

DEVELOPMENT OF PREDICTION MODEL FOR SUPPORT IN DECISION-MAKING PROCESS IN FOOTBALL ACADEMIES – LITERATURE REVIEW

ROK VRBAN

Univeristy of Maribor, Faculty of Organizational Sciences, 4000 Kranj, Slovenia, e-mail:
rok.vrbani@student.um.si

Abstract Talent development process in football is considered as a process of providing the optimal environment for identifying and realising the maximum ability of young athletes. A multidimensional approach to analysing factors that influence junior to senior transition can produce much better support for coaches and management to distinguish elite and non-elite players. With development of digital technologies and artificial intelligence, more clubs are able to perform detailed analysis of their youth development programme. In this paper, we focus on identifying good practices in connecting digital technologies with talent development process in sports. Based on established methods and techniques used by experts in a field of data mining within sports, we want to select an appropriate methodology and approach in discovering knowledge from the data for the doctoral dissertation. Literature review presents a first step in a hollistic process of identifying key attributes in junior to senior transition. The findings suggest that the comprehensive approach towards analysing data in sports, results in better identification of skills and attributes of young athletes. Consequently, data mining in sports is becoming more and more important in assessing important characteristics on every level within talent development process.

Keywords:

talent
development,
data
mining,
digital
technologies,
literature
review

1 Introduction

European football market size has doubled over the past decades. Many private investors from Asia, Middle East and Russia have joined the European football landscape in that period. This inflow of funding raised transfer activity and many clubs started searching for better options on market rather than in their own youth academy. Financial capabilities allow big clubs to invest money in the more developed players, while clubs with limited financial abilities need to go through the entire junior-to-senior transition (JST) process in order to transition them to the first squad or sell them. As most of the clubs depend on selling players, they must adapt to current trends and development of new technologies. Among these are comprehensive statistical analysis, which, combined with use of sophisticated technology and artificial intelligence, create a new understanding of the development process and evaluation of the players. Measuring performance of individuals during training sessions and league games has increased in recent years together with technological improvements. Artificial intelligence and modern statistical approaches in discovering knowledge in data (data mining) have brought new ways of understanding key factors in measuring performance, skill, attributes, talent development and other important aspects, which play an important role in the world of football. Most of the small and medium sized European football clubs evaluate their youth prospects solely qualitatively or by using basic statistical methods, which turn out to be insufficient in accurate monitoring of talents' professional development. Moreover, the collected data has small impact on decision-making process as coaches and managers have limited or no professional support from data analysts. Consequently, the data present a burden rather than the advantage for the club.

2 Methodology

The central question of the research project is: Which athlete's attributes play a key role in successful transition from youth to professional football player?

In order to answer the central question, the purpose of the study is to create a model, which will help us recognize importance of factors in youth development in football. By creating a model, which will be based on quantitative research of sub-elite and elite youth academies across various European football clubs, we want to encourage

sport clubs around Europe to start using advanced data methods and modern technological equipment to improve their youth academies and to understand the value behind the work they put in raising youth to become professionals. First step is literature review where we will overview the topic, compare different views of authors about the existing problem and examine the gap in the research. Articles in the literature review will be obtained through various online databases, such as Web of Science (WoS) and other bibliographic databases, mainly connected to sports, talent development and digital technology. The key words used for literature review included “talent development in sports”, “talent development and artificial intelligence” and “data mining in sports”. The data for the research will mostly be collected through interaction with high profile subjects within management of the clubs and the players of the observed club and through different research methods. We plan on using a mixed method of design science research (DSR) and multiple case study (MCS). The main methodological approach will be DSR, while we intend on using MCS for collecting and analysing the data for model development. Quantitative data will be collected through observation and measurement tests of pre-defined mental, physical, tactical and technical skills. The data will also include other attributes and characteristics of each individual. Qualitative data will be collected through interaction with managers and coaches within youth academy and by interviewing observed players (semi-structured interviews). While some data are publicly available, the rest is to be collected in agreement with the selected clubs and individuals who are willing to provide their personal opinions on requested topics. The proposed duration of collecting the data is a year and a half. Each of the selected clubs will be evaluated once every 6 months in a span of a week. Clubs will be selected according to groups and sub-groups classified by the ECA evaluation system. Proposed sample involves up to 3 football clubs in each division (i.e. max. 12 clubs). Each club will be requested to assign their youth teams (between 13 – 17 years old) for the tests. Tests will be performed by coaches and training specialists according to instructions given by the researchers. Altogether the sample should involve between 600 - 1000 talents. The data will be analysed with modern statistical methods and tools. Model will be based on data mining process and evaluated through CRISP-DM process model. Anticipated methods include algorithms such as KNN, AdaBoost, Naïve Bayes and Random forest.

