

Words Don't Matter

A machine-learning study of 780,494 news-market pairs finds that a prediction market's *own price geometry*, not the content of the news, ranks which markets are about to move.

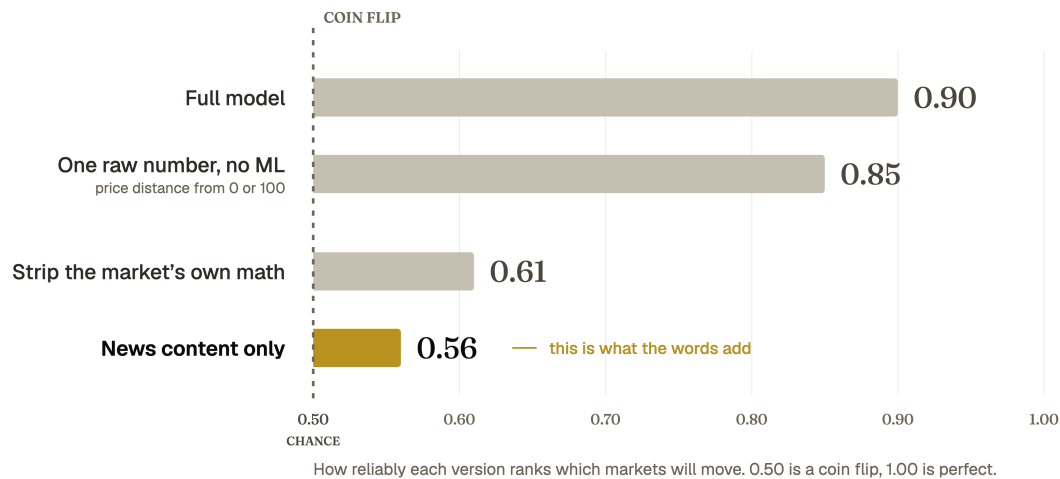


Figure 1 · Ranking power (AUC) by what the model is allowed to see. Higher is better; 0.50 is a coin flip. The words-only model barely clears chance, while a single number about the market's own price outranks everything the headline contains.

■ THE FINDING

Prediction markets are supposed to run on information. A headline breaks, traders read it, and the price adjusts. If that were the whole story, the words themselves would carry the signal, and a model that reads the news should see the move coming before the market finishes making it.

It does not. We gave a machine-learning model only what a headline contains: its sentiment, the entities and tickers it names, the type of source, and its length. Asked to rank which markets were about to move, that words-only model scored **0.56**, barely above the 0.50 you would get from a coin flip. Handed everything else we could measure, the same model reached **0.90**. Almost none of that lift came from the words.

■ THE METHOD

The study spans 780,494 news-market pairs over eight weeks, from April 29 to June 26. For every pair we asked one question: did the market's price move at least two cents within the next hour? The score we report is AUC, a measure of ranking, not accuracy, on a 0.50-to-1.00 scale. A model earns a high AUC by reliably sorting the markets that are about to move above the ones that will sit still, not by calling any single outcome right, and a serious reader should hold those two ideas apart.

To find where the 0.90 came from, we removed each input in turn and shuffled it across the eight weeks, keeping the feature's shape but destroying its meaning. Sentiment, entity counts, tickers, headline length, and source type each carried approximately zero weight; scrambling them changed nothing. The signal was somewhere else entirely.

*Prediction markets trade their **own geometry**, not the news.*

■ **MARKETS TRADE THEIR OWN GEOMETRY**

The signal was in the market's own math. Three numbers describe a prediction market without reference to any news: how far its price sits from 0 or 100, the price level itself, and the momentum it already carried. Strip those three from the full model and **0.90** collapses to **0.61**. Run the test the other way, and a single one of them, the distance of the price from 0 or 100, with no machine learning at all, scores **0.85**, more than every word-based feature combined.

That is the finding in a line: prediction markets trade their own geometry, not the news. A contract pinned near 0 or 100 has little room and little reason to move. One loitering in the middle with momentum behind it is already telling you what it is about to do. The headline that runs alongside is, for the most part, decoration.

■ **WHAT IT MEANS FOR A TRADER**

An entire industry is built on the opposite assumption. The market for alternative data, which sells traders non-traditional inputs including news feeds and social sentiment, was worth **\$14.2 billion** in 2025. This study puts a hard question to that spend: if sentiment, entities, tickers, and source type each rank movement no better than chance, what is the sentiment feed actually buying?

The part that predicts is already public. It is the market's own price, sitting in plain sight on the order book, free. Before paying for someone else's reading of the news, a prediction-market trader should first exhaust what the price is saying about itself. On this evidence, that is where nearly all of the signal lives.

THE TAKEAWAY

WORDS DON'T MATTER · SUMMARY OF ANALYSIS

- 01** A **words-only** model (sentiment, entities, tickers, source type, headline length) scored **0.56** AUC, barely above a 0.50 coin flip.
- 02** The **full model**, given everything, scored **0.90** AUC.
- 03** Removing the market's own math (price distance from 0 or 100, the price level, and momentum) dropped the full model from **0.90** to **0.61**.
- 04** One raw number, the price's distance from 0 or 100, with **no machine learning**, scored **0.85**, more than every word-based feature combined.
- 05** In leave-one-out testing across all eight weeks, sentiment, entity counts, tickers, headline length, and source type each carried **approximately zero weight**.

For scale, the case runs against a real market: the **alternative-data industry, which sells news and social-sentiment inputs to traders, was a \$14.2 billion market in 2025**. On this evidence, a prediction market's own price already holds the part that predicts.