Research Reflective Report

Importance and Impact of the Research

My research project investigated whether continuously updating prediction markets can provide forecasting value beyond what a machine-learning model can extract from structured variables. Using NBA games as the main setting, I compared predictions across different pre-game horizons to see how forecasting performance changed as more information became available closer to the event.

The results did not show a statistically significant difference between the forecasting methods. While this was different from what we initially expected, it is still a useful finding. It suggests that prediction markets may not always provide additional value beyond structured data and also highlights the high level of uncertainty involved in predicting NBA games. However, these findings may depend on the type of event being predicted, so applying the same approach to other areas could produce different and useful insights.

The use of multiple forecasting horizons was also an important part of the project. It allowed us to show that forecasts generally became more informative closer to the event as more relevant data became available. The structured game-horizon dataset is another methodological contribution and could provide a useful framework for similar forecasting problems in other domains.

More broadly, the project demonstrates that prediction markets can be viewed as practical forecasting tools rather than only as platforms for betting or profit. They combine the judgement of many participants and respond quickly to new developments. In this study, Polymarket produced results comparable to the ML models, while requiring less effort from the end user to gather and process data. This suggests that prediction markets could offer a quick and accessible way to support decision-making in areas such as finance, elections, and public health

Personal Impact of the Research

This research helped me develop my technical skills significantly. I worked with large amounts of historical data from several different sources, transformed them into a consistent format and learned how to train and evaluate a Logistic Regression and XGBoost ML model. Throughout the process I learned that building a useful model is not only about choosing an algorithm but about checking what information is the most important to be included.

Working across many datasets improved my organisation and data-management skills. At the beginning, when I was still deciding which features to include, I created many versions of raw and transformed datasets. I quickly realised that without a clear system it would become difficult to know where each variable came from or which version was being used. I therefore became more disciplined about separating raw data from processed features, naming files consistently and tracking changes. This made the project more reproducible and reduced the risk of mistakes.

Alongside this, I also developed a stronger understanding of methodological rigour, particularly around bias and data leakage. Since I was comparing forecasts at different pre-game horizons, every prediction had to use only information that would genuinely have been available at that point. This meant thinking carefully about timestamps, injury updates, completed games and lineup information rather than simply using the most complete data available.

Collecting the data actually took a lot longer than initially expected. This meant I really had to work on my time-management skills. I started setting weekly goals for myself and treated my supervisor updates almost like personal deadlines, since I wanted to have something concrete to show rather than vague progress. I also kept a running checklist of smaller tasks, which made an overwhelming project feel more like a series of manageable steps.

Probably the biggest shift for me was learning to actually question my own results instead of just accepting them. I was still new to ML models, so I could not just trust an output because it looked reasonable. I remember being genuinely surprised early on that the ML models were performing at similar accuracy levels to Polymarket. I had assumed a market with so many traders would win easily, since it could pick up on information a structured model might completely miss. To check whether my results made sense, I let the model predict the outcome using just one feature: which team had the better season win percentage at that time. That alone gave an accuracy close to 67% five days before tip-off, surprisingly close to my full models. Having that simple benchmark gave me something to sanity-check against and made me more confident in the final results.

It also reminded me that sport is inherently uncertain. When I was surrounded by data, models and graphs every day, it became easy to forget that even the best dataset will never predict every game correctly. That uncertainty is part of what makes sport interesting to watch in the first place. This changed how I thought about forecasting. I stopped seeing the goal as trying to “predict the future perfectly” and started thinking more about whether the probabilities were sensible and well calibrated.

The project also taught me to be more realistic about research scope. Initially, I wanted to compare prediction markets and ML models across US elections, film awards, NBA games and US economic indicators. Retrieving market probabilities would be relatively straightforward, but I underestimated how difficult it would be to construct comparable ML datasets for each domain. Some data were limited, expensive or difficult to match consistently. Narrowing the project to NBA games was therefore an important decision. The NBA ultimately provided binary outcomes, abundant historical data and a fast-moving information environment that suited the research question well. Focusing on one domain allowed me to explain the methodology more clearly and conduct deeper analysis. I learned that completing one idea thoroughly is often more valuable than investigating several ideas superficially.

Writing the report itself was another challenge. I found it difficult to explain quantitative methods in enough detail for an unfamiliar reader while also staying within the word count. I dealt with this by keeping the main report focused and concise, while moving examples and supporting tables into the appendices. This prevented the report from becoming too overwhelming, while still giving the reader the chance to view more detail in the appendix. Overall, I feel happy to have improved my communication skills and have also realised that good research is not only about doing the analysis, but also about making it understandable to someone else.

Leadership Skills Gained

One of the main leadership skills I developed was taking ownership of a complex and open-ended project. There was no single dataset or predefined method that could answer my research question, so I had to make decisions about which data to collect, how to represent difficult information such as injuries, which models and evaluation measures to use, and how to respond when parts of the methodology did not work as expected. Managing these different components strengthened my ability to break a large problem into smaller tasks, prioritise what was most important and continue making progress despite uncertainty.

Future Path

This research has shown me the range of possibilities in working with data. Forecasting is not unique to sports. It appears in finance, politics, economics, public health and many other fields. Seeing how many questions can be explored through data has made me want to analyse more of them and has opened up the possibility of pursuing a career as a data analyst in the future.

Working with a supervisor on an open-ended research question also gave me much more confidence for tackling my fourth-year thesis and greater comfort with academic research in general. I can see myself extending some of the ideas from this project in a more mathematical direction, for instance investigating whether prediction-market price paths satisfy the martingale property implied by market efficiency.