

StockTwits Sentiment Classification and Stock Returns

M.-A. Divernois, D. Filipović

EPFL & Swiss Finance Institute
<https://github.com/marcaureledivernois>

November 19, 2021

Idea

Analyze predictive power of social media for stock markets

- 1 Collect & classify 90 million messages from StockTwits from 2009 to 2020
- 2 Build firm-specific daily sentiment measure (polarity)
- 3 Test predictive power of polarity on stock returns in general and around events

Keywords : NLP, Social Media, Investor sentiment, Event study

Message items

The screenshot displays a vertical list of four messages. Each message includes a user profile picture, a name, a timestamp, a ticker symbol (SAAPL), a text body, a sentiment label, and interaction icons (thumbs up, thumbs down, retweet, and share).

- Message 1:** User **satkaru** at 02:27 PM. Ticker: **SAAPL**. Text: "that's that" with five thumbs up and a rocket emoji. Sentiment: **Bullish** (green label).
- Message 2:** User **Smoothsailing** at 02:27 PM. Ticker: **SAAPL**. Text: "jesus. I'm still holding my 300 calls I bought yesterday AM." Sentiment: **Bullish** (green label).
- Message 3:** User **JohnBer** at 02:26 PM. Ticker: **SAAPL**. Text: "SPARTA!!!" Sentiment: **Bullish** (green label).
- Message 4:** User **Etrading** at 02:26 PM. Ticker: **SAAPL**. Text: "Let see how much more this Junk POP !!!! They will miss earning due to Corono Virus.. so this is going again ATH? Ridiculous". Sentiment: **Bearish** (red label).

- Ticker symbol
- Timestamp
- Text message (body)
- Sentiment label (bearish / bullish / none)
- Others (32 in total)

Sentiment classification

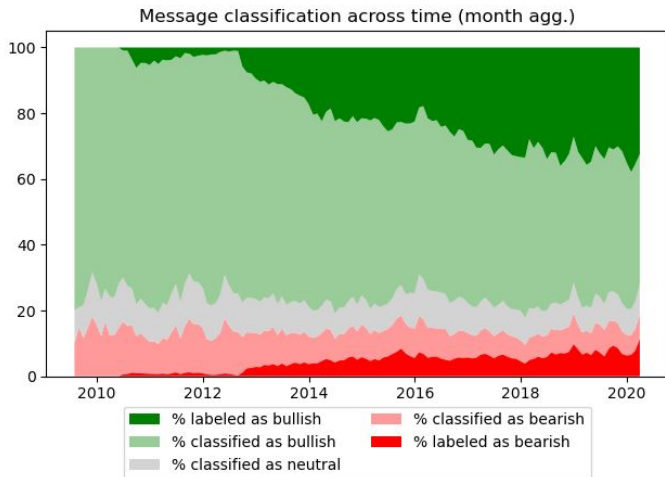
- 1 Run a logistic regression of the sentiment labels on the TFIDF features of user-labeled messages. Gives a score

$$TFIDF \mapsto s(TFIDF) = \text{prob}[\text{sentiment} = \text{bullish} \mid TFIDF]$$

- 2 Run two binary classifications based on this score $s(\cdot)$ on user-labeled messages: the first (second) classifier sets bullish (bearish) as positive class and non-bullish (non-bearish) as negative class.
- 3 Every message m is then mapped on either

$$m \mapsto \begin{cases} (\text{non-bullish}, \text{bearish}) & =: \text{bearish} \\ (\text{bullish}, \text{bearish}) & =: \text{neutral} \\ (\text{non-bullish}, \text{non-bearish}) & =: \text{neutral} \\ (\text{bullish}, \text{non-bearish}) & =: \text{bullish} \end{cases}$$

Sentiment-classified messages



Build a daily aggregate sentiment measure for every firm and the market:

Polarity

- Polarity of firm i on day t

$$P_{i,t} = \frac{\# \text{ bullish} - \# \text{ bearish}}{\# \text{ bullish} + \# \text{ bearish}} = \frac{\sum_{j=1}^{M_i^{(t)}} (\mathbf{1}_{C_{i,j}^{(t)}=1} - \mathbf{1}_{C_{i,j}^{(t)}=-1})}{\sum_{j=1}^{M_i^{(t)}} (\mathbf{1}_{C_{i,j}^{(t)}=1} + \mathbf{1}_{C_{i,j}^{(t)}=-1})}.$$

