

## NLP-Driven Sentiment Analysis on the Financial Market

### Popular Online Platforms and APIs for Financial Sentiment Analysis

#### 1. [MarketBeat](#)

- Covered Assets: U.S. Stocks
- Sentiment Indicators: Algorithmic media sentiment rating; List of trending stocks
- Methodology: Most likely a mix of fundamental (earnings, revenues, etc.), technical (price movements), and NLP analysis (meaning, keywords, etc.).

#### 2. [Sentdex](#)

- Covered Assets: U.S. and London Stocks, Australian Securities, FOREX, Bitcoin
- Sentiment Indicators: NLP-based sentiment classification
- Methodology: It identifies the subject from the reading (from Reuters, WSJ, CNBC, etc. for stocks) by named entity recognition, groups text bits into noun-phrases by chunking, and decides the attitude based on adjectives. [Negative (sell): -3~-1; Neutral: 0; Positive (buy): 1~6]

#### 3. [Alternative.me](#)

- Covered Assets: Bitcoin and major cryptocurrencies.
- Sentiment Indicators: Crypto Fear & Greed Index
- Methodology: Besides technical analysis, it tracks hashtag activity and engagement on Twitter—which positively correlates with greed; used to poll retail investor sentiments; and analyzes changes in search volumes of Bitcoin-related keywords. It translates risk aversion as fear and momentum-chasing as greed. [0 Fear~100 Greed]

#### 4. [StockTwits](#)

- Covered Assets: Equities and crypto.
- Sentiment Indicators: Stocktwits Sentiment Index; Stock Sentiment Classification
- Methodology: It calculates a rolling average of retail sentiment from user-labeled bullish/bearish posts, updated daily after market close . [Greed: 0~1; Fear: 1~3]

#### 5. Online news sources Pushshift API / PRAW

- Covered Assets: Any asset discussed
- Sentiment Indicators: Custom NLP analysis outputs
- Methodology: It implements custom NLP via Python NLP libraries, such as FinBERT, to analyze comments and posts scraped from websites (such as r/stocks and r/wallstreetbets on Reddit, etc).

### Accuracy and Statistical Validation

Below are sources with published statistical validation or backtests.

#### a) Sentdex

- Backtest Process: The following simple, sentiment-based strategy is run on Quantopian: buy up to ten stocks with a sentiment score of 6 using 10% of the portfolio's available cash at a stop-loss of 0.5%, hold each stock until its sentiment hits -1 or lower.

- **Results:** The sentiment score is reflective of imminent market behavior, crafting a profitable strategy at 0.11 alpha, 1.96 Sharpe Ratio, and 12.8% Max Drawdown

#### b) Crypto Fear & Greed Index

- **Backtest Process:** The following Crypto Fear & Greed Index Enhanced Dollar-Cost Averaging (DCA) Strategy is run and compared to a pure DCA Strategy from March 2018 to September 2023: Every 7 days, invest \$150 if “Extreme Fear”, \$100 if “Fear”, \$75 if “Greed”, and \$25 if “Extreme Greed.”
- **Results:** The index makes good long-term strategies, potentially by correlating with price turning points, raising the ROI from 124.81% to 151.68% and the PNL from \$35694 to \$38866.

#### c) Stocktwits

- **Backtest Process:** The Bull-Bear sentiment ratio for each trading day of 2020 and Exponential Moving Averages (EMA) of various lookback periods are calculated. Then on each day, if the Bull-Bear ratio is higher than the EMA, a 100% net long position is held and vice versa.
- **Results:** The market sentiment has fair predictive power, especially when the sentiment is strong. The strategy fittingly was net short during Feb-March crash and September’s correction. It does particularly well for TSLA.

#### d) FinBERT

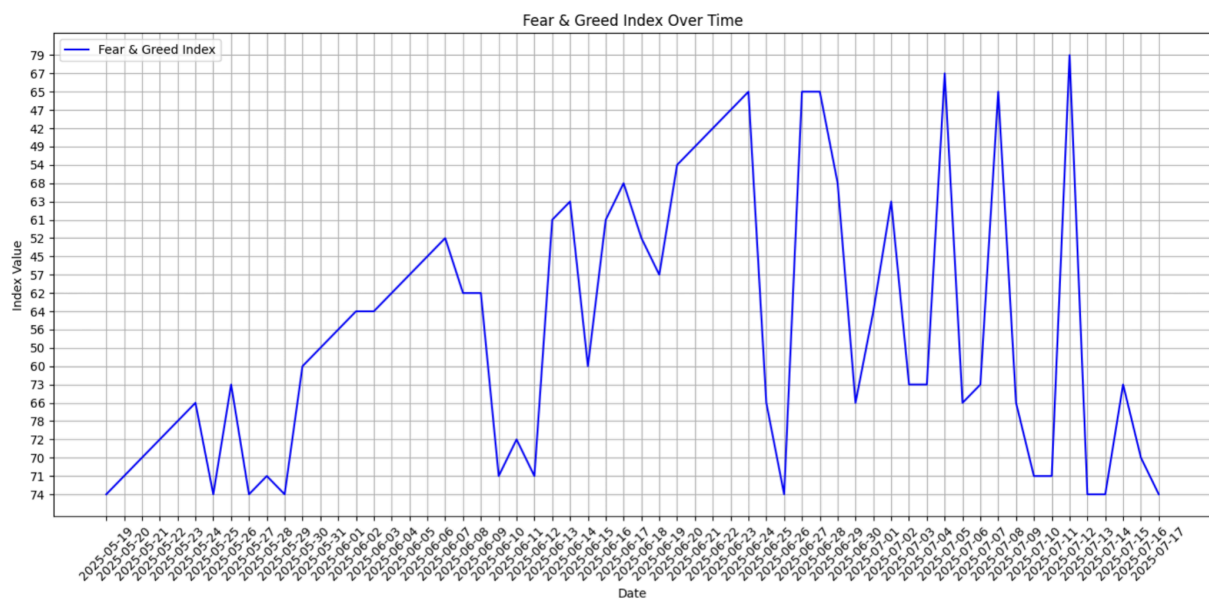
- **Backtest Process:** The model is trained, validated, and tested on the Financial PhraseBank dataset for classification—which contains thousands of financial news sentences labeled as positive, neutral, or negative at various levels of annotator agreement—and, for regression tasks, the FiQa Sentiment dataset—which contains tweets with sentiment scores between -1 and 1—and compared to other models, including LSTM and ULMFit.
- **Results:** FinBERT outperforms all other models in capturing sentiments, with an accuracy and F1 Score of 0.86 and 0.84 respectively across all data.

