Ozempic Sentiment and Stock Price Analysis

DS2500 | Laney Strange | April 15 2025

Rishona Shelby Israel (israel.r@northeastern.edu)

Kathleen Lautenbach (lautenbach.k@northeastern.edu)

Mahika Modi (modi.mah@northeastern.edu)

Erika Sohn (sohn.e@northeastern.edu)

Background

Novo Nordisk (NYSE: NVO), one of the world's most valuable companies, holds significant value due to its pharmaceutical products, particularly Ozempic, a diabetes and GLP-1 inhibitor. Given Novo Nordisk's role in global markets, we expect shifts in Ozempic sentiment and Novo stock to have widespread market implications. Recently, Ozempic has garnered public attention for its secondary effects as a weight-loss drug, leading to increased media coverage and shifting public perception. Our project aims to analyze two key factors: Ozempic sentiment on Reddit and changes in Novo Nordisk's stock price. Further, we aim to analyze the correlation between both factors using predictive modeling. We will analyze data from subreddits like r/Ozempic and r/stocks to capture broader sentiment. We chose Reddit because it provides more organic sentiment compared to platforms like Twitter or news outlets, with less influence from marketing.

Problem Statement

The sentiment surrounding Ozempic, especially within social media communities like Reddit, may provide valuable insights into public sentiment and its potential predictive relationship with

stock price movements. This study analyzes sentiment data from individual posts on subreddits, including r/Ozempic and r/stocks, to explore whether sentiment fluctuations influenced changes in Novo Nordisk's stock price in March. The efficient market hypothesis (EMH) suggests that stock prices reflect all available information, making it theoretically impossible to consistently outperform the market through stock selection or market timing. However, the weak form of EMH allows for the possibility that sentiment analysis of public opinion might capture information not yet fully incorporated into stock prices. Our research seeks to fill the gap in the research in understanding how shifts in the public's sentiment may influence broader market trends and reflect underlying economic behavior, potentially challenging or supporting aspects of market efficiency theory in the context of pharmaceutical stocks driven by consumer sentiment. This raises the question: If Ozempic sentiment is more positive, will Novo Nordisk's stock prices increase? Conversely, if sentiment is negative, will stock prices decrease? What is the correlation between public sentiment and stock prices?

Hypotheses

Set 1:	Set 2:
H0: There is no significant correlation between Ozempic sentiment on Reddit and investor sentiment on r/Stock regarding Novo Nordisk.	*H0:* There is no significant relationship between sentiment on Reddit and future movements in Novo Nordisk’s stock price. *H1:* Positive sentiment on Reddit is associated with increases in Novo Nordisk’s stock price,

H1: A significant correlation exists between Ozempic sentiment on Reddit and investor sentiment on r/Stock regarding Novo Nordisk.	while negative sentiment is linked to decreases.

Data Sources

Our analysis relies on two primary data sources: the Reddit Posts gathered using the PRAW API and manually downloaded daily stock price data from Bloomberg. Using the PRAW API, we collected posts from the subreddits r/Ozempic, r/stocks, r/Diabetes, r/biotech, r/biotech_stocks, wallstreetbets, and r/WeightLossAdvice, focusing on sentiment related to Ozempic and Novo Nordisk. To filter relevant posts, we used the keywords "Ozempic," "Novo Nordisk," "NVO", "weight loss," "side effects," and "skinny." These posts were then analyzed for sentiment using the VADER sentiment analysis tool. The second data source was a file of daily closing stock prices for Novo Nordisk, obtained from Bloomberg via the HP (historical prices) function and converted from an Excel to a CSV file for ease of analysis.

Regarding ethical considerations, Reddit is a public platform, and the data was collected without violating user privacy. However, to protect anonymity, we excluded personal identifiers such as usernames. We also acknowledged potential biases in the data, as Reddit discussions reflect the opinions of a self-selecting user group, which may not represent the general population's demographics (i.e, race, gender, and age) or their views. Despite the potential biases, the data still provides relevant and meaningful information for gauging general sentiment in the context of our analysis.

Data Science Approaches

Our data analysis explores the relationship between user sentiment and stock price movement. We performed sentiment analysis as a form of unsupervised learning on consumer and investor opinions in the subreddits using VADER, an NLP model designed for analyzing sentiment in social media text. We found VADER to work well with our PRAW because it is effective with short and informal content like Reddit posts, further capturing nuances such as emoticons, slang, and negations.

