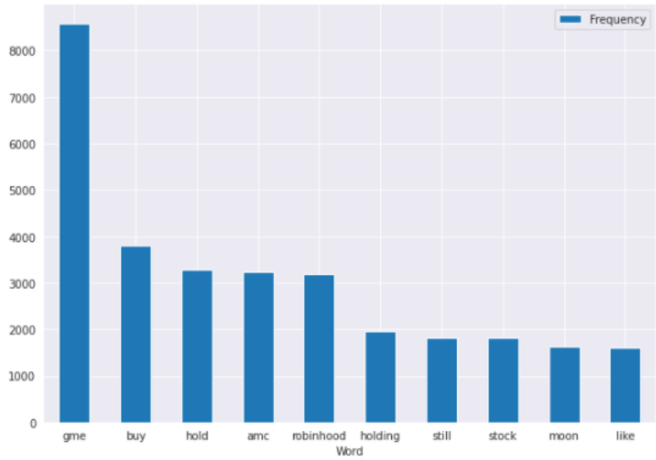



*Figure 2: Frequency histogram of most popular words on WSB*

### *6.3 Emotion Detection*

Using the Text2Emotion library, we assembled an aggregated time series of trends in emotions by day. Figure x shows that 'Fear' is by far the most dominant emotion in the posts we have examined. This supports the hypothesis as fear of missing out, or 'FOMO' is a well-recognised phenomenon that would increase buying if the stock is performing well, whether as fear of substantial losses if the stock is dipping would be an incentive to sell.

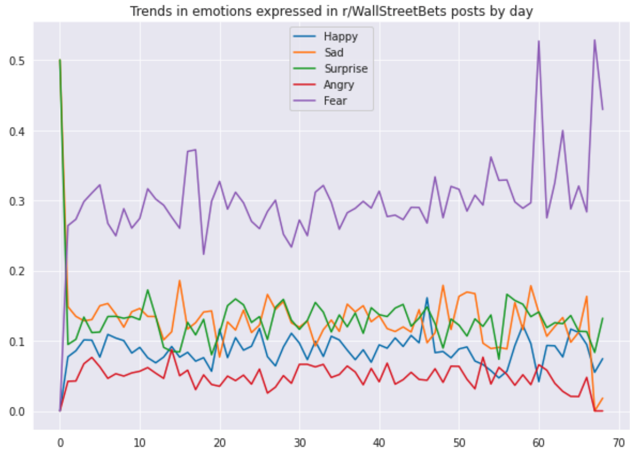



*Figure 3: Aggregate time series of 5 emotional states*

### *6.4 Stock Price*

This delta graph (delta is the difference between closing price and opening price) illuminate the volatility of the stock, confirming to us why this is a phenomenon that is worth trying to analyse.

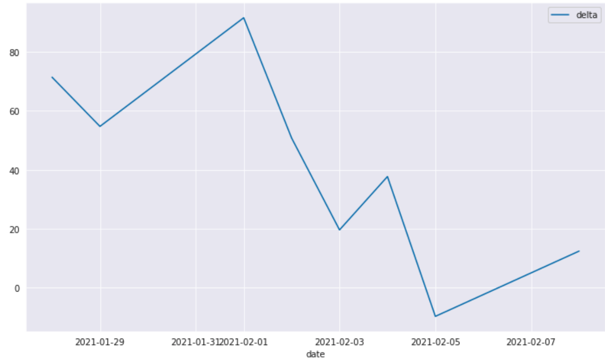



*Figure 4: Delta Representation of change in stock price.*

## 7. Data Cleaning

Machine Learning requires that the data be cleaned of any extraneous details to perform properly. To render the data appropriate for feeding into our choice models we must perform the following tasks on the Pandas data frame:

### *7.1 Drop Redundant Columns*

Firstly, we examine our datasets and select the attributes that are of importance. For the purposes of this investigation, we only need title and timestamp from the reddit data, and the price point was stored as a variable. `Reddit_wsb_labelled.csv` is a curated dataset for this purpose and had already been cleaned (title, datetime and polarity with no null values in the correct format) so these steps only apply to `reddit_wsb.csv`

```
del df['id','body',]
```

### *7.2 Dealing with Missing Values*

Once the frame has been stripped of unnecessary attributes, any tuples that do not contain a title are dropped.

