**NLP 2021 - Final Project Proposal**

**IDC MSc**

**Date:** July 5, 2021

**Authors**: Ofer Gross (ID 304997067), Yuval Katz (ID 311132468), Ofir Nesher (ID 204502926)

**Title: Does social media affect the stock markets?**

**Introduction:**

A meme stock is any stock that's seen excessive trading volume from retail investors who've targeted it on social media. In other words, this stock has “gone viral” on social media and has seen its price skyrocket as a result [[1]](https://ca.finance.yahoo.com/news/meme-stocks-rich-170043638.html).

During the last few months, there’s a hype around meme-stocks (like GameStop and AMC) in the stock market inspired by Reddit forums and Twitter [[2]](https://www.forbes.com/sites/jonathanponciano/2021/05/25/amc-and-gamestop-skyrocket-as-reddit-traders-rally-around-meme-stocks-anew/).

Stocks are great for analysis because we can truly see whether the prediction is in fact a great indicator of the market movement [[3]](https://link.springer.com/article/10.1007/s12559-021-09819-8).

**Goals**:

The project will focus on **Sentiment Analysis** of tweets about a few stock symbols (tickers) in order to predict the trend around them and suggest whether to buy or sell them.

Our process is as follows: we scrape top-rated comments from top-rated posts on Twitter and Reddit and analyze the content for a few stock ticker mentions. We then run the comment through a NN, and the sentiment analysis is performed. The stock sentiment gets a positive value from 1 to 100, or a negative value from -1 to -100 (by scaling the hyperbolic tangent, tanh, function, for example).

To perform this, the NN can use a "bag of words" pre-processing methods (like the N-gram model for example), meaning that word order and sentence structure are lost.

**General approach:**

1. Using Twitter/Reddit-’s API to fetch the data about the relevant tickers we’re interested in
2. Pre-process all data collected
3. Running a NN on our data
4. Visualize sentiments per ticker
5. Suggest a buy/sell signal for each ticker

**Resources:**

* Twitter/Reddit (posts and API)
* Tweepy - An easy-to-use Python library for accessing the Twitter API [[4]](https://www.tweepy.org/)
* Yahoo Finance (stock prices)
* Our own code and lecture slides from the semester (from the NLP course and the homework assignments)

**Final work to be submitted by August 20th, 2021**

<https://docs.google.com/presentation/d/17OE2gWufxBhoUBeRArkEmX1p4K_S8QOpNT5qOaJAApU/edit#slide=id.p> - Kfir’s slides