### Python Code and Libraries for Financial Sentiment Analysis

For full code implementation and visual analysis, please refer to the accompanying .py and .ipynb files. Sentdex and StockTwits used to offer API services, but the former is now discontinued and the latter is under maintenance until further notice. The following is an overview of the available Python code and packages.

#### a) Crypto Fear & Greed Index

Alternative.me provides an [API for fetching the Fear & Greed Index](#) of the crypto market at different timestamps.



## b) FinBERT

The FinBERT model, already pre-trained on financial text, can be imported from HuggingFace, which houses models and tools for NLP in its Python library called Transformers. (Usually, the application of FinBERT is prefaced with webscraping of comments or posts online or from social media, such as Twitter/Reddit, via tools like Pushshift API or PRAW, but for demonstration purposes and the scope of this project, the [Financial PhraseBank dataset collected on Kaggle](#) is directly used.)

```
"According to Gran , the company has no plans to move all production to Russia , although that is where the company is growing .": ('neutral',
'88.57% Confident'),
'Technopolis plans to develop in stages an area of no less than 100,000 square meters in order to host companies working in computer technologies and telecommunications , the statement said .': ('neutral',
'53.38% Confident').
```

The example code is adopted from a contributor on [Kaggle](#).

## Summary & Personal Insights

This project highlights the diversity of platforms and methods used to derive financial sentiment, ranging from simple retail-driven metrics (e.g., StockTwits) to highly specialized models (e.g., FinBERT). While all tools capture market sentiments in various capacities, not all provide transparent methodologies or publish backtested results.

Among these, **FinBERT** stands out as the **most statistically robust NLP-based sentiment model**. It is the superior choice of model as it most precisely interprets financial language, thanks to its domain-specific pretraining. Its reported accuracy (0.86) and F1 score (0.84) in classification tasks on the Financial PhraseBank and FiQA datasets, along with outperforming legacy models like LSTM and ULMFiT, suggest a high level of reliability and intuitive consistency. Thus, it can be trustingly used for texts scraped for all kinds of assets, especially stocks.

Another strong performer (though only for the crypto market) is the **Crypto Fear & Greed Index**, which, while more heuristic-driven, demonstrates solid alignment with retail market psychology. It **enhances pure DCA strategy**, allocating more during periods of extreme fear and less during greedy periods, by generating a 27% improvement in ROI. Even a single straightforward index, when grounded in behaviorally sound theory and enriched with online activity data (Twitter hashtags, search trends), can yield effective tactical signals.

From an implementation standpoint, the above two sources are the most readily accessible, with well-established frameworks and open APIs or models. Yet, the other sources are still helpful by offering insightful, comprehensive sentiment measures.

Meanwhile, **Sentdex** provides a sentiment scoring methodology based on adjectives and subject identification. Despite the API now being discontinued, historical backtesting (0.11 alpha and 1.96 Sharpe ratio) implies that the underlying approach had **predictive power**.

**StockTwits** is less rigorous due to its user-labeled structure but similarly shows **decent predictive ability** in aggregate. The 2020 backtest strategy using Bull-Bear sentiment ratio and EMA switches correctly captured market reversals like the COVID-19 crash. However, since it reflects crowd sentiment, it is reactive in nature and may suffer from echo chamber effects.

Less could be concluded about MarketBeat since the mechanism of its sentiment analysis relies on propriety aggregation, though given the breadth and detail of its other offerings, its sentiment data appears to be thorough.

Going forward, combining FinBERT's textual granularity with sentiment aggregators like StockTwits could potentially form a hybrid sentiment indicator with enhanced market timing power. Such an approach might capture both the depth of language nuance and the breadth of crowd behavior—offering a more complete view of market sentiment.

## References

[MarketBeat](#)

[Sentdex](#)

[Alternative.me Crypto Fear & Greed Index](#)

[StockTwits](#)

[Back testing Sentdex Sentiment Trading Signals — Harrison Kinsley](#)

[DCA strategy using the Fear & Greed Index — skogsraw](#)

[Mining StockTwits Retail Sentiments for Momentum Trading — Jay Lin](#)

[FinBERT: Financial Sentiment Analysis with Pre-trained Language Models — Dogu Araci](#)

[Sentiment Analysis for Financial News — Ankur Sinha](#)

[😊 Huggingface FinBERT for Sentiment Analysis — Maxwell](#)