Project report for Sports Analytics

Predicting biathlon shooting outcomes and interactive modelling of results

Sports Analytics - 753A01

Nikolaus Stratil-Sauer - nikst363@student.liu.se

IDA - Department of Computer and Information Science
Linköping University

22-06-2023

Contents

1. Introduction

Biathlon is an olympic winter sport combining the disciplines of cross-country skiing and shooting. Athletes compete in multiple disciplines each with a different number of shootings and amount of distance skied. The best biathletes win races by skiing faster than the rest of the field and also having a high target hit rate. Since each miss during shooting results in a significant time loss, it is interesting to consider the factors that contribute to a successful shooting. While biathlon is not yet very relevant in the world of sports analytics and traditional methods prevail in the field, some research has been released trying to use statistical and machine learning methods to investigate biathlon races and outcomes. These range from analysis of existing data to trying to predict outcomes of future races and events. One particularly interesting study was released by Maier, Meister, Trösch and Wehrlin [3]. In their paper they train machine learning models in order to predict the outcome of a single biathlon shot. Their models show some predictive power, however there remained a high degree of randomness which could not be explained by the models.

In this project my first goal is to build on the base provided by Maier et al. and try to create an improved machine learning model, since the performance of their models was not ideal and the authors considered improvements to be possible. There are several avenues I pursued in order to increase performance:

1. **Incorporating more features.** While the model by Maier et al. included 48 model variables, there were still some potential factors that could be modelled as a feature. One important area is weather data. Every regular watcher of biathlon races can testify that wind and weather can play a big role to the outcome of a race, thus it could also be interesting to a ML model. I also identified some other potentially relevant new features, like features relating to the style of the race (against the clock or head-to-head).

2. **Improving existing features.** I identified some minor areas in which the existing features could be improved to provide more clarity and relevance to the model.

3. **Providing more data.** The authors suggested that more data could help the model to achieve more predictive power. In my project I also included more varied data from more seasons and from different levels of competition, while Maier et al. only used data from the World Cup (the highest competition level).

In the following chapter I will describe the added features and changed methods in more detail. In the chapter Results, I will show the results of the improved model.

Using predictions generated by machine learning to improve the broadcasting and viewer engagement is popular in many different sports. The second part of my project is intended as a first step in this direction for the sport of biathlon. I implemented a framework where users can generate predictions for a biathlon shooting, based on parameters that can be defined by the user interactively. This dashboard was implemented in R shiny and published on `https://nikostra.shinyapps.io/biathlon_pred/`. The chapter *Interactive dashboard* describes the implementation and methodology of this part of my project.

2. Method

2.1 Data

There are two primary data sources that freely publish biathlon data. The first is the official datacenter of the International Biathlon Union (IBU) [1]. Here comprehensive data from all biathlon races in all competition levels is available in PDF format. The other data source I used is the archive of HoRa Systemtechnik GmbH, a major supplier of biathlon target systems. They provide shooting data for all races they are responsible for, which amounts to around 50% of World Cup races. This data set is available here[2] in a handy .xlx format, which makes processing much easier.

In my project I included all individual competitions from the seasons 2018-19 to 2022-23 from the two highest competition levels (World Cup and IBU Cup) that are available in the HoRa data set. Relay competitions are excluded because they have slightly differing rules for the shooting. In total this amounts to 205,600 indidual shots. This means that I provide roughly 25% more data than Maier et al. in [3]. 10% of the data was used as test data while the rest was used for training. There was a small number of rows with missing values for some of the variables. These I decided to remove from the data set instead of for example imputing values, since it was only a small number, not likely to strongly impact the final result.

