Polymarket Trending News Politics Sports Crypto Log In Sign Up

$0 Vol.

Prediction Markets vs Machine Learning

Anna Demasure
Supervisor: Alessio Benavoli

Sports · NBA

Evidence from NBA Game Forecasting Across Pre-Game Horizons

FULL REPORT

Background:

Prediction Markets:

- Trade contracts on future events, where prices (0–1) reflect the market's implied probability.
- Traders are financially incentivised to incorporate new information into prices.

→A continuously updating, crowd-generated forecast.

Machine Learning Models:

- Models trained on historical and contextual data to learn patterns and generate predictions.

→A structured-data-based forecast

Key difference: Markets can incorporate news, judgement and rumours whereas ML is limited to the structured data provided.

Motivating question: Do markets add predictive value beyond ML model predictions?

Why the NBA?

Rich historical data → well-suited to ML training.
High market activity → meaningful prediction-market prices.
Binary win/loss outcomes → direct comparison between classification ML and market forecasts.
Fast-moving information → ideal for studying how forecasts evolve toward tip-off.

Data

Market Data - Polymarket
- 1,019 NBA games from 2025/26 season

ML training data
- 2018/19–2025/26 NBA data
- ML features: Team form, Efficiency, Injuries, and Lineups & rotations

Horizons: 5d, 3d, 2d, 1d, 12h, 6h, 3h, 1h.

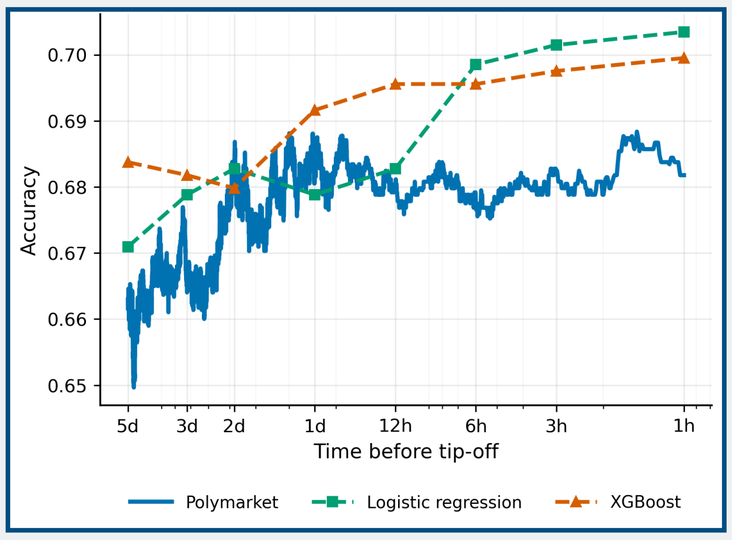
ML Models

Logistic Regression — Simple baseline model.

XGBoost — Flexible tree-based model capturing nonlinear relationships.

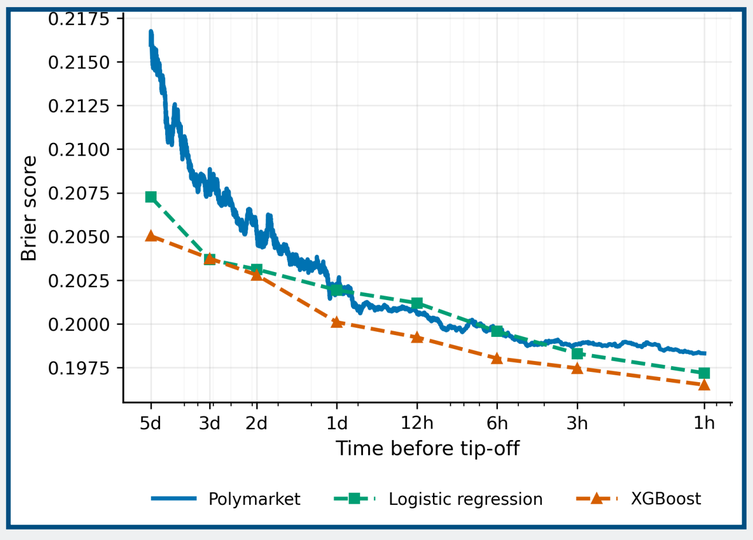
Key Results

Accuracy by Pre-Game Horizon



All methods improve toward tip-off.
Polymarket rises from ~**0.66 → 0.68**,
Logistic Regression exceeds **0.70 near 3 hours**.
XGBoost remains **consistently strong**.

