Tweet-Informed Prediction Market Forecasting

Vijay Sundarapandiyan vsundar1@umd.edu
Mazin Karjikar mkarjika@umd.edu
Omkar Pathak omkarp07@umd.edu
Aditya Menachery amenache@umd.edu

1 Introduction

Prediction markets represent a structured mechanism for aggregating dispersed information, where prices reflect the collective expectations of participants regarding future events. These markets have long been studied as a practical manifestation of the "wisdom of the crowd," offering probabilistic forecasts that often rival expert opinion. However, the forces that drive short-term directional changes within prediction markets remain less well understood.

This project seeks to explore whether public sentiment signals, particularly from the social media platform Twitter, can serve as a leading indicator for such shifts. The central hypothesis is that the real-time discourse, speculation, and narratives circulating within online communities not only mirror collective expectations but may also foreshadow adjustments in market consensus. By systematically collecting and analyzing tweets related to specific prediction market contracts, whether tied to political outcomes, economic indicators, or broader social events, we aim to quantify sentiment dynamics and examine their relationship to subsequent price movements.

At the core of our methodology is a large language model (LLM), which enables richer interpretation of unstructured text than traditional sentiment classifiers. The LLM is employed to classify sentiment and stance in relation to specific prediction market contracts, to extract themes and narratives by clustering semantically similar discussions, and to identify directional bias by mapping the polarity of discourse to potential upward or downward shifts in market pricing. These outputs are transformed into time-series sentiment indicators, which are then statistically correlated with directional changes in prediction market prices. By combining LLM-driven text understanding with econometric analysis and short-horizon predictive modeling, we aim to evaluate whether spikes in sentiment intensity, shifts in narrative focus, or sustained discourse trends anticipate subsequent market movements.

By bridging the fields of computational social science, NLP, and market forecasting, this project addresses a fundamental question: can the crowd's own chatter—interpreted through an LLM—enhance our understanding of the crowd's aggregated bets? Ultimately, the goal is to assess whether incorporating sentiment-driven insights into prediction market models can improve the accuracy of short-term directional forecasting, offering a richer picture of how information, opinion, and belief flow between unstructured online discourse and structured market outcomes.

2 Related work

Several studies have been conducted on prediction market microstructure, sentiment analysis of social media data to predict asset price moves, large language models to understand financial signals, and mapping social media news to prediction market price changes. The seminal paper in prediction market microstructure is "Accuracy and Forecast Standard Error or Prediction Markets," which shows that prediction markets are generally accurate markers (as opposed to polls) of market prices at both long and short horizons. Recent studies on newly developed prediction markets (Kalshi, Polymarket), primarily "Makers and Takers: The Economics of the Kalshi Prediction Market."

Next, there are several papers linking social media, specifically Twitter, to sentiment analysis that can accurately predict financial asset returns. A 2023 Federal Reserve Study also showed that overnight Twitter sentiment may be an accurate predictor of stock prices in the following day. Further, more recently, large language models have been used to extract financial signals based on economic and financial text. Further, language models have been used to classify communications distributed by central banks to forecast financial asset price moves. As there have not been any studies linking large language model use to price discovery in prediction market venues, this gives significant exigence to the project idea discussed in this proposal.

3 Your approach

1. Model Selection. We will select a large language model capable of sentiment classification and few-shot reasoning.

2. Data Collection. We will use TweePy and Twitter API to collect real-time tweets relevant to a specific event on a prediction market, either Kalshi or Polymarket.

3. Labeling and Synthetic Data Generation. We will manually label a subset of collected tweets with respect to each binary feature. Then we may generate synthetically labeled tweets, depending on the data quality.

4. Data Preprocessing. We will create a pipeline to clean the collected data of URLs, mentions, and extraneous characters. Emojis may be helpful in feature classification.

5. Model Finetuning. We will finetune the LLM on the labeled dataset and optionally the synthetically generated data.

6. Sentiment Analysis Evaluation. We will create a baseline using known feature classifications, and evaluate the LLM on generating those features correctly.

7. Feature Extraction and Engineering. We will use large language model to extract features from the tweet, for example emotion, confidence, logical fallacies, betting direction, and use machine learning to weight these features into the prediction for the short-term market shift direction.

8. Integration with Prediction Market Data. We will combine the extracted features and learned weights into a simple predictive model of short-term market shift.

9. Market Shift Prediction Evaluation. We will evaluate the accuracy of market shift predictions made by our model.

What baselines will you use? We will use a naive sentiment analysis tool to generate a single number and use that to directionally evaluate the market movement (single feature baselines).

Justify why your group has enough compute to carry out your project We can feasibly rent a GPU for around a dollar an hour through Vast.ai but because finetuning is cheap we can simply use Unsloth in a Google Colab.

3.1 Schedule

Data Acquisition and Preprocessing (Weeks 1–2)

- Collect tweets and related social media data using Tweepy or the Twitter API.

- Retrieve prediction market price data through APIs such as Polymarket.

- Clean, timestamp-align, and store datasets in a structured format for analysis.

Model Development and Fine-Tuning (Weeks 3–7)

- Fine-tune open-source large language models using Unsloth and LoRA for domain-specific sentiment and stance detection.

- Build model pipelines for semantic clustering, scoring, and narrative extraction using Python frameworks and libraries.

- Generate time-series sentiment indicators and integrate them with market data.

Evaluation and Error Analysis (Weeks 8–9)

- Perform quantitative analysis comparing sentiment indicators with market movements using correlation and statistical tests.

- Conduct error analysis to identify misclassifications or weak correlations.

- Refine fine-tuning parameters or data filtering based on findings.