### *7.3 Cleaning Text*

Using Python's `lower()` function, all capital letters in the posts are transformed to lowercase letters for continuity. Using Python's regular expression module - handlers, URL's, single characters and extra spaces are all replaced with an empty string using the `re.sub()` command. Special characters are dealt with by the `re.findall()` command.

## 8. Modelling

For the purposes of this report, we are interested in the binary classification of either positive or negative sentiment, as our hypothesis revolves around the core concept of volatility being a result of heightened emotional reaction, hence we have removed neutral labels during data preprocessing. Our three supervised learning models have been optimised to predict a post being of one of these two categories. All models were drawn from Scikit-learn's sklearn library.

Although we are working with text, we need to transform the data into information that the algorithm can manipulate and draw tangible results from. To do feature selection, we used `CountVectorizer` from the sklearn library to transform words into numbers, since that is what algorithms work with.

### *8.1 Logistic Regression*

Logistic Regression in Data Analysis is a means of streamlining data into binary values, informed by a given boundary set between 0 and 1. In this case 0 and 1 represent negative and positive sentiment respectively in the predictor space.

#### 8.1.1 Justification

For stock market analysis applications, logistical regression is a computationally inexpensive means of gaining a high-level overview of the sentiment data. Although the scope is limited to binary values, it can be used as a means of ongoing analysis due to its quick processing time, triggering alerts for a finer grained (I.e., emotion detection) approach if there is a substantive swing from positive to negative sentiment.

#### 8.1.2 Model Implementation

The model was implemented using LogisticRegression classifier from sklearn library:

```
from sklearn.linear_model import LogisticRegression
```

The data was split into two, 80% for training and 20% for testing:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
```

The data was fitted the model using:

```
lr.fit(X_train_res, y_train_res)
```

The fit function will apply needed pre-processing and regularization for better accuracy. We get a fit score of ~87%. After this the model is executed.

### 8.2 Linear SVC

LinearSVC is a form of Support Vector Model (SVM) that maps data points that are not linearly separable to a space where they are.

#### 8.2.1 Justification

Within the scope of the report, this means that we should be able to differentiate between different polarities even if the posts contain words that might suggest a different sentiment. With WSB post data, this is a particularly useful tool as posts are often shrouded in irony or sarcasm.

#### 8.2.2 Model Implementation

The LinearSVC classifier from sklearn library was used to implement the model:

```
From sklearn.svm import LinearSVC
```

The data was split into two, 80% for training and 20% for testing:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
```

The data was fit into the model using:

```
lsvc.fit(X_train_res, y_train_res)
```

The fit function will apply needed pre-processing and regularization for better accuracy. We get a fit score of ~87%. After this the model is executed.

### 8.3 Multinomial Naïve Bayes

This method uses frequency analysis in the training stage to identify words and how often they appear in texts of either polarity. This will then predict the sentiment of a statement by multiplying the prior probability of the whole statement belonging to a classification by the proportional probability of each word in the statement. The sentiment is then classified by the highest posterior probability.

#### 8.3.1 Justification

Naïve Bayes is sometimes disparagingly referred to as 'Idiot Bayes' due to its rigid and simplistic methodology. However, its continued use in modern data science has shown that it continues to produce results that fare surprisingly well in accuracy measurement and cross validation tests such as F1 score.