2.2 Feature engineering

My goal was to add new features to the model in order to make more accurate predictions. To achieve this I combined weather data from the IBU datacenter with the existing data set. Weather readings were taken 30 minutes after the start of each race, so they only account for a general trend and not for precise analysis of each shot. Still this data could be interesting to the model. Especially relevant for the shooting process is the wind speed. I included two features for the wind speed, one is the numerical wind speed and one is a categorical translation of the wind speed according to the beaufort scale [2]. The categories are "calm" with a wind speed between 0 and 1.5 m/s, "moderate" with a wind speed between 1.5 and 3.3 m/s and "strong" with a wind speed over 3.3 m/s. We can see in figure 2.1 a clear trend where a higher wind speed corresponds to a lower hit rate. Other weather related features that were added are air temperature, weather conditions (sunny, overcast, fog, snowing or raining) and the snow condition (compact, hard packed, powder, wet).

In addition to the weather features I also tried to improve the precision of the model by performing some minor changes of the features used by Maier et al. and adding some new ones. I added features for the shot number of this series and a scaled version of the total number of shots in the race in order to account for differences in the disciplines, as some races have 10 total shots, while others have 20. Another interesting new feature was one indicating if a shot is the last shot of a race or not. Since a big part of biathlon is the mental game there is a lot of pressure on biathletes, as a single missed shot can ruin an otherwise perfect race. This pressure is particularly strong on the last shot, as biathletes often describe that this is the moment where the thoughts

[1] `https://biathlonresults.com/#/start`

[2] `https://hidrive.ionos.com/share/75yh8.f027#$/`

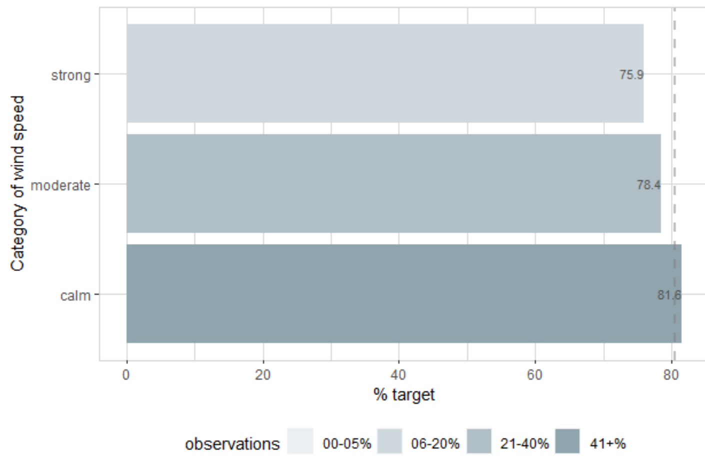

Figure 2.1: Comparing wind speed category and target hit rate.

and doubts come in. I investigated this effect and indeed we can see a difference in hit rates between these last shots and other comparable shots, as seen in figure 2.2.

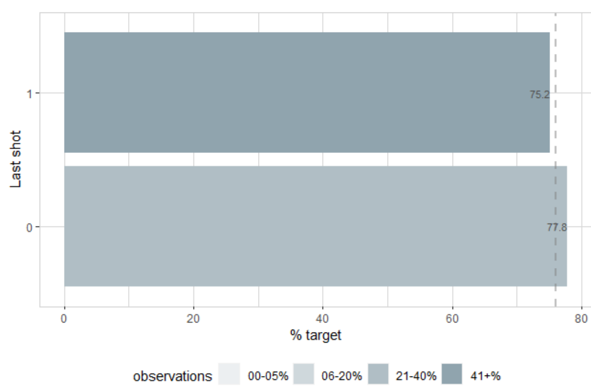

Figure 2.2: Comparing the hit rates of the last shot of a race with other comparable shots (standing shooting and shot number 5 in series).

After encoding categorical features using one-hot encoding I end up with 46 total features for the model. See table 2.1 for a detailed list of all features.

2.3 Prediction

To predict shot outcomes I use the Xgboost framework [1] in my project. This model was chosen for a number of reasons. It proved that it performed well in many industry applications and in machine learning competitions. In addition it was the best performing model of the ones tested by Maier et al. [3]. Also it is lightweight and does not have high performance requirements. Since many of our features cannot be assumed to uncorrelated and some are heavily correlated, it is useful to use a model like Xgboost which is robust to these correlated

Table 2.1: Table of model features. Features that were already used by Maier et al. have a white background, while new features added in this project have a green background color. N denotes the total number of features after one-hot encoding of categorical features.