Reddit posts were collected in dictionary form, where the key was the subreddit title and the value was a dataframe of all posts for selected keywords in March. The function was limited to 50 posts per keyword to ensure successful execution with a reasonable runtime. To accurately align subreddit sentiment with Novo Nordisk's stock closing prices, post dates were adjusted. Posts made after market close were attributed to the following day's sentiment, and those made after Friday at 5 p.m. EST, as well as on Saturday and Sunday, were included in the following Monday's sentiment score. This alignment allowed for the proper analysis of how Reddit sentiment influenced sequential stock behavior. The dictionary was merged into a single dataframe for calculating overall daily sentiment averages, while subreddit-specific data was kept separate for heat map analysis. Each response on our platform was assigned a sentiment score based on positive and negative words, weighted by the number of upvotes it received, reflecting how strongly a sentiment was endorsed by the community.

To analyze the relationship between sentiment and stock performance, we conducted a Time-Series analysis. This method tracks sentiment changes over time and compares them with Novo Nordisk's closing stock prices. Time-series analysis is effective in identifying trends over

sequential data points, making it suitable for correlating stock price movements with sentiment fluctuations.

Additionally, we used Linear Regression as a supervised learning method to model the impact of Reddit sentiment on stock prices. Linear regression helps quantify the relationship between sentiment scores (independent variable) and stock price (dependent variable). By calculating the R^2 value, we assess the strength of this relationship. Thus allowing us to interpret the direction and magnitude of sentiment's effect on stock price movements.

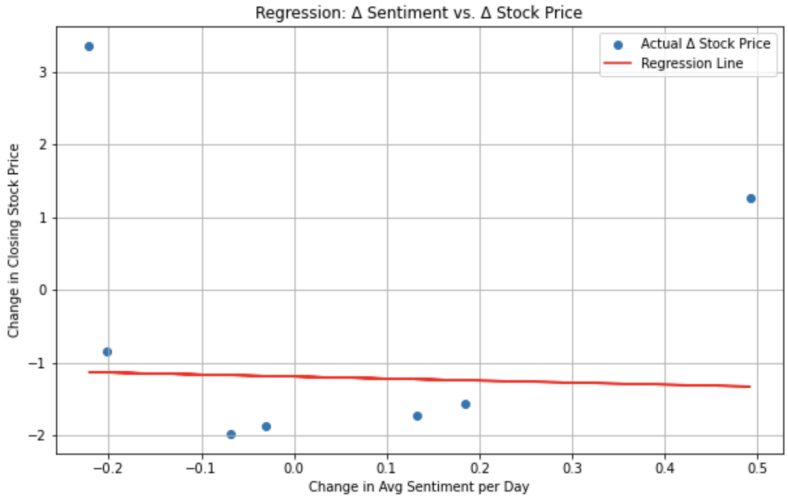
Results and Conclusions

Our analysis finds no significant relationship between Ozempic sentiment on Reddit and stock price ($r = -0.0138$, $p = 0.9526$). The low negative r value challenges our hypothesis, and the regression model shows randomly scattered points, further confirming no significant correlation between sentiment and stock price.

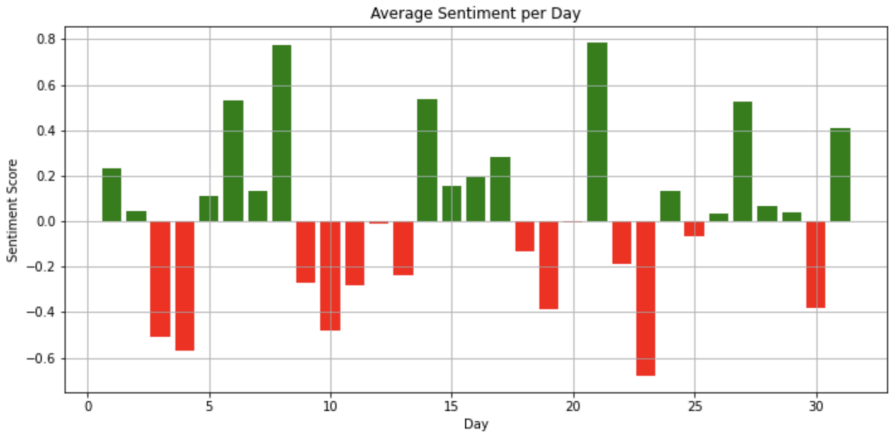
Initially, we expected to find a positive correlation between Ozempic sentiment and stock price, assuming that increased positive sentiment would drive stock price growth. However, the analysis revealed no significant relationship, suggesting that factors beyond online sentiment likely play a more critical role in stock price movements.

