

Usefulness of Data Mining in sport management research – The case of sport clubs

Contact details

Name author(s): Pamela Wicker & Christoph Breuer

Institution(s) or organisation(s): Institute of Sport Economics and Sport Management, German Sport University Cologne

City and country: Cologne, Germany

Email address for correspondence: wicker@dshs-koeln.de

Aim of paper and research questions

Sport clubs are often objects of investigation in sport management research. Usually descriptive analyses are carried out regarding the presentation of the studies' results. Thereby, differentiations between sport clubs according to number of members or offered sports are undertaken. However, these features are not always the distinctive ones. As sport clubs are very heterogeneous, it is difficult to find out structures beside the mentioned ones and which features are really the distinctive ones. Therefore a more explorative procedure of data analysis should be used. Using Data Mining, several procedures and algorithms are applied simultaneously which lead to the detection of patterns and structures in mass data that would not have been identified using conventional statistical analyses (Cios, Pedrycz & Swiniarski, 1998; Petersohn, 2005). Compared to conventional statistical analyses the factors are identified within the procedure and not determined before the procedure. The aim of this paper is to show some advantages of the application of Data Mining in sport club research by identifying distinctive features of sport clubs. Thus this is rather a methodological paper.

Literature review

Data Mining was often applied in several fields of marketing research like customer relationship marketing, analysis of shopping behaviours, selection of target groups for marketing campaigns, customer-lifetime-value, and social network analysis (Hippner, Küsters, Meyer & Wilde, 2001). Moreover, Data Mining was already used in astronomy and molecular biology (Grob & Bensberg, 1999). Although Data Mining is a well known method in marketing research and other areas of research, there is a research gap concerning the application of Data Mining in sport management research and especially in sport club research.

Research design and data analysis

Data basis is the Sport Development Report 2007/08 (Breuer & Wicker, 2008), a nationwide online survey of sport clubs in Germany (N=13,068). Data Mining is applied in order to find out the distinctive features of specific types of sport clubs. Therefore several decision trees (C&R-Tree, CHAID) are estimated to reveal which factors determine (1) the proportion of immigrants, (2) the proportion of women, (3) the amount of membership fees, (4) the reception of public subsidies, (5) and the prevention of doping in sport clubs. (6) Moreover, the severity of several problems is analysed in all sport clubs and in specific sport clubs to find out protection and risk factors. These sport clubs are those with a high proportion of immigrants (at least 20 %, (7)) and competitive sport clubs (at least one national top athlete, (8)).

Results

The estimated decision trees revealed several interesting results. Some of them are presented in the following. First, the decision tree for the proportion of immigrants in sport clubs reveals that the most important factors are whether the sport club has undertaken specific activities regarding the integration of immigrants, the year of foundation, the proportion of immigrants in the federal state, and whether soccer is offered or not. Second, the offered sports are essential regarding the proportion of women in a sport club. Third, the amount of membership fees is strongly influenced by the existence of paid staff. Fourth, sport clubs with a high number of children and youths receive more often public subsidies. Fifth, the number of national top athletes is very decisive for the fact whether a sport club undertook measures in order to prevent doping. Sixth, the problem of the adherence and acquisition of youth top athletes is a bigger problem in sport clubs with fewer than 340 members and in shooting clubs. The problem of the adherence and acquisition of volunteers is smaller in sport clubs where a high percentage of members participated in convivial gatherings. Seventh, in sport clubs with a high proportion of immigrants, the adherence and acquisition of coaches is a bigger problem in soccer clubs with a high number of children and youths. Eighth, the financial situation in competitive sport clubs is a bigger problem in those sport clubs with a high number of national top athletes per member.

Discussion and conclusion

The application of Data Mining revealed some interesting correlations. However, some correlations are not surprising; but it is surprising that the noted factors are the distinctive ones. As sport clubs are very heterogeneous, Data Mining helped to identify new patterns and structures which would have been undetected applying conventional analyses. Thus, the application of Data Mining can be recommended in sport management research and especially in sport club research.

References

- Breuer, C., & Wicker, P. (2008). *Sports clubs in Germany. Analysis of the sports clubs' situation in Germany*. Cologne: German Sport University Cologne.
- Cios, K., Pedrycz, W., & Swiniarski, R. (1998). *Data Mining Methods for Knowledge Discovery*. Boston: Kluwer.
- Grob, H.L., & Bensberg, F. (1999). *Das Data-Mining-Konzept*. Münster: Westfälische Wilhelms-Universität Münster.
- Hippner, H., Küsters, U., Meyer, M., & Wilde, K.D. (2001). *Handbuch Data Mining im Marketing*. Braunschweig, Wiesbaden: Vieweg.
- Petersohn, H. (2005). *Data Mining. Verfahren, Prozesse, Anwendungsarchitektur*. München: Oldenbourg.