

Title: From Data to Dollar – Using the Wisdom of an Online Tipster Community to Improve Sports Betting Returns

Abstract:

With thousands of (online) bookmakers accepting wagers on sporting events, sports betting has become a billion-dollar business worldwide. Therefore, researchers and practitioners have gathered interest in investigating the “wisdom-of-crowds” effect in online tipster communities to predict the outcomes of sports events. We extracted 1,534,041 tips of 3,484 tipsters from *Blogabet.com* and used this user-generated content to investigate whether there is wisdom in online tipster communities that can be used to improve betting returns. We applied state-of-the-art data mining and natural language processing techniques and tested our hypotheses using quantitative research methods. Our results demonstrate that there is indeed wisdom in such online tipster communities that can improve sports betting returns. Tipsters won 3.29% more tips than the *implied win probability* set by bookmakers and produced averaged *yields* of 3.97%. We further identified four characteristics that are significant indicators for smarter sub-crowds within the overall crowd of an online tipster community.

Keywords: *data mining; natural language processing; online communities; sports betting; sports innovation; user-generated content; wisdom-of-crowds.*

Reference to this paper should be made as follows: Gruettner, A., Wambsganss, T. and Back, A. (2020) ‘From Data to Dollar – Using the Wisdom of an Online Tipster Community to Improve Sports Betting Returns’, *European J. International Management*, Vol. XY, No. XY, pp. XY-XY.

Biographical notes: Arne Gruettner is a Research Associate at the Institute of Information Management at the University of St.Gallen (HSG), Switzerland. He is the “Head of Research Group Sports Digitalization”. His research focusses on the practical applications of digital technologies to support various stakeholders in the Sports Industry. He is currently involved in

research projects on Data Mining, Machine Learning (ML) and Natural Language Processing (NLP) that investigate how user-generated content can be leveraged to gain valuable insights for e.g. bookmakers, sports federations and sports clubs.

Thiemo Wambsganss is a Research Associate at the Institute of Information Management at the University of St.Gallen (HSG), Switzerland. His research focuses on the design and application of Machine Learning (ML) and Natural Language Processing (NLP) techniques for artificial Intelligence (AI)-powered adaptive learning tools.

Andrea Back is a Full Professor of Information Systems at the Institute of Information Management at the University of St.Gallen (HSG), Switzerland. Her research areas are Strategy and Management Methods for the digital age with a focus on agile innovation management, especially in the area of mobile business and IoT solutions.

Arne Gruettner Institute of Information Management University of St.Gallen arne.gruettner@unisg.ch	Thiemo Wambsganss Institute of Information Management University of St.Gallen thiemo.wambsganss@unisg.ch	Andrea Back Institute of Information Management University of St.Gallen andrea.back@unisg.ch
---	--	--

1. Introduction

Sports betting has become a billion-dollar business around the globe. In Germany, for example, each day people place bets worth EUR 25 million. An increase of 21 percent over the previous year (Deutscher Sportwettenverband, 2020). Online tipster communities, such as *Betadvisor.com*, *Blogabet.com* or *Oddsportal.com*, offer semi-professional sports bettors, so-called tipsters, the opportunity to publish, share and explain their carefully elaborated tips over the internet. Community members, on the other hand, can comment and discuss those publicly available tip recommendations. Online tipster communities can be seen as a new type of sports-based entrepreneurship (see Ratten, 2011) driven by the emergence and rise of innovative digital technologies in the sports industry that cover a wide range of sports and have a lively community (Ratten, 2017; Gruettner, 2019). Considering the forecasting power (e.g. in terms of diverse knowledge and expertise), recent academic studies published in scholarly journals have shown interest in investigating the underlying dynamics of online tipster communities (e.g. Brown and Reade, 2019). In this vein, the user-generated content of tipsters – which we will refer to as *tipster-generated content* (TGC) in this study – offers the potential to become a revealing data source to improve sports betting returns.

TGC has proven to be valuable in predicting the outcomes of sports events, as it not only contains concrete predictions of match results but also (often) background information about the tipsters or even detailed textual match analyses. Existing studies have been published on the “wisdom-of-crowds” effect of Surowiecki (2004) (e.g. Schumaker, Jarmoszko and Labeledz, 2016; O’Leary, 2017; Peeters, 2018; Brown and Reade, 2019). The wisdom-of-crowds effect operates on the premise that an averaging of forecasts eliminates individual prediction errors and thus leads to greater accuracy. In other words, large groups of individuals are better at making predictions than individuals are. The effect has tremendous practical implications: First, it suggests that decisions made by the majority rule (or by averaging opinions) will