

Tweets to Trades

Use Sentiment Analysis on Tweets to Predict Stock Price Movement

Section 26 Team 1



Lilo Hsiung



Xixi Zhong

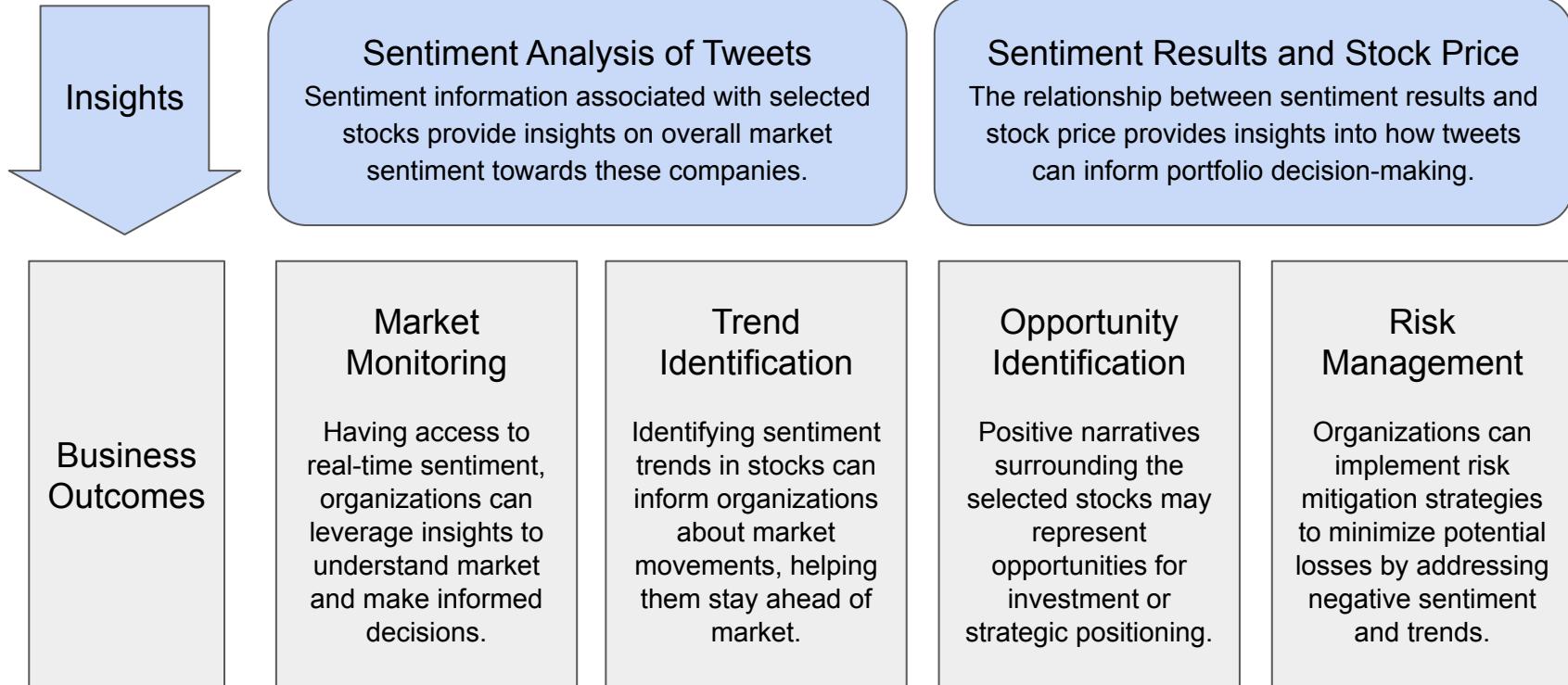


Bosan Hsu



Kuan Lin

The project aims to leverage sentiment analysis of tweets related to selected stocks to equips businesses with timely insights to spot trends, seize opportunities, and manage stock market risk.



Use Tweets and Stock Price Data to Run Prediction

Twitter-Financial News Sentiment

Source: Hugging Face

Features:

Date: 2015-2020

Tweet: comment to selected stock

Company: Apple, Tesla, Amazon...

Data Size: 110K tweets

Bias: Sample Bias, Noise, Temporal Misalignments, Fake News

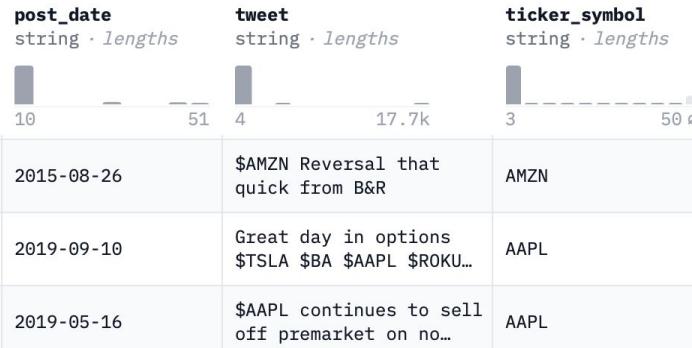
Historical Daily Stock Data

Source: yfinance (Yahoo Finance)

Features: Date, Stock Name, Close Price, Volume

AAPL

Date	Close	Volume
2019-04-17	50.782501	115627200
2019-04-18	50.965000	96783200
2019-04-22	51.132500	77758000
2019-04-23	51.869999	93292000
2019-04-24	51.790001	70162400
2019-04-25	51.320000	74172800
2019-04-26	51.075001	74596400
2019-04-29	51.152500	88818800



Dig into Tweets with Two Methods: Lexicon and LDA

Step 0

Data Normalization

Lexicon Method

Reasoning

Tweets have its own nature, so we adopt AFINN Sentiment Lexicon as it was originally created for microblogs.

Steps

- 1. Split the dataset:** Training data (80%) and Testing data (20%).
- 2. Get Sentiment Score:** Map the words to sentiment score to get the sentiment of the tweets.
- 3. Train and Test Data:** Train the data with sentiment score as independent variable and stock price as dependent variable to get the relation between tweets and stock price. Then put testing data into the formula generated by training data.

Evaluation

Regress the Stock Price rise-and-fall with Tweets sentiment and interpret the outcome.

LDA Method

Reasoning

LDA is ideal for unlabeled Tweets dataset as it identifies underlying topics and simplifies analysis by reducing text data dimensionality.

Steps

- 1. Model Construction:** Utilized LDA to discern the underlying thematic structures within our corpus of tweets data.
- 2. Model evaluation:** Log-likelihood, Perplexity and Coherence Scores.
- 3. New Document Topic Extraction:** (Optimism, Neutral, Pessimism)