#### 8.3.2 Model Implementation

The model was implemented using LinearSVC classifier from sklearn libary:

```
from sklearn.naive_bayes import MultinomialNB
```

The data was split into two, 80% for training and 20% for testing:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
```

The data was fit into the model using:

```
nb.fit(X_train_res, y_train_res)
```

The fit function will apply needed pre-processing and regularization for better accuracy. We get a fit score of ~83%. After this the model is executed.

### *8.4 Comparing Models*

As can be seen in figure 5, all performed well in accuracy testing, but the Multinomial NB showed a subtle reduction in accuracy when compared to the other 2. The difference in accuracy is negligible and doesn't make a particularly strong case for use of Linear SVC or Logistic regression.

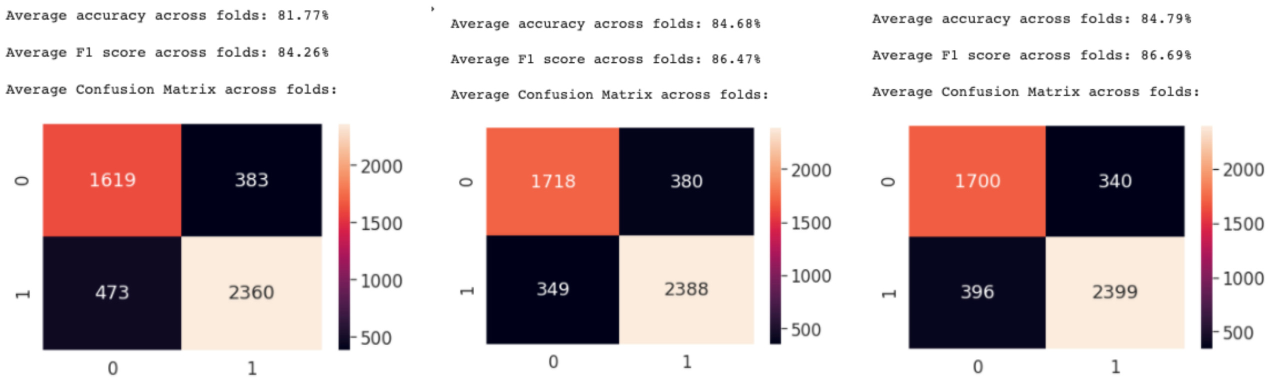



*Figure 5: (From left to right): Confusion matrices for Multinomial NB, LinearSVC and Logistic Regression respectively*

## 9. Results, Visualisation and Analysis

In order to produce an easily understandable time series that would highlight the polarity differences between models clearly, we split the mean values by date from Jan 28th to Feb 5th.

For each model, we input the separated dates to predict the polarity of each discrete period. We can see that `y_pred` contains the prediction of every row of the input dataset. To simplify plotting, we take the mean of that array using numpy.

