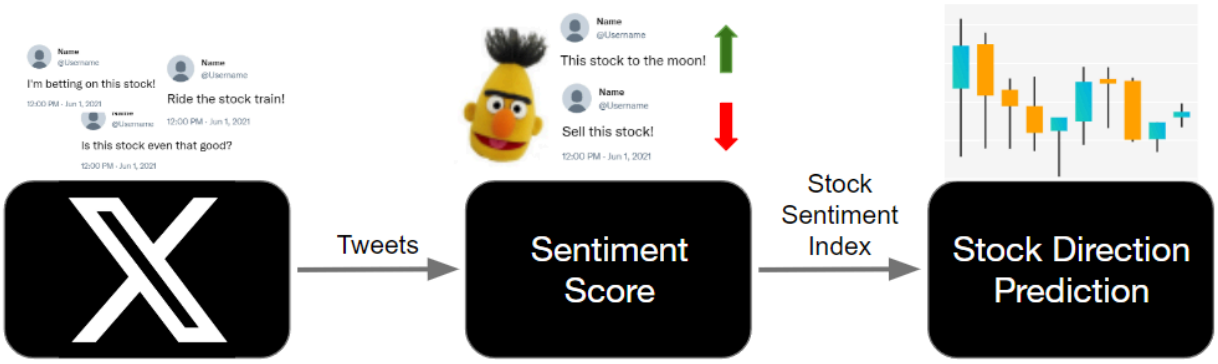
The team worked with the CTC in order to investigate the feasibility of integrating social media sentiment in trading strategies. To address this question, the team worked on creating a proof of concept prediction pipeline, as illustrated below.

Figure 1: Concept for prediction pipeline



First, the team researched methods of sourcing social media (X/Twitter) data. The data was obtained from both curated datasets used in the training of other LLMs, as well as tweets pulled in real-time using the official X API and an unofficial API called ReTTiwt.[1]

Second, the team used and fine-tuned LLMs to classify the sentiment of stock-contextual tweets. The team found success using a model called TwHIN-BERT[2] - a model initially trained on a corpus of tweets - which after fine-tuning with sourced X data improved accuracy from roughly 80% to 90%.

Lastly, the team explored ways to connect sentiment to stock returns. The team experimented with creating sentiment indices as an hourly exponential moving average of tweet sentiments. The team constructed autoregressive models as a baseline to predict next hour returns. Overall, these models showed great promise and feasibility in incorporating sentiment into trading strategies.

[1] *Rishikant181, Rettiwt-API, https://github.com/Rishikant181/Rettiwt-API*

[2] *Zhang et. al.. (2023). TwHIN-BERT: A Socially-Enriched Pre-trained Language Model for Multilingual Tweet Representations at Twitter. In KDD 2023 - Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 5597-5607). Association for Computing Machinery. https://doi.org/10.1145/3580305.3599921*