**Influence of tweet sentiments on stock market prices**

Claim

|  |  |  |
| --- | --- | --- |
| **Twitter Sentiment** | **Correlated with** | **Stock Price trend** |

Claim Conditions

|  |  |
| --- | --- |
| Key | Value |
| Twitter sentiment analysis | VADER sentiment scoring is used to assess public opinion on a scale from -1 to 1. |
| Stock price trend | Compare the close stock price of the next day and the current day if it goes up then label it as 1, if it goes down then label it as 0. |

Figure (5 panels max, separate files, URLs only, one plot or diagram per panel)

1. Daily Sentiment Count for Tesla ([*https://ibb.co/XyzGfpk*](https://ibb.co/XyzGfpk))
2. F1 scores for all the timeseries split folds ([*https://ibb.co/TWhfXw7*](https://ibb.co/TWhfXw7))
3. ROC AUC scores for all the timeseries split folds ([*https://ibb.co/X3DT8HX*](https://ibb.co/X3DT8HX))
4. Feature importance ([*https://ibb.co/1qq7GyZ*](https://ibb.co/1qq7GyZ)*)*

Highlight

This study uses Twitter sentiment analysis to predict Tesla stock trend and explore whether sentiment has an impact on stock trend. We found that sentiment scores lagged by 7 days have significant predictive power on stock trends, indicating the importance of social media sentiment in stock market analysis.

Description (3,000 characters)

This research aimed to assess whether sentiment expressed on Twitter has a measurable impact on stock market prices. We analyzed sentiment data from tweets and combined it with financial factors to predict the stock price trend and investigated if public opinion shared on social media correlated with stock performance.

**Data Collection and feature selection**

We used datasets obtained from Kaggle, which provided both financial stock data and corresponding tweet data for companies of interest. We only chose to study the case on Tesla specifically due to its significant presence on social media and according to previous study, technology industries affected significantly [1]. The stock price data included daily closing prices from 30-09-2021 to 30-09-2022, while the Twitter dataset contained tweets relevant to the companies, tagged by keywords and hashtags.

Sentiment analysis on these tweets was performed using the VADER (Valence Aware Dictionary and sentiment Reasoner) model, which assigns a sentiment score between -1 (negative) and 1 (positive), and counts the number of positive and negative tweets each day.

To predict the stock price trend (based on whether it goes up or down compared to the next day’s closing price), we selected the following: price fluctuation, price gain, total valuation, average sentiment score, sentiment score with a 7- day lag, positive and negative tweet counts.

**Dataset split**

We used a time series split for training and testing our models to maintain the temporal order of the data, and we divided the dataset into four folds based on potential seasonal patterns in stock prices and trends.

**Model Selection**

1. **Logistic model**

We employed the Logistic Regression model for our analysis first, which is a suitable method for binary classification problems. However, based on our results, the model did not perform well. The mean F1 score was 0.35 and ROC AUC score was 0.41, indicating the model cannot capture the pattern in data.

1. **Random Forest Model:**

We also employed a Random Forest model to predict the stock trend and to assess feature importance. Random Forests are ensemble learning methods that rank features based on their contribution to the model's prediction accuracy.

The Random Forest model performed better than the logistic model, The F1 and ROC AUC score were better in the first three folds, especially for the third fold, which were 0.75 and 0.67 respectively. However, it performed worse in the last fold, with an F1 score of only 0.5. This could be caused by the change in market conditions, such as Musk’s twitter acquisition and the three-for-one stock split [2], which implies that it is important to consider all the macro and market dynamics in stock prediction.

**Results**

Based on the feature importance analysis using the Random Forest model, we found that the sentiment with a 7-day lag showed the most significant impact in predicting the stock trend. This provides insight for investors when incorporating social media data into stock analysis, especially for persistent sentiment within a week.

**Future research**

For future study, it would be interesting to incorporate more macro features into the model and apply event study methods to relate the trend with real-world events. Moreover, in the context of Tesla, it would be important to analyze the impact of Elon Musk’s tweets. Given that he has acquired Twitter (now rebranded as X), his influence on social media has significantly increased, and his personal statements could lead to substantial market reactions.

References

[1] Teti, E., Dallocchio, M., & Aniasi, A. (2019). The relationship between Twitter and stock prices: Evidence from the US technology industry. *Technological Forecasting and Social Change*, 149, 119747. <https://doi.org/10.1016/j.techfore.2019.119747>

**[2]** Tesla Announces Three-for-One Stock Split. (2022). Tesla Investor Relations. <https://ir.tesla.com/press-release/tesla-announces-three-one-stock-split>

Code

[**https://github.com/PRARTHANA-G01/Sentiment\_Stock\_Relation**](https://github.com/PRARTHANA-G01/Sentiment_Stock_Relation)

Data

<https://www.kaggle.com/datasets/equinxx/stock-tweets-for-sentiment-analysis-and-prediction/data>