Group	**Variables**	**N**
Competition	Season, Location, Discipline, Gender	16
Shooting	Lap, Lane, Mode, Shot number, Aiming time	5
Preceding hit rates	In this race, Overall (10, 50, 200), mode-specific (10, 50, 200), mode and shot number specific (200)	8
Weather	Wind speed, Wind category, Air temperature, Weather, Snow condition	11
other new features	Last Shot, Head to Head, Shot number scaled, Competition level, Long aiming time (categorical)	6
Target	Result of shot (hit/miss)	1

features. Finally due to it's decision tree structure Xgboost easily handles NA values in the data, which is important for the second part of the project, where we will try to predict future events with some variables left undefined.

In order to tune the model and find the best hyperparameters I used cross validation on the training data. The parameters to be tuned are the learning rate, the maximum depth of the trees and the number of boosting iterations. In order to find the best model I used the AUROC (Area Under the Receiver Operating Characteristics) metric, as it is well suited for our imbalanced classification problem. The best performing model is tested on the test data to get a final accuracy score.

3. Results

The best performing model was using the following parameters: learning rate (eta) = 0.15, maximum number of iterations (nrounds) = 150, maximum tree depth (max_depth) = 3. Using these hyperparameters I trained a model on the training data and tested it on the test data. This achieved an AUROC of 0.647. The corresponding ROC curve is shown in figure 3.1. This means that given two shots, a hit and a miss, the model is able to correctly classify them in 64.7% of cases. The optimal decision threshold lies at 0.8072 according to our model. When the predicted probability lies above this value, a shot should be classified as a hit, otherwise as a miss.

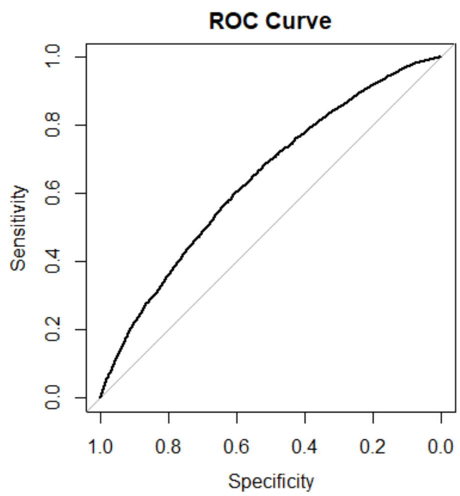

Figure 3.1: ROC curve for the best performing model with AUROC = 0.647

3.1 Discussion of results

Now it is interesting to compare our findings with the model produced by Maier et al.[3]. In their report they considered three models. A logistic regression model, a neural network and a Xgboost model. These models achieved AUROC scores between 0.60 and 0.62. As we can see my model improves on their score by a few percentage points. How did this improvement happen?

To answer this question we consider the feature importance table 3.1, where the ten most important features are listed. Here we can see that the prior hit rate is by far the most important feature. Specifically the model relies heavily on the average hit rate of the prior 200 mode-specific shots, as the importance score for this feature is much higher than for all other features. This effect was also described by Maier et al. Other features play only a smaller role, but variables like the shooting mode (standing or prone) and the shooting time are also considered by the model and show up in the top ten most important features.

Table 3.1: Feature importance for the model. New features noted in green.

Feature	**Gain**	**Importance**
pre_hit_rate_200_mode	0.40024789	0.40024789
pre_hit_rate_200	0.08293004	0.08293004
shooting_time	0.08018220	0.08018220
wind_speed	0.07629648	0.07629648
pre_hit_rate_50_mode	0.05480529	0.05480529
mode	0.04528312	0.04528312
pre_hit_rate_200_mode_shotNr	0.02467484	0.02467484
pre_hit_rate_50	0.02392918	0.02392918
air_temp	0.02207018	0.02207018
shot_number_race_scaled	0.02041007	0.02041007

We can also see that some newly added features are taken into consideration by the model. The wind speed seems to be quite useful as it still has a significant importance value for the model and shows up in fifth place. The variables air temperature and scaled shot number also appear in the ten most important features. These new features and the improved data set can explain the uplift in performance for our new and improved model.