Final Report, Visualization, and Presentation (Week 10)

- Summarize findings and visualizations of sentiment–market relationships.
- Prepare final report, presentation slides, and documentation of methodology and results.

4 Data

The primary text data for this project will consist of publicly available posts from Twitter that are directly or indirectly related to prediction market contracts. These include tweets discussing political outcomes, economic indicators, and major social events that prediction markets seek to forecast. Such discourse provides timely signals of speculation and opinion, making it well suited for testing whether sentiment dynamics can anticipate short-term changes in market pricing. To align these discussions with specific market instruments, we will also collect contract descriptions and event metadata from platforms such as Kalshi and Polymarket.

Data will be obtained through the Twitter API through Tweepy. This approach ensures access to both historical and real-time data streams. In addition, we will incorporate curated datasets of tweets from prior research on elections, economic releases, and other salient events. Prediction market data, including contract prices and event timelines, will be sourced directly from Kalshi and Polymarket APIs or public feeds, enabling temporal alignment of sentiment measures with observed market outcomes.

For model development, we will adopt a hybrid annotation strategy. A large language model will be employed to classify sentiment, detect stance, and identify thematic clusters in the collected tweets. To guide and calibrate this process, we will hand-annotate a small subset of tweets, not as a full benchmark, but as seed examples. These examples will be used to construct few-shot prompts for a powerful LLM, enabling the generation of additional synthetic labeled data. This augmentation approach allows us to capture a broader range of discourse patterns while limiting the manual annotation effort to a manageable scale.

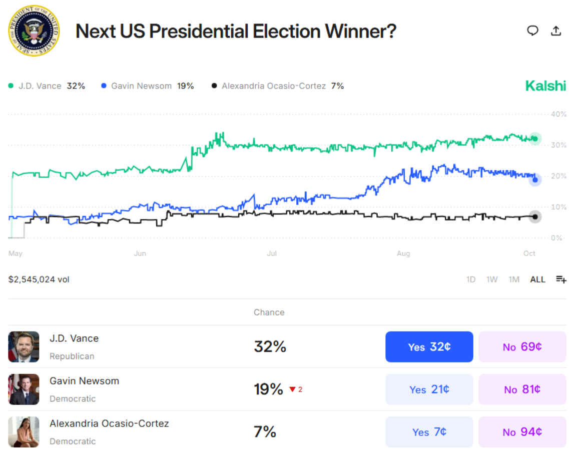

Figure 1: An example of contemporary prediction markets on Kalshi relating to the next US Presidential Election

The chosen data sources are both accessible and directly aligned with the research objectives. Twitter provides a high-volume, real-time environment for capturing shifts in sentiment and narrative framing, while Kalshi and Polymarket offer structured, transparent measures of collective expectations. Together, these datasets provide a rigorous foundation for evaluating whether online discourse serves as a leading indicator of short-horizon prediction market movements.

5 Tools

This project will leverage a combination of natural language processing and machine learning toolkits to implement the full data collection and analysis.

For data acquisition, we plan to use the Tweepy or snscrape libraries to collect tweets and scrape other online data related to the selected prediction market contracts. These tools will allow retrieval of real-time as well as historical posts with the relevant timestamps and user information. We will access prediction market data through APIs such as Polymarket's, using python libraries for cleaning. We will try using different open-source LLMs instead of settling on one now. Computational resources and needs will play a factor in this decision.

For language model adaptation, we will employ Unsloth and LoRA (Low-Rank Adaptation) for fine-tuning of large language models on domain-specific sentiment. This approach will allow the model to capture the linguistic, tone, and narrative structures in social media discussions about these prediction markets. We will perform the fine-tuning on a subset of tweets, enabling the model to classify sentiment polarity and market stance with higher contextual accuracy than general-purpose LLMs.

6 AI Disclosure

- Did you use any AI assistance to complete this proposal? If so, please also specify what AI you used.
  - Yes, ChatGPT

If you answered yes to the above question, please complete the following as well:

- **Free response:** Describe your overall experience with the AI. Did you use it to generate new text, generate research ideas, check your own ideas, or rewrite text? How helpful was it? Did it just directly give you a good output, or did you have to edit it? Was its output ever obviously wrong or irrelevant?
  - We outlined our proposal as a set of bullet points describing each part of the proposal in-detail. We then gave the AI (ChatGPT) this outline to put the bullet points together and create paragraphs from our work. The model was helpful in creating paragraphs, but we edited some parts to ensure accuracy and alignment with our original ideas. The model output was never obviously wrong or irrelevant, as we had given it ideas and prompted it to use only our ideas when putting the bullet points into paragraph form. In our prompts, we also told the model to refine our writing style and improve the grammar / flow of our writing from the original set of bullet points in the outline.

7 References

References in **figure 2**

Adams, T., Ajello, A., Silva, D., Vazquez-Grande, F. (2023). More than Words: Twitter Chatter and Financial Market Sentiment. arXiv preprint arXiv:2305.16164.

Berge, J., Nelson, F., Rietz, T. (2003). Accuracy and Forecast Standard Error of Prediction Markets. https://www.biz.uiowa.edu/faculty/trietz/papers/forecasting.pdf?utm_source=chatgpt.com%20citation

Bollen, J., Mao, H., Zeng, X. (2010). Twitter mood predicts the stock market. arXiv preprint arXiv:1010.3003.

Ng, Hunter and Peng, Lin and Tao, Yubo and Zhou, Dexin, Price Discovery and Trading in Prediction Markets (June 30, 2025). Available at SSRN: https://ssrn.com/abstract=5331995 or http://dx.doi.org/10.2139/ssrn.533199

Figure 2: References

References