Automatic Media Processing - Algorithmic Market Predictor (2-AMP) Project Proposal

Brody Curry
Department of Electrical Engineering and Computer Science
University of Tennessee, Knoxville
Knoxville, Tennessee
bcurry8@vols.utk.edu

George Evans
Department of Electrical Engineering and Computer Science
University of Tennessee, Knoxville
Knoxville, Tennessee
gevans16@vols.utk.edu

Ryan Perry
Department of Electrical Engineering and Computer Science
University of Tennessee, Knoxville
Knoxville, Tennessee
rperry21@vols.utk.edu

Nicholas Dawson
Department of Electrical Engineering and Computer Science
University of Tennessee, Knoxville
Knoxville, Tennessee
ndawson2@vols.utk.edu

Tanishq Somani
Department of Electrical Engineering and Computer Science
University of Tennessee, Knoxville
Knoxville, Tennessee
tsomani@vols.utk.edu

***Abstract*— This paper discusses the development of a proposed program that aims to improve automated trading accuracy. This program would find, scrape, and analyze relevant news articles to predict movements in the stock market better than traditional historical market analysis. This paper outlines the division of labor, timeline, and deliverables for this project and serves as a guide for group work going forward.**

Keywords— Finance, Data Science, LLM, Automated Trading

I. Objective

The US stock market is a massive source of data, and whoever can best predict future movements of the market has an edge above the competition. While many existing automated trading systems have already utilized AI/ML to trade based on historical stock performance, this method fails to consider real-world events; these systems can only predict stock movements once external stimuli is already affecting the market. However, it may be possible to predict the impact of real-world events before they impact the stock market by gathering newly published news articles and processing their contents to determine potential market impact. With this extra data, our system would theoretically have greater trading efficiency compared to traditional historical market analysis. This proposed system would scrape relevant news articles from various sources, determine their relevancy on the target stock/industry, generate a holistic sentiment score that represents the news' predicted impact on the stock, and make real-time trading decisions based on this news data and existing historical market analysis.

II. Motivation

Stock prediction is a highly sought after market, and has proven to be extremely difficult, even with the use of AI and other machine learning methods. There is great monetary and educational benefit to successfully predicting how a stock will behave given current trends. While not every stock has a plethora of available articles and updates, we want to evaluate the effectiveness of our approach by selecting a limited scope of stocks to start.

For instance, the Nvidia stock is one of the most popular on the US stock market, and due to its popularity it receives constant updates and articles from a variety of sources. In regards to some other possible options, the Nvidia stock is also relatively stable, which makes it enticing for a proof of concept experiment such as the one proposed.

III. Data

The data in mind comes from news articles released by various media outlets and sites that create content pertaining to recent tech events. Such sources include but are not limited to Forbes, IGN, and Wired. The project will take these articles and scrape and study the text and subject within them, to determine how they relate to stocks. Data will likely have popular stock metrics to determine if the stock will be positively or negatively affected. The project will also utilize past news releases and historical market data to form a training dataset for our AI predictor. Many datasets containing historical market data sourced from Yahoo Finance and official Nvidia reports are publicly available online from sources like Kaggle.

IV. Responsibilities

Ryan Perry will be handling the scraping of the sites, planning to use RSS and similar tools. Nic Dawson will be creating the article finder and ensuring ways to find its relevancy to the chosen stock. George Evans will be handling LLM and general troubleshooting in regards to it. Brody Curry will be handling general text writeup and notes alongside supporting the LLM development, as well as general documentation. Tanishq Somani will help in finding suitable articles to run through and test the LLM as well as helping in the scraping of sites. All members of the team will be expected to keep in contact and discuss any and all adjustments and discoveries they make during the process of research and actual work. Furthermore, all of us will be

responsible for learning and improving our skillsets via this project. All team members will do their best to help with general documentation and deliverables for each stage of the project and planning process.

V. Timeline

The first three weeks will largely be put toward researching and understanding the process of analyzing data and LLM integration. During this research period, team members will evaluate prospective LLM models and databases for stock training data, as well as discuss how to create our AI and use existing LLM's to make our own specialized version. Furthermore, this time will be spent developing internal standards for data storage and labelling in order to streamline linking our finder, scraper, analyzer, and predictor systems.

Weeks 4-6 will begin building each aspect of the LLM, as well as the analyzer. In building the LLM, we will use our chosen model and begin training it using data from our stock data source and article source, with hopes that the model will be able to provide accurate predictions by the end of the sprint. The final weeks will be spent troubleshooting and refining any issues or problems.

VI. Expected Outcome

To keep the scope of the project feasible for the class and timeframe, our efforts will go to scraping different news sites for usable data, determining how relevant the article is to different stocks or industries and generating a sentiment score based on the content within the article. The desired outcome would likely consist of a functional run from scraping a news site for relevant articles, retrieving information from those articles, calculating a sentiment score from the text, and generating a stock prediction based on the output for at least one stock and news source.

Ideally, this work would be a part of a larger project to create a fully automated stock trading system that can predict and trade without human input. This larger system would be able not only to trade on Nvidia, but would be generalized to other stocks and industries.