***Project Proposal***

Predicting Short-term Stock Price Movement using WallStreetBets Reddit Data

***Group Members***

* Xizheng Li: responsible for collecting and organizing the Reddit data and helping explore the potentially good algorithm/models.
* Xueli Cao: in charge of the financial data collection and processing and trying out different algorithms/models that we can leverage.
* Haoning Gong: focused on developing machine learning models and conducting data analysis.

***Data Source***

The primary data source for this project will be WallStreetBets Reddit data, which will be collected using the Reddit API. In addition, we will also collect historical equity price, options volume, short interest, institutional ownership, and insider activity data from various sources such as Nasdaq data, Bloomberg, SEC filings, and traditional technological indicators used in industry

***Project Plan***

1. *Explain what you intend to study with your project.*

Our main objective is to develop a predictive model to forecast short-term US equity prices. Specifically, we aim to use this project as an opportunity to study and apply the following techniques:

* Natural language processing (NLP) for analyzing the tone and sentiment of social media posts related to individual stocks.
* Real-time web data collection methods to ensure we have up-to-date information for analysis.
* Time-series data modelling to account for historical stock price trends and market conditions.
* Apply Deep Learning / Machine Learning models to predict stock price movement

By applying these techniques, we hope to gain insights into the predictive power of social media data in the stock market and uncover new trading opportunities.

1. *What is the ultimate objective?*

Our ultimate objective is to develop a predictive model to accurately forecast short-term US stock price movements using social media and financial data. We hope to identify trends and predict short-term stock price movements not captured by traditional financial models.

1. *What types of models are you considering?*

We plan to develop several models, including:

* Sentiment analysis models that analyze the tone and sentiment of Reddit posts related to individual stocks.
* Predictive models that use natural language processing and machine learning algorithms to analyze the data and predict short-term stock price movements.
* Explore how the Deep Learning models help to predict stock price movement as opposed to other machine learning baseline models.
* Time-series models (ARMA, GARCH) that account for historical stock price trends and market conditions.

***Additional Questions:***

1. *Why is this project interesting?*

This project is interesting because it combines cutting-edge machine learning techniques with social media data to potentially generate alpha in the stock market. By analyzing sentiment and discussions found in WallStreetBets Reddit threads, we may be able to identify trends and predict short-term stock price movements not captured by traditional financial models. For example, the recent GME/AMC short squeeze led by WallStreetBets retail investors highlighted the potential impact of social media on the stock market. This project has the potential to discover new trading opportunities by uncovering market sentiment and trends in real time.

1. *What challenges and obstacles might you anticipate with this project?*

One major challenge will be filtering out useless posts in WallStreetBets, as not all posts are relevant to our analysis. Additionally, applying natural language processing techniques to unstructured social media data can be complex and require significant computational resources. We will need to carefully consider the accuracy and interpretability of our models and ensure that our methods are transparent and compliant with relevant regulations. In addition, we don’t really know the predicting power of social media text data after we include other traditional predictors. However, this is how research is conducting and why it’s interesting and exciting.