```
np.mean(y_pred)
```

This process is repeated for all the three models. The end-of-day (EOD) share price is taken from the merged `historical_prices` dataset we obtained using TiingoClient API and plotted on a time series. On this graph we also plot the mean values outputted from the model for each day.

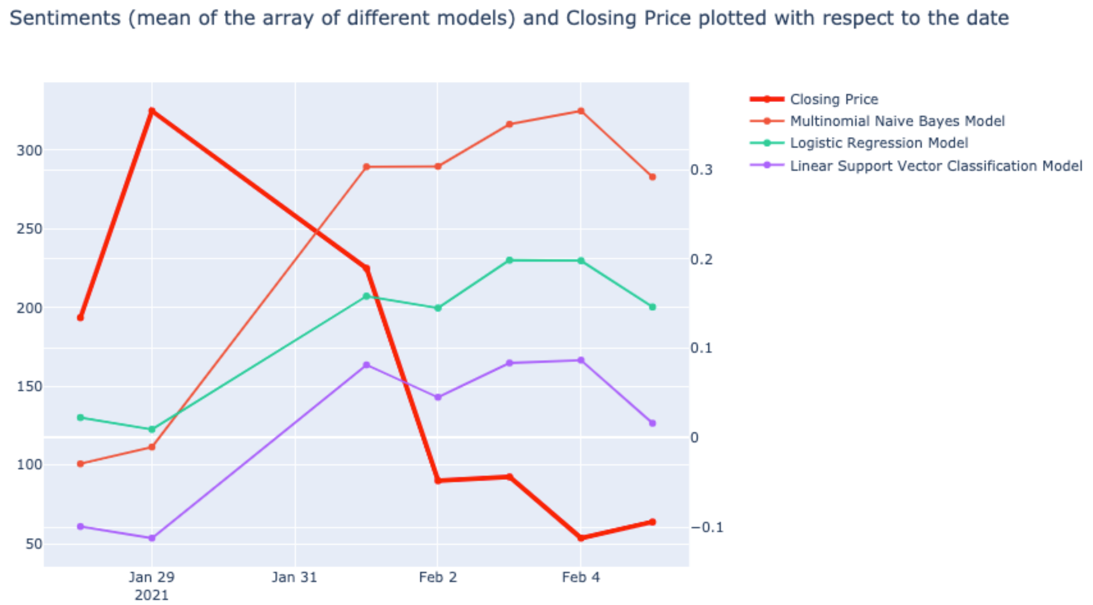



*Figure 6: Combined*

From the graph we can see that Multinomial NB model was more positive than the LR during the last day (Feb 5th) even though it was more negative in the beginning (Jan 28th). As we anticipated the LSVC and LR had an almost identical curve and the average tended toward zero(neutral).

## 10. Successes, Challenges & Validity in Business Applications

### *10.1 Successes and Challenges*

Our data exploration was insightful and of a good standard, with a well-informed picture of the quirks of the idiosyncratic environment we were trying to capture. It provided us with unexpected results, giving us the opportunity to intuit interesting theories using our personal knowledge of the situation. Having used 3 machine learning algorithms, we were able to provide a comprehensive comparison of accuracy measurements. Also, the accuracy across all three models was very good with high F1 scores across the board.

In terms of limitations to our approach, there are several important factors to consider. Firstly, WSB is a forum heavily biased toward the experiences of retail investors, who make up a relatively insignificant proportion of the marketplace. Consequently, the results might not reveal the full truth about the direction of sentiment of all players involved, an important consideration when looking at price volatility as a product of sentiment. The dataset we used was very limited given the scope of a forum that hosts millions of posts. One key aspect to the phenomenon was use of emojis, used in a high proportion of posts and usually express a very clear and strong sentiment polarity: Diamonds, apes, bananas etc were almost universally expressions of positivity, while paper, bears, rainbows etc were used almost exclusively to express negativity. If we had found a way of tokenizing these emojis it would have been sure to improve the quality of our analysis substantially.

Social media is an unreliable source by nature, especially an anonymous platform like Reddit. The platform is particularly susceptible to bots and bad-faith actors dishonestly expressing opinions, skewing the results in a potentially dramatic fashion.