In evaluating our regression model, we calculated an R^2 score of -0.1438 and an MSE score of 4.0581. The negative R^2 indicates a poor predictive fit, highlighting the weak relationship between sentiment and stock price. The high MSE further underscores the inaccuracy of our predictions, reflecting significant deviation between predicted and actual stock prices. The poor

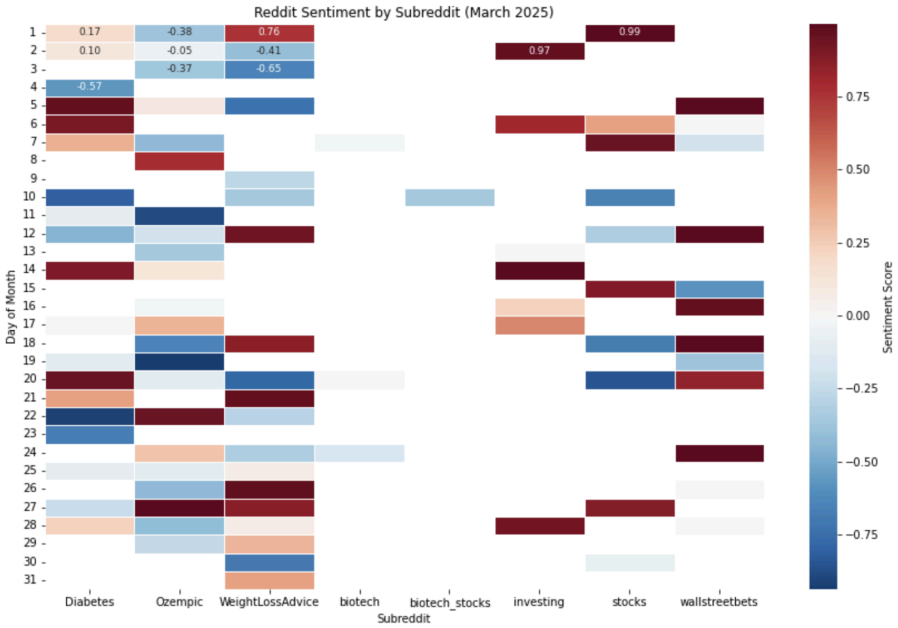
fit is likely due to external factors beyond sentiment, such as macroeconomic trends, company R&D, and CSR policies, which significantly influence stock price movement.



The stock price line graph and sentiment bar plot below show the movement of average sentiment and Novo Nordisk's closing stock price over time. When compared, there is no visible relationship between the two variables, but both graphs show significant spikes and drops. This volatility may indicate emerging concerns within Novo Nordisk and offer valuable insights, such as areas for improvement in product development.



The heatmap reveals a wide range of sentiment across subreddits, with red indicating positive sentiment, blue showing negative sentiment, and white/grey representing neutral sentiment. Health-focused subreddits, like r/Ozempic and r/WeightLossAdvice, tend to show more negative sentiment, while finance-related subreddits, such as r/investing and r/stocks, display more positive sentiment. This polarization highlights how sentiment varies across communities and provides insights into how different populations or industries are affected by Ozempic, helping identify the most relevant subreddits for meaningful analysis.



Future Work

Given that Reddit sentiment data isn't a reliable predictor of Novo Nordisk's stock price movements, future work could address different predictors, such as macroeconomic factors, earnings call transcripts, company event memos, analyst reports, etc. Dates for sentiment analysis could strategically be picked to include days when rate cuts/increases are announced, quarterly reports are released, or other company/industry-specific announcements. This would allow us to explore sentiment relating to the companies' potential future performance. Future work could also include expanding the date range, increasing the number of Reddit forums used, using other online posting platforms (Twitter and/or LinkedIn), and also increasing the number of posts used for a more accurate sentiment analysis. To add another layer of nuance, we could analyze sentiment correlation patterns during different market conditions (bull vs. bear markets or even high vs. low volatility periods) to identify conditional relationships. Implementing more advanced natural language processing techniques beyond VADER, such as BERT or RoBERTa, could also better capture nuanced sentiment expressions specific to financial and medical contexts.