But we also have to consider the fact that the improvement in performance was only minor and there is still a high degree of randomness in the prediction that cannot be explained by the model. Adding and improving the features and the data set helped the model, but its accuracy and predictive power is still far from optimal. While this is not necessarily an unexpected outcome, it also strengthens the theory proposed by Maier et al. that most of the predictive information lies in the preceding hit rate of an athlete, combined with a high degree of randomness for each shot.

4. Interactive dashboard

In this part of the project I propose a way to use the Xgboost model to generate predictive shooting outcomes for future races. This is interesting as it could be used to make live broadcasts of biathlon races more interesting and engaging for the viewer, Imagine a situation where a group of biathletes head into the final shooting together, the broadcast could display simulated shooting outcomes or a clean shooting probability for each athlete in order to build up excitement and inform the viewer.

4.1 Method

To simulate these outcomes we use the Xgboost model. Since we want to predict outcomes for a future situation, some of the model variables will be unknown, while others can be defined. We will know the biathlete who takes the shot (and thus also his preceding hit rates), the location, discipline and the lap and mode of the shooting. Other variables are harder to define prior to the fact like the hit rate in this particular race. These could be imputed, but since this would be imprecise at best and Xgboost is robust against NA values I chose to leave them undefined. This way the model produces a predicted hit probability for each shot we want to simulate.

Now the question is how we can use these hit probabilities to simulate a biathlon shooting. One way would be to use the optimal classification thresshold we got from the AUROC metric. This lies at 0.8072. Every shot with a higher probability would be classified as a hit, otherwise as a miss. The problem with this approach is since the probabilities are heavily influenced by the preceding hit rate of an athlete this would lead to situations where for some athletes with lower average hit rates, all shots would be classified as a miss, while for more proficient shooters every shot would be classified as a hit. This is far from a perfect situation and does not lead to realistic predictions.

Instead of that approach I chose to use binomial draws using the model probabilities. The users inputs several parameters (location, discipline, athlete, lap) based on which I construct the model variables. Then I use the model to get the probabilities of the 5 shots of the shooting series and do a binomial draw to simulate the outcome. This ensures more realistic and reasonable outcomes, than when I'm using the probabilities and the decision threshold.

4.2 The dashboard

The dashboard was built using R shiny. Users have a section where they can define the input parameters. When a prediction is simulated, two values are displayed for each of the five shots of the shooting series. Firstly the hit probability from the model is displayed in text form. Secondly a graphic is displayed, that shows the simulated outcome from the binomial sampling of the probabilities. The graphic uses the traditional biathlon targets to show if a shot was a hit or a miss. A white disk shows a hit, while if the target remains black, a miss was simulated for that shot. This way all information is available to the user in the dashboard, the precise predicted hit probability and also the simulated outcome.