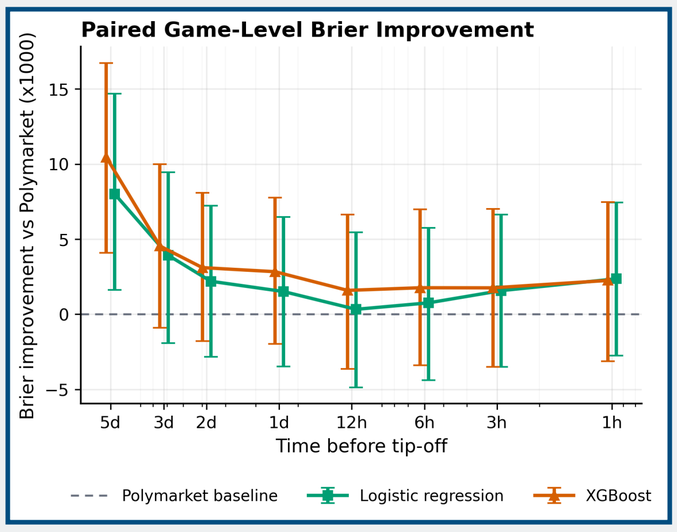
Brier Score by Pre-Game Horizon



(Lower Brier score = better forecast quality)

All methods outperform the **0.25 no-skill baseline.**
ML models hold an **early advantage.**
Polymarket narrows the gap near tip-off.

Statistical Testing



5d: ML advantage is statistically significant
~3d: Advantage begins to narrow
Near tip-off: No statistically clear difference

Specific Game Types

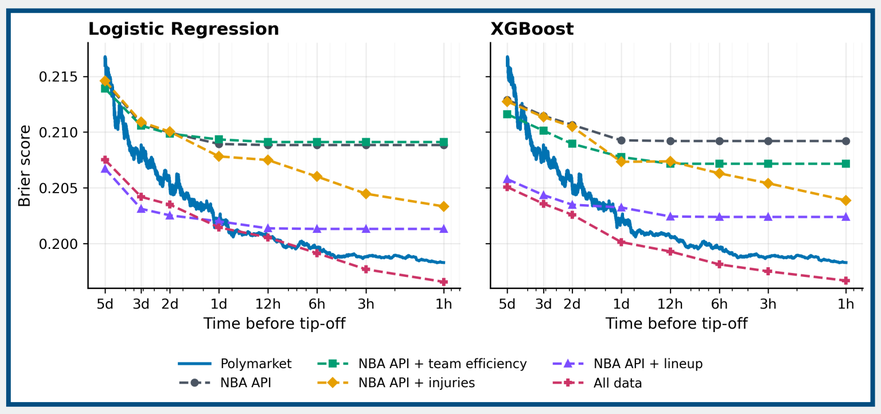
Forecasts compared across three game types: star-injury shocks, favourites that win, and favourites that unexpectedly lost.

Injury shocks: Markets react fastest to new injury information → suggests markets absorb informal injury news before it's officially reported.

Clear favourites: Polymarket is more confident than either model.
→ a bigger payoff when favourites win, a bigger cost when they don't.

Which Data Sources Matter Most?

Brier Score by Data Availability



Efficiency → Injuries → Lineups/Rotation
Full feature set: **~+2% accuracy** for both models in comparison to team form features only.

Discussion

- Neither machine learning nor prediction markets are consistently superior.
- ML leads early, but the gap narrows toward tip-off.
- Markets react quickly to informal information but may be overconfident.
- Information matters more than model complexity.
- Future models could combine market probabilities with ML features.
- Results may differ in sports with less data or lower market activity.

Key References:

Wolfers, J. & Zitzewitz, E. (2004). *Prediction Markets.* Journal of Economic Perspectives, 18(2), 107–126.
Page, L. & Clemen, R.T. (2013). *Do Prediction Markets Produce Well-Calibrated Probability Forecasts?* Economic Journal, 123, 491–513.
Thabtah, F., Zhang, L. & Abdelhamid, N. (2019). *NBA Game Result Prediction Using Feature Analysis and Machine Learning.* Annals of Data Science, 6.
Murphy, A.H. (1973). *A New Vector Partition of the Probability Score.* Journal of Applied Meteorology and Climatology, 12(4), 595–600.

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

LAIDLAW SCHOLARS

1H 6H 1D 1W 1M ALL