Research questions:

Q1: Is it possible to use twitter data to accurately predict the stock price of a company or other financial asset?

Q2: Is there a correlation between sentiment of the financial community on twitter and the long-term price of an equity asset?

Q3: Can textual twitter data specifically from financial institutions be used to predict asset or market returns based on classification of content?

Literature Review

1. <https://www.sciencedirect.com/science/article/abs/pii/S0957417416306571>

* By separating a group(s) of twitter users that could form a representative sample of the general stock market, based on metadata that would indicate their affiliation to the financial community, a sentiment analysis can be carried out to determine what current events affect the sentiments of the stock market, and hence company stock prices
* By following a 5 step architecture, the paper achieves reasonable precision by calculating a normalized sentiment score in the 4th step
* Normalization helps account for increases in tweet volumes over time as twitter’s user base grows. This is analogous to the size of stock market growing with the population and technology access.

1. <https://www.sciencedirect.com/science/article/pii/S0957417419304270>

* Sentiment extraction carried out on 9-years worth of microblog text data using dictionary approach to carry out sentiment analysis related to financial market events over 5 economic regions
* Approach was augmented from previous models by using domain specific topic word lists and creating a negation list to account for reversal in sentiment polarities
* Overall the model’s effectiveness lacked in areas of positive event sentiment, however, during periods of negative market returns significant gains could be made using the predictive model

1. <https://www.semanticscholar.org/paper/Product-Sales-Prediction-Based-on-Sentiment-Using-Gaikar-Marakarkandy/beec0b7ed38a6a9408724e87b142608ec0c8e586>

* A modified approach to estimating company-specific top line income as opposed to financial metrics based off financial participants
* Key assumption is that twitter users can represent a sample cross section of a company’s customer base, and hence sentiment toward company products can be used to predict sales
* Variables of Purchase Intention and Attitude were analysed separately to create an aggregated sentiment score toward the company

1. <https://www.researchgate.net/publication/337880061_Liquidity_Risk_and_Investors%27_Mood_Linking_the_Financial_Market_Liquidity_to_Sentiment_Analysis_through_Twitter_in_the_SP500_Index>

* An analysis on market participants’ effect on market volatility through twitter
* Bid-ask spread of a cross-section of stocks comprising the S&P500 was used as a proxy for market liquidity
* Sentiment scores calculated from unstructured tweet set was were then regressed against general bid-ask spreads
* For daily spreads the model showed fair predictive capability, however over longer periods this did not hold

1. <https://www.sciencedirect.com/science/article/pii/S2405918817300247>

* Lexicon based sentiment classification was made of a tweet set belonging to the U.K. using Umigon
* Research was augmented by focusing on political tweets and events of political nature under the presumption that a portion of stock market movements could be attributed to political climate shifts
* FTSE 100 market returns were used as a proxy to general market movements
* Sentiment scores were calculated and then regressed for tweets classified as being political in nature – while volume of tweets proved to have lower correlation statistics than were expected, correlation between overall market sentiment and daily volatility and returns also turned out to be weak

**Rejection**

* We rejected the idea because in order to add value to the already existing set of research, we would have to create models that might be too complex given our constraints
* Research has focused on using twitter data to predict singular, direct financial measures such as revenue or expected growth
* We wanted to narrow down our analysis to a user base of financially literate users e.g. public investors such as Warren Buffet or equity traders or asset management firms such as J.P. Morgan
* We would then calculate sentiment scores off text they would publish on the platform and classify their outlook on a stock as either Neutral, Bullish (good) or Bearish (bad)
* However, to forecast a company’s stock price or other relevant financial metrics relating to its health would be complicated because there are many factors that determine said prices and said metrics that are typically not captured in textual analysis, e.g. interest rate term structures or capital structure/cash-flow expectations
* There exist non-textual models that can form these predictions with greater accuracy, hence potentially rendering our contributions immaterial