CS 410 Final Project Proposal October 2021

Topic: Stock Sentiment Analysis Using Twitter Feed

Team: Porcupine Tree

Team Captain: Zachary Oldham

Zachary Odham	NetID: zoldham2
Yogeswara Rao Lekkalapudi	NetID: yrl3
Luis Mariano Ovalle Castaneda	NetID: lo22

1. **Introduction**

For the final project, we choose the topic of sentiment analysis using the Twitter feed. We will be using a Twitter API to get the documents related to a specific stock selected by the user and then using Text Access and Mining techniques learned during the course to analyze the social sentiment of the stock. This system will be beneficial for those wanting to understand the sentiment surrounding a stock prior to investing.

2. **Overview**

Our plan is to build a system that takes as input the name of a stock. It will then build a query, likely by augmenting the stock name with the name of the company associated with that stock, though this approach may change. We will then apply the techniques that we have learned to build and optimize the query vector. Once we have a query vector, we will rank all tweets within a time window by relevance to the query vector, and then select the top k tweets. We will perform sentiment analysis on each of these tweets in order to obtain probabilities that each tweet is positive or negative. We will then have two lists, one of positive tweets and one of negative tweets. We will reorder them based on their ranking with the query vector, and some heuristic based on the popularity of the tweet. We will then present the user with the two lists of tweets with the goal of giving them an idea of how the public feels about the stock. We will also compute some heuristic sentiment score that attempt to describe the general sentiment surrounding the stock and present this to the user.

3. **Motivation**

Social media provides a dynamic and massive amount of information that is difficult to interpret by the user. This project aims to harness this data set to give helpful information to the user for their research and help make investing decisions. Our software should quantify a specific stock user's social sentiment using three different categories; Positive, Neutral, or Negative.

4. **Work To Be Done**

We will need to figure out how to work with the Twitter API, we need to do some investigation into what tweets we want to retrieve, we need to implement a query vector improvement system as well a ranking algorithm, we need to implement sentiment analysis which may use either the NLTK or a recurrent neural network, we need to implement a heuristic

ranking algorithm, and we need to implement a heuristic sentiment summary. Across the three of us, this will likely take at least 20 hours per person.

5. **Evaluation**

We can compare our sentiment analysis against the Fidelity tool that provides a sentiment score around a stock. We can also compare the judgement against a manually generated judgement that we create by manually looking through twitter. We may also attempt to see if there is any correlation between our analysis and the performance of a stock.

6. **Tools**

- Python 3.5
- MeTA Toolkit
- Tweepy, Twitter API v1.1
- NLTK Natural Language Toolkit.
- Tensorflow