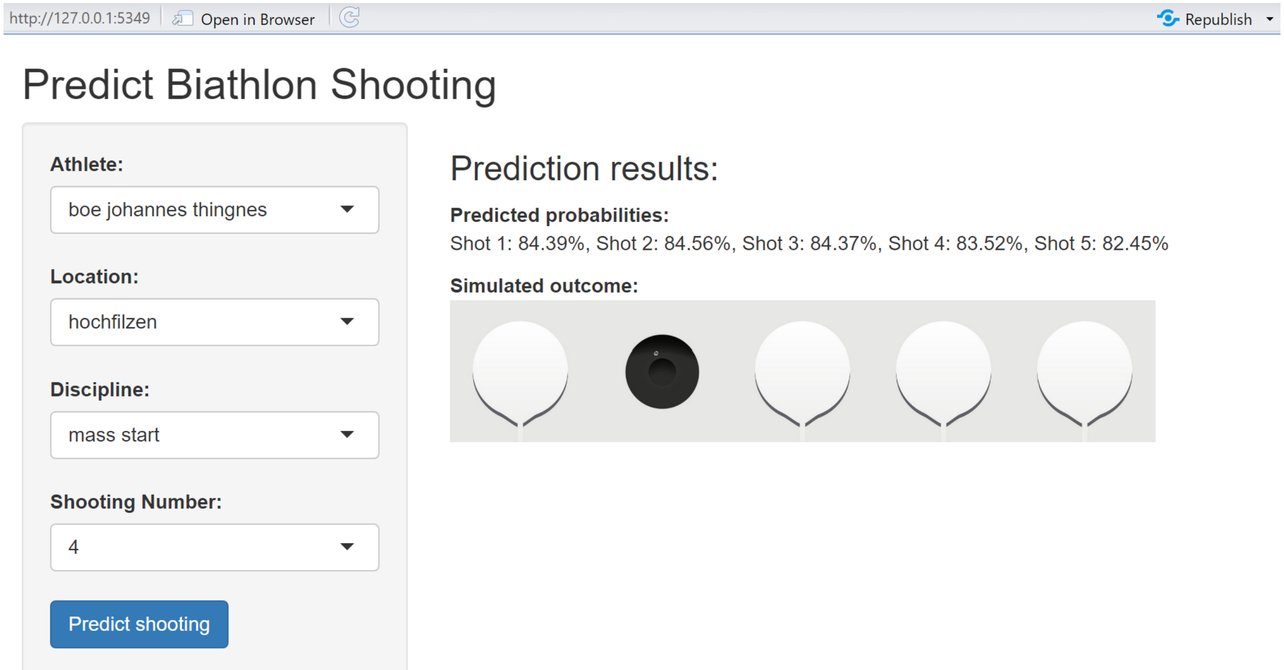

Figure 4.1: Simulation for the final shooting of Johannes T. Bø (World Cup winner 2023) in a mass start race in Hochfilzen.

The dashboard is available on `https://nikostra.shinyapps.io/biathlon_pred/` using the free tier of the platform shinyapps [1], which makes free hosting possible (with a limit of 25 hours uptime per month).

[1] `https://www.shinyapps.io/`

5. Future

The prediction performance is still not ideal, despite the improvements implemented in this project. More data could improve the model, especially data from the locations that are not covered by the shooting provider HoRa. Data from all World Cup locations and more competitions of lower competition levels might offer some additional insight for the model in the patterns behind biathlon shooting. This data is available in the IBU data center, but it would need to be parsed and combined from individual PDFs into a more usable format. Extracting the information from these PDFs computationally also poses a challenge because of the structure of these documents, but it could be worthwhile since they also contain interesting information about skiing times and intermediary rankings which could be used as model features.

Weather data showed some relevance for the model. If more precise weather data could be obtained, it could help the model to perform better. Especially interesting would be wind speed data at the time of a shot, for each shot instead of having just one wind speed value for the whole race as it is right now. This could help to improve accuracy by better accounting for gusts and changes in weather during a race.

The dashboard implemented in this project could be seen as a proof of concept for showing more advanced metrics and predictions in biathlon coverage and live broadcasts. Predicted shooting outcomes could also be used by athletes to find areas of weakness and improve them in training and by coaches to inform their decision when nominating the squad for a relay race. Winter sports in general and biathlon in particular are still at the beginning of the exploration using sports analytics and it is exciting to see what the future holds in this space.

Bibliography

[1] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[2] Scott Huler. *Defining the wind: the Beaufort scale and how a 19th-century admiral turned science into poetry*. Crown, 2007.

[3] Thomas Maier, Daniel Meister, Severin Trösch, and Jon Peter Wehrlin. Predicting biathlon shooting performance using machine learning. *Journal of sports sciences*, 36(20):2333–2339, 2018.