- Market polarity on day t

$$P_t^M = \text{weighted average} = \frac{\sum_i M_i^{(t)} \cdot P_{i,t}}{\sum_i M_i^{(t)}}.$$

Can polarity predict next day returns?

$$R_{i,t} = \alpha + \beta \cdot P_{i,t} + \epsilon_{i,t}$$

$$R_{i,t+1} = \alpha + \beta \cdot P_{i,t} + \epsilon_{i,t}$$

	$R_{i,t}$	$R_{i,t+1}$
Constant	-0.0047*** (0.000)	-0.0002 (0.000)
$P_{i,t}$	0.009*** (0.000)	0.0003 (0.000)
R^2	0.012	0.000
No. Obs.	34100	34100

- Polarity is positively associated with contemporaneous returns
- Polarity has no predictive power on next day stock returns in general

Firm events - AAPL

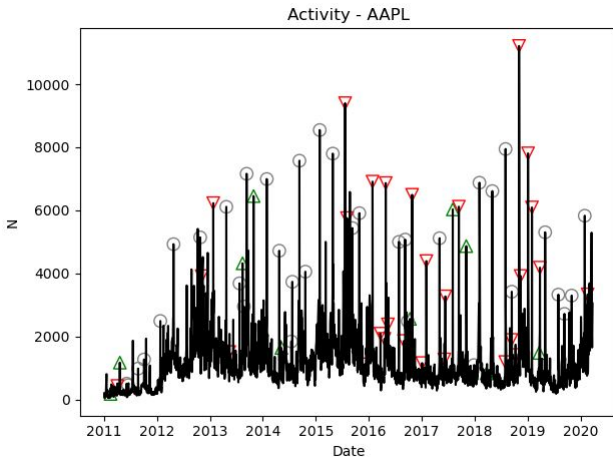
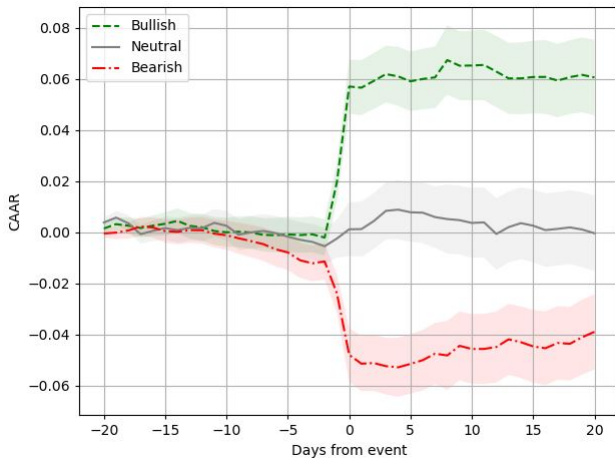
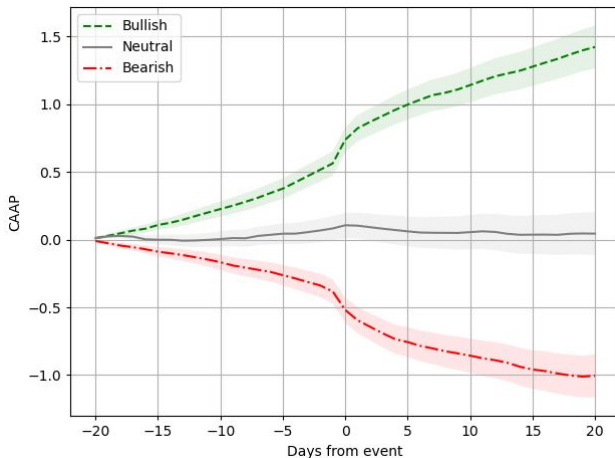


Figure: Daily volume of messages for Apple. Events are days with an unusual high volume of messages (total 1131). Green upper-triangles show bullish events, gray circles are neutral events and red down-triangles represent bearish events.

Cumulative returns around firm events



Cumulative polarity around firm events



- We extract a daily sentiment measure (polarity) for firms from StockTwits
- Polarity is positively associated with contemporaneous stock returns, but this result loses its significance against next-day returns.
- Around firm events, abnormal polarity has significant predictive power on the type of the event, as opposed to abnormal returns.
- Application: Conditioning on (or knowing) an event will happen in the near future (e.g., earnings announcement), portfolio and risk managers can get an idea of the direction of the market.