### *10.2 Business applications*

If further refined to include different social media platforms and an appropriate size dataset, this model could enjoy applications in risk analysis, a fundamental practice in stock trading. A swing in sentiment polarity could be predictive of volatile movements in price and this could be used to trigger alerts to individual investors depending on their given taste for risk.

# 11. Concluding Remarks

In this report, we attempted to examine if there was any correlation between swings in the polarity of sentiment and price volatility in particularly visible stocks. In keeping with the EMH referenced in chapter 2, we must conclude that predictions using our models have **not** shown significant link between the two factors. We can clearly see that closing price of GME was not directly related to the mean of polarity of sentiments.

Overall, the polarity of posts in WSB tended toward positive rather than negative sentiment, as shown in our final sentiment/price point time series graph. Although the LR and LSVC models performed slightly better than the MNB (in terms of accuracy), we can't conclusively state that any model would provide better performance. This lack of difference suggests that there is much to be improved upon in our data modelling approach.

# 12. References


1. Cowles 3rd, A. (1933) "Can Stock Market Forecasters Forecast?" Econometrica: *Journal of the Econometric Society*, 1, 309-324.
2. Cowles 3rd, A. (1944) "Stock Market Forecasting." *Econometrica*, 12: 206-214
3. Fama, Eugene F. (1970) "Efficient Capital Markets: A Review of Theory and Empirical Work." *The Journal of Finance*, 25, 2: 383–417.
4. Teeter, Preston; Jörgen, Sandberg. (2017) "Cracking the Enigma of Asset Bubbles with Narratives." *Strategic Organization*, 15, 1: 91–99.
5. Mäntylä, Mika & Graziotin, Daniel & Kuutila, Miikka. (2016). "The Evolution of Sentiment Analysis - A Review of Research Topics, Venues, and Top Cited Papers." *Computer Science Review*. 27. 10.1016/j.cosrev.2017.10.002.
6. B. Pang, L. Lee, S. Vaithyanathan (2002)." Thumbs up? Sentiment Classification using Machine Learning Techniques"
7. Read, J. (2005) "Using Emotions to Reduce Dependency in Machine Learning Techniques for Sentiment Classification."