3 Existing literature

As football is the most popular and played sport in the world with estimated 3,5 Billion fans all around the world, much research has been done in the academic and non-academic field. With more and more technology involved in sports, many other professions, traditionally less connected to football, find their common points of research with football. A lot of research has been done connecting football and medicine (such as Hawkings et al., 2001, Krusturup & Krusturup, 2018, Kramer, 2010) or football and tracking systems (e.g. Link et al., 2016, Schütz et al., 2019). With development of football related topics, clubs learn and implement new business models to their youth academies. On the other hand, it is important to understand the perspectives of youth players themselves. Chamorro et al. (2016) suggest, that youth football players, who are not only interested in becoming professionals but also give importance to education and private life, appear to be more resourceful to cope with the transition to professional football. According to Franck (2018, p.55) JST is very complex and therefore, she suggests the following: “I recommend that future research should continue combining quantitative methods with in-depth interviews in a mixed-methods approach when exploring the JST. This type of approach provides detailed pictures of different pathways and life experiences. In the future, research exploring how social agents in the environment perceive their roles in assisting JST athletes could also be beneficial.« Beside athletic predispositions, it is also crucial for youth to compete on high level before moving to senior level. Hollings (2013) claims, that there is a higher probability for youth athletes, who have already competed on high level, to become elite senior players. Lorenzo et al. (2009) came to the same conclusion in basketball. Interviewed youth players pointed out tougher demand and seriousness when they train with the senior team. They required more mental and physical effort. In order to train more, they had to sacrifice their leisure time and consequently they wasted more time for studying. The transition from amateur to professional can differ from sport to sport. One of the most prominent researchers in the youth development process in sports, Stambulova (2010), presents a 5-step career planning strategy as a guideline for each individual to estimate different spheres in his/her current life and structure the future by thinking forward and adjusting priorities to the wished result. Bennie & O'Connor (2006) claim, that transition from youth to senior stage is also influenced by other forces, such as psychological, social, economic and political forces. This claim is consistent with the prediction, that development of youth players is not only

connected to one's motivation and existing sportive components, but also challenges and problems of a daily life. Another important factor in the youth development process are talent's parents. They are important for young athletes, but can often be too pushy (Stambulova et al., 2009). To check intensiveness of pushing young talents by their parents, coaches or other authorities, authors many times rely on the push pull anti-push anti-pull model. Ekström & Sundqvist (2009) use the model to investigate the reasons of continuing or dropping out of the sport at Swedish youth hockey players in transition. In order to understand why and how much an athlete improves, we must create a continuous process of measuring and evaluating individual's physical, tactical and technical attributes. Growing sport industry has become more demanding and consequently, professional sport teams hardly compete on an international level without a full support of advanced technology. Yokesh & Kumar (2015) state, that technology changes the way we see sports. It influences what we play and how we play it. It improves ways to predict and treat injuries. De Koning (2010) agrees with that by claiming, that technological improvements could be the key to future world records. Author based his claims on comparing world record times in 1500-m speed skating for men and development of technology. Cortsen & Rascher (2018) focus on technology and data within football. They claim that data inside sport have much higher value than many would think and are much more difficult to use effectively. The problem is not just technical but lack of financial and human resources as well. They also argue the ownership of the data, which is crucial for profitability. Yet, profitability is only possible if the quality of the data is sufficient and if the owner knows how to commercialize it properly. Rein & Memmert (2016) agree with these claims and add that big amount of data may pose a problem if the club does not have a sufficient organizational structure and knowledge behind using modern technology. Machine learning process in football can be valuable, but it must be in collaboration with sport scientists as the process of understanding big data within tactical and technical progress is extremely complex. With implementation of machine learning process Musa et. al. (2020) present multiple performance-related parameters in beach football and sepak takraw. Key performance indicators that define performance were found in beach