Abstract

Contents

Chapter 1

Introduction

1.1 Motivation

1.2 Aims

1.3 Methodology

Chapter 2

Literature Review

This chapter aims to present the current state of research in two different domains. The first is about predicting sporting events using machine learning. The latter examines sports betting with a particular focus on betting odds and how these can help to predict events in the future.

The established guidelines of Brocke et al., 2015 and Webster and Watson, 2002, were used to determine the research status and respectively document the literature search process. As stated by Webster and Watson, two types of literature reviews exist. This literature review belongs to the second type, which is, according to Webster and Watson, in general, shorter and where *'authors [...] tackle an emerging issue that would benefit from exposure to potential theoretical foundations'* (Webster and Watson, 2002, p. 14). First, as recommended by Brocke et al., the literature search process is documented as accurately as possible to facilitate future research on this topic. Then, the literature found is summarised in a concept matrix according to Webster and Watson and examined according to specially selected criteria. On this basis, research gaps get identified, and finally, the research question for this thesis gets formulated.

According to Brocke et al., in order to find relevant literature on the research areas dealt with in the thesis, the topic must first be divided into individual concepts. These concepts help to find literature in scholarly databases using keyword search. The keywords searched for in this thesis were *'fantasy football'*, *'machine learning'*, *'prediction'* and *'betting odds'*. The keywords were entered in every existing constellation to find articles that do not correspond to all keywords. Based on the research of Gusenbauer, 2019, *Google Scholar* and *Microsoft Academic*, the most extensive academic search engines were used for the literature search. When selecting the results from this search, attention was paid to the currently awarded VHB journal rankings (see V., 2015) for the sub-field of business informatics to ensure that the literature researched is of high quality. One journal that would be less considered following this approach, but seems extremely relevant to the research in this thesis, is the *Journal of Quantitative Analysis in Sports* (JQAS). This journal gets published by the American Statistical Association (ASA), which according to themselves, *'is the world's largest community of statisticians'* (see *About ASA* 2021). Using the papers from the JQAS and journals highly ranked by the VHB, the remaining literature got found using backward search and forward search suggested by Webster and Watson.

In the process mentioned above, 22 papers were examined and compared in a concept matrix as required by Webster and Watson. Due to its size and for the sake of readability, this matrix is in the appendix. Nevertheless, the concepts used to examine the papers will be briefly discussed from left to right in this paragraph.

The year of publication, the VHB ranking and the distinction in which form the paper was published serve to evaluate the quality of the literature. That is to ensure that primarily the most recent papers in renowned peer-reviewed journals were analysed. The sport discipline helps to notice similar approaches in different sports. While sports differ, some are more related than others. The main idea behind this is that there may be viable approaches from a similar sport that would have been unconsidered otherwise.

During the research, to the best of my knowledge, no publication was found which deals precisely with the problem at hand. For this reason, the research had to be focused on similar approaches, objectives or tasks. The solving approaches vary from more straightforward approaches such as mixed integer programming to more complex multi-hierarchical Bayesian models. Some publications used a combination of several methodologies, which are strongly dependent on the task to be solved. A distinction was therefore made between optimisation and prediction tasks. Although almost all papers unanimously had the goal of setting up a team that would score as many points as possible, they came at the solution differently. The matrix distinguishes between publications that optimised only the team performance as a whole and those that predicted the performance for each individual player and then combined the best players into a team. At the same time, it investigated which papers relied on betting odds or another form of prediction markets.

Lastly, the data used in each publication was analysed. Due to the always different data, a generalised view was applied, which examines whether time-series data is used, whether the home advantage was taken into account and whether betting odds were used.

Chapter 3

SPITCH

3.1 General

3.2 Game Rules

3.3 Scoring System

3.4 Competitors

Chapter 4

Data

4.1 Used Data

4.2 Descriptive Analytics

4.3 Data Preparation

Chapter 5

Modelling

5.1 Optimization

5.2 Prediction

Chapter 6

Machine Learning

6.1 Feature Selection

6.2 Model Selection

6.3 Model Training

Chapter 7

Evaluation

7.1 Combinatorial Model

7.2 Machine Learning Model

7.3 Performance Comparison

Chapter 8

Conclusion

List of Figures

List of Tables

Source Code

Bibliography

[1] *About ASA*. URL: https://www.amstat.org/ASA/about/home.aspx?hkey=6a706b5c-e60b-496b-b0c6-195c953ffdbc (visited on 07/20/2021).

[2] Jan vom Brocke et al. "Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research". en. In: *Communications of the Association for Information Systems* 37 (2015). ISSN: 15293181. DOI: 10.17705/1CAIS.03709. URL: https://aisel.aisnet.org/cais/vol37/iss1/9/ (visited on 07/16/2021).

[3] Michael Gusenbauer. "Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases". en. In: *Scientometrics* 118.1 (Jan. 2019), pp. 177–214. ISSN: 0138-9130, 1588-2861. DOI: 10.1007/s11192-018-2958-5. URL: http://link.springer.com/10.1007/s11192-018-2958-5 (visited on 07/16/2021).

[4] VHB e. V. *VHB-JOURQUAL3: Wirtschaftsinformatik*. 2015. URL: https://vhbonline.org/fileadmin/user_upload/JQ3_WI.pdf.

[5] Jane Webster and Richard T Watson. "Guest Editorial: Analyzing the Past to Prepare for the Future: Writing a literature Review". en. In: (2002), p. 11.

Appendix A

A.1 Diagrams

A.2 Tables

A.3 Screenshots

A.4 Graphs

Decleration of Authenticity

I declare that I wrote this thesis on my own and did not use any unnamed sources or aid. Thus, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person except where due reference is made by correct citation. This includes any thoughts taken over directly or indirectly from printed books and articles as well as all kinds of online material. It also includes my own translations from sources in a different language. The work contained in this thesis has not been previously submitted for examination. I also agree that the thesis may be tested for plagiarized content with the help of plagiarism software. I am aware that failure to comply with the rules of good scientific practice has grave consequences and may result in expulsion from the program.

Berlin, 13/09/2021

Jakob Heine