8. Acheampong, F. A.; Wenyu, C.; Nunoo-Mensah, H. (2020) "Text-based emotion detection: Advances, challenges, and opportunities"


# 13. Appendices

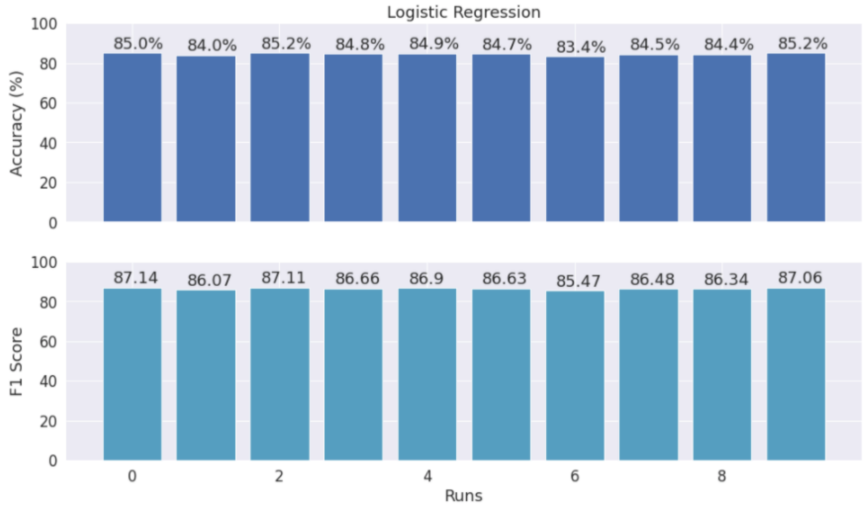
## *13.1 Notebooks*

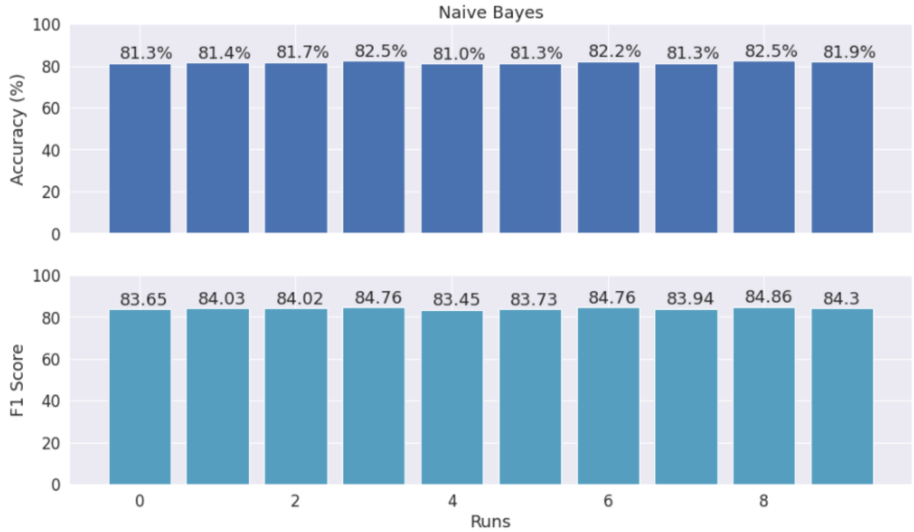
Data Exploration -https://colab.research.google.com/drive/1UcK-N9yj0B9mnYOspEVVCnPM5X3aL3ou?usp=sharing

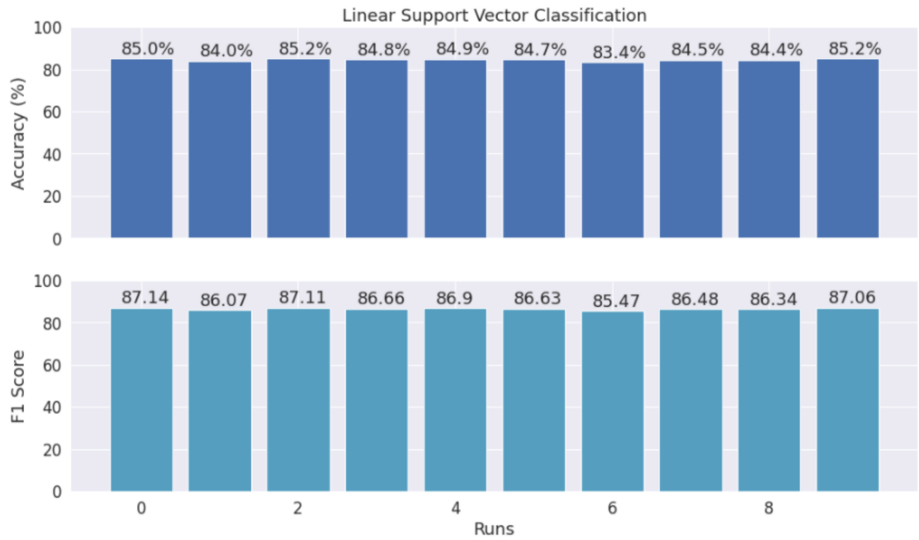
ML Model and Analysis - https://colab.research.google.com/drive/1e7eecMTrn_IENQ_oxw3d6GQMPJubjkB9?usp=sharing

**Cross Validation**

The accuracy and F1 Scores




The data from reddit_wsb.csv shown as a table using pandas

| | title | score | comms_num | body | timestamp |
|---|---|---|---|---|---|
| **0** | It's not about the money, it's about sending a... | 55 | 6 | NaN | 2021-01-28 21:37:41 |
| **1** | Math Professor Scott Steiner says the numbers ... | 110 | 23 | NaN | 2021-01-28 21:32:10 |
| **2** | Exit the system | 0 | 47 | The CEO of NASDAQ pushed to halt trading "to g... | 2021-01-28 21:30:35 |
| **3** | NEW SEC FILING FOR GME! CAN SOMEONE LESS RETAR... | 29 | 74 | NaN | 2021-01-28 21:28:57 |
| **4** | Not to distract from GME, just thought our AMC... | 71 | 156 | NaN | 2021-01-28 21:26:56 |

The pre-processed and cleaned data given as input for the model

| | clean_comment | category |
|---|---|---|
| 0 | it not about the money it about sending message | 0 |
| 1 | math professor scott steiner says the numbers ... | -1 |
| 2 | exit the system | 0 |
| 3 | new sec filing for gme can someone less retard... | -1 |
| 4 | not to distract from gme just thought our amc ... | 1 |
| ... | ... | ... |
| 44265 | caught these mfs in 8k | 0 |
| 44266 | enjoy | 1 |
| 44267 | took an outside bet on bidu 2 weeks later | 0 |
| 44268 | degiro restricting order to market for gme amc... | -1 |
| 44269 | ask fiona cause coty still doesn fucking know | 0 |

44270 rows × 2 columns

The historical stock price data retrieved from TiingoClient

| | index | date | close | high | low | open | volume | adjClose | adjHigh | adjLow | adjOpen | adjVolume | divCash | splitFactor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 2021-01-28 | 193.60 | 483.0000 | 112.2500 | 265.00 | 58815805 | 193.60 | 483.0000 | 112.2500 | 265.00 | 58815805 | 0.0 | 1.0 |
| 1 | 1 | 2021-01-29 | 325.00 | 413.9800 | 250.0000 | 379.71 | 49414294 | 325.00 | 413.9800 | 250.0000 | 379.71 | 49414294 | 0.0 | 1.0 |
