

# My findings on using machine learning for sports betting: Do bookmakers always win?

A naive money-oriented idea?



Manuel Silverio · Aug 26 · 4 min read ★



Photo by [Kevin Ku](#) on [Unsplash](#)

One afternoon, in the middle of my holidays the thought of using machine learning to predict football results in the premier leagues came to my mind. I have never bet on sports myself because I do not like to dispose of the money, I make that way. However, I entertained the idea. I thought that if I can design an algorithm that gives me over 60% accuracy, I could spread the risk and bet on multiple matches, thus making constant revenue.

I thought the following, if I start with £100 (I live in the UK), and bet £10 pounds in 10 different matches, the odds are that I could win around 5 out of 10 matches. If I make

sure to have a reliable algorithm, I thought I could make an absolute fortune. Here is the important part, I have never placed a bet in my life, and I do not know the details and intricacies of the betting industry. I knew I had to do my research to understand better what was I missing.

I knew I was missing something; I could not be the first person to think of this, it is never that easy. I proceeded to do a small research to understand better what I could find around this topic. The results were very interesting as I found how things really work.

First, I found a couple of journal papers which allowed me to assemble a small literature review on this field. And yes, apparently, this is a whole research area in which professionals in the field of Artificial Intelligence dedicate their time and effort to improve their Machine Learning (ML) models. According to Bunker et al. (2019) Although several studies have considered statistical sports prediction, the use of the Neural Networks paradigm is a more recent approach.

In my research I found the percentage of accuracy of various algorithms for different sports:

Purucker (1996) achieved 61% predicting accuracy for results in the National Football League (NFL) using a Neural Network Model.

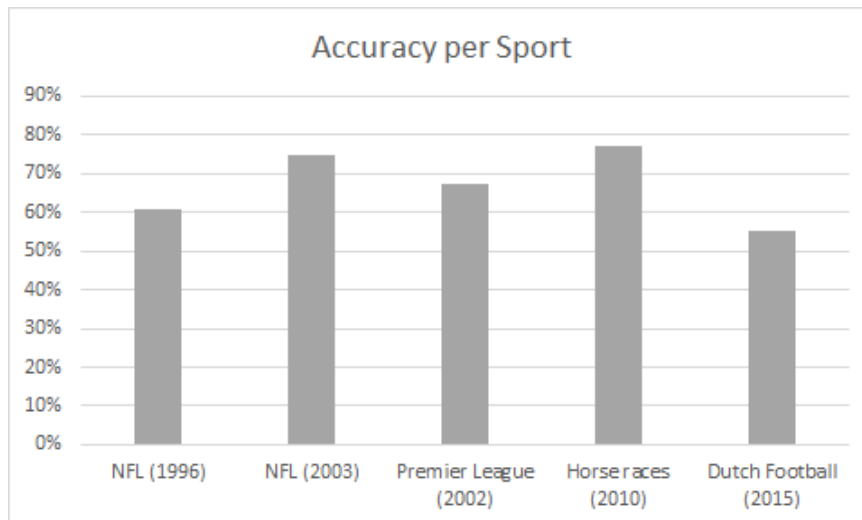
Kahn (2003) expanded the work of Purucker (1996), achieving 75% accuracy across the matches of week 14 and 15 of the NFL. For this data on 208 matches in the 2003 season were collected.

McCabe and Trevathan (2008) studied sports prediction in four distinct sports, namely, NFL (Rugby League), AFL (Australian Rules football), Super Rugby (Rugby Union), and English Premier League Football (EPL) using data back to the year 2002. The average performance of the NN algorithm was 67.5%.

Davoodi and Khanteymoori (2010) attempted to predict the results of horse races, using data from 100 races at the Aqueduct Race Track held in New York during January of 2010. Although the algorithm developed by Davoodi and Khanteymoori (2010) required exhaustive training time, it achieved 77% accuracy.

Tax and Joustra (2015) used data from Dutch Football competitions to predict the results of future matches. In this case the authors also considered the betting odds as variables for their Machine Learning models. While their models achieved an accuracy of 54.7%, the model which used only the betting odds achieved 55.3% of accuracy. This fact

made me realise something. Bookmakers have their own data science team. If the odds of a team winning are 10/1, then probably that team is going to lose.



Accuracy of ML predicting models per Sport — Adapted from literature review

After finding out that the accuracy of algorithms in previous research is around 50 to 70% (depending on the sport), I still thought that it would be feasible if the odds for the matches would be at least 2/1.

Finding out why no one has become a millionaire with this yet

Before I write the first line of code I was determined to find out if this was really feasible. At some point, I thought that maybe it was not legal to use your own algorithms, to which a simple Google search answered that it is allowed. Then I thought about bookmakers and how they regulate or limit the amount you can bet. I found a dissertation named **“Beating the bookies with their own numbers — and how the online sports betting market is rigged 2017”**.

This dissertation is where my research stopped. This paper explained how the authors attempted to use their algorithm to monetize and found two main barriers. First:

- ***bookmakers use their own machine learning algorithms to generate the odds of the match.***

Therefore, as your ML model points you towards the more certain results, you might always end up with a low benefit. Second, and even more important:

- ***Bookmakers discriminate against successful clients.***

Consequently, when you start to win often, bookmakers will start discriminating against you and restraint the amount of money you can bet.

Besides these two barriers, the paper I found, explained that over a 5 month period they made \$2,086 in 672 bets, with a return of 6.2%. That is good (It's not that good for the amount of effort you have to input), but it is hard work. You have to dedicate a lot of time and effort to make many bets and withstand being flagged by bookmakers.

My conclusions are that developing ML models for sports betting is good only for practice and improvement of your data science skills. You can upload the code you make to GitHub and improve your portfolio. However, I do not think it is something that you could do as part of your lifestyle in the long term. Because at the end bookmakers never lose. Ultimately I ended up not doing a single line of code in this project. I hope that my literature review helps illustrate others.