How to assess Twitter's impact on financial markets

Social media channels have great potential as a source of alternate data for investors. But they present new challenges as well

BY PRODIPTA GHOSH

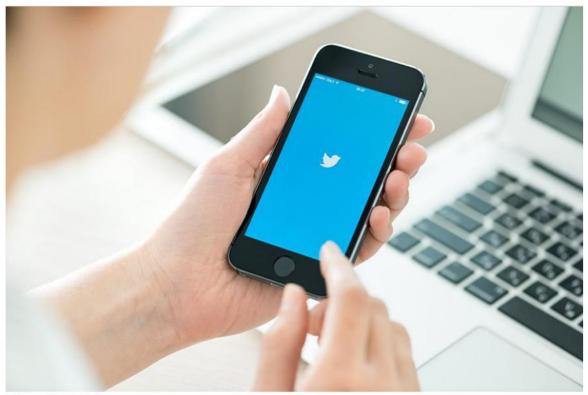4 min read

UPDATED: May 28, 2019 02:42:11 PM UTC

Image: Shutterstock

In August 2018, Elon Musk tweeted about taking Tesla private.

> **Elon Musk** ✔
> @elonmusk
>
> Am considering taking Tesla private at $420. Funding secured.
>
> ♡ 86K   10:18 PM - Aug 7, 2018
>
> 💬 21.6K people are talking about this

That led to a roller-coaster ride for Tesla investors over the next couple of months. In 2016, when US president Donald Trump tweeted about cost overruns in Lockheed Martin's projects, the stock took an immediate beating of 5 percent.

Donald J. Trump ✔
@realDonaldTrump

The F-35 program and cost is out of control. Billions of dollars can and will be saved on military (and other) purchases after January 20th.

♡ 52.7K 6:56 PM - Dec 12, 2016 ⓘ

💬 22.1K people are talking about this ›

Not just CEOs and Presidents-–in February 2018, when influencer-turned-entrepreneur Kylie Jenner posted a negative review about Snapchat, the company's shares closed 6 percent down on the day.

Kylie Jenner ✔
@KylieJenner

sooo does anyone else not open Snapchat anymore? Or is it just me... ugh this is so sad.

♡ 355K 3:20 AM - Feb 22, 2018 ⓘ

💬 71.6K people are talking about this ›

Many studies have corroborated this anecdotal evidence. One study, 'The Effects of Twitter Sentiment on Stock Price Returns', correlates tweets volumes and sentiments with returns on stocks from the Dow Jones index. This is true for both expected (for example around earnings announcements etc) and unexpected tweets and lasts for several days, it finds. An earlier study reported evidence that the sentiments on Twitter have predictive power on market returns. It also affirmed that the predictive accuracy correlates to investor under-reaction.

And it is not just about sentiments. A study from Stanford and the University of Michigan reports that company-initiated news via Twitter often leads to improved liquidity of that company's stock. This is especially true for small-cap firms, which are often under the radar of analysts from big houses. Leveraging social media in predicting the future is now an active area of research, as suggested by 'A Survey of Prediction Using Social Media', ranging from box office revenues and elections to stock markets.

Social media channels have great potential as a source of alternate data for investors. But they present new challenges as well. The sheer volume of tweets on a particular topic or hashtag might mean that it is in the realm of the machines to process these data. But computers are not usually great with such unstructured data. Advances in machine learning and natural language processing (NLP) come to our rescue here, to distill the deluge of data into sentiments and insights.

Wikipedia defines NLP as a subfield of computer science, information engineering and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyse large amounts of natural language data.

The first step is to filter the data. Hosing the tweet stream directly into a trading strategy is not going to work. At the very basic outset, we need filtering based on recency and credibility (for example, removing bots and fake accounts). Network features (for example, number of followers) and temporal features (for example, speed of re-tweets) can be used to augment this.

Second, we need to have a mapping system between tweets and the underlying stocks. This can be as straightforward as filters based on specific hashtags or stock symbols, like #AAPL. Or it can leverage advanced constructs like knowledge graph to generate semantic associations, like mentions of iPhone or Tim Cook.

The core sentiment analysis has two major approaches. One way is using NLP. The most common here is the bag-of-word method, where we parse the text and compare the count of positive vs negative words (polarity) from a known set of words (dictionary of lexicons). A common folly of such methods is getting confused with the quirks of the English language, like double negation, sarcasm etc. A more advanced method is the VADER model, specifically attuned to analyse feeds from micro-blogging sites like Twitter.

The other approach uses machine-learning methods to learn from features of the text, often from lexical analysis, such as like N-grams (An N-gram is a contiguous sequence of n items from a given sample of text or speech). It can be more versatile, but the drawback here is the need for a large labelled dataset to train the model.

With all their promises, some still remain sceptical. Sentiment-based strategies can often misfire spectacularly. Social media is open to manipulation and false rumours. On April 23, 2013, a hacked account of the Associated Press falsely tweeted about an explosion in the White House, leading to a flash crash in Dow and S&P. While such instances have been rare for the broader market indices, for individual stocks, especially small caps, it has been not so uncommon. The US market regulator, SEC, came up with an investor alert in 2015, after a series of such cases. Since 2018, the market regulator in India, SEBI, also stepped up its efforts to curb misinformation and rumour mongering on Twitter and WhatsApp. Algorithms are becoming increasingly sophisticated to handle such cases.

Microblogging sites like Twitter, with its succinct messaging, have launched a new technology race among hedge funds and sophisticated investors. The future is a place where sophistication and speed of data analytics, rather than a monopoly on the data source, will dictate alpha.

The author is the vice president at QuantInsti