| 2 | 2 | 2021-02-01 | 225.00 | 322.0000 | 212.0000 | 316.56 | 37382152 | 225.00 | 322.0000 | 212.0000 | 316.56 | 37382152 | 0.0 | 1.0 |
| 3 | 3 | 2021-02-02 | 90.00 | 158.0000 | 74.2201 | 140.76 | 76663691 | 90.00 | 158.0000 | 74.2201 | 140.76 | 76663691 | 0.0 | 1.0 |
| 4 | 4 | 2021-02-03 | 92.41 | 113.3999 | 85.2500 | 112.01 | 42698511 | 92.41 | 113.3999 | 85.2500 | 112.01 | 42698511 | 0.0 | 1.0 |
| 5 | 5 | 2021-02-04 | 53.50 | 91.5000 | 53.3300 | 91.19 | 62427275 | 53.50 | 91.5000 | 53.3300 | 91.19 | 62427275 | 0.0 | 1.0 |
| 6 | 6 | 2021-02-05 | 63.77 | 95.0000 | 51.0900 | 54.04 | 81345013 | 63.77 | 95.0000 | 51.0900 | 54.04 | 81345013 | 0.0 | 1.0 |
| 7 | 7 | 2021-02-08 | 60.00 | 72.6600 | 58.0200 | 72.41 | 25687282 | 60.00 | 72.6600 | 58.0200 | 72.41 | 25687282 | 0.0 | 1.0 |

The combination of historical stock price and the post title made

| | title | score | comms_num | body | timestamp | Weekday | alltexts | happy | angry | surprise | sad | fear |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | It's not about the money, it's about sending a... | 55 | 6 | NaN | 2021-01-28 21:37:41 | Thursday | It's not about the money, it's about sending a... | 0.0 | 0.00 | 0.00 | 0.00 | 1.00 |
| 1 | Math Professor Scott Steiner says the numbers ... | 110 | 23 | NaN | 2021-01-28 21:32:10 | Thursday | Math Professor Scott Steiner says the numbers ... | 0.0 | 0.25 | 0.25 | 0.25 | 0.25 |
| 2 | Exit the system | 0 | 47 | The CEO of NASDAQ pushed to halt trading "to g... | 2021-01-28 21:30:35 | Thursday | Exit the system The CEO of NASDAQ pushed to ha... | 0.0 | 0.00 | 0.00 | 0.00 | 1.00 |
| 3 | NEW SEC FILING FOR GME! CAN SOMEONE LESS RETAR... | 29 | 74 | NaN | 2021-01-28 21:28:57 | Thursday | NEW SEC FILING FOR GME! CAN SOMEONE LESS RETAR... | 0.0 | 0.00 | 0.00 | 1.00 | 0.00 |
| | Not to | | | | | | Not to | | | | | |

| | Happy | Angry | Surprise | Sad | Fear | date | hour | post_count | index | close | high | low | open | volume |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.0 | 0.00 | 0.00 | 0.00 | 1.00 | 2021-01-28 | 21 | 1 | 0 | 193.6 | 483.00 | 112.25 | 265.00 | 58815805 |
| 1 | 0.0 | 0.25 | 0.25 | 0.25 | 0.25 | 2021-01-28 | 21 | 1 | 0 | 193.6 | 483.00 | 112.25 | 265.00 | 58815805 |
| 2 | 0.0 | 0.00 | 0.00 | 0.00 | 1.00 | 2021-01-28 | 21 | 1 | 0 | 193.6 | 483.00 | 112.25 | 265.00 | 58815805 |
| 3 | 0.0 | 0.00 | 0.00 | 1.00 | 0.00 | 2021-01-28 | 21 | 1 | 0 | 193.6 | 483.00 | 112.25 | 265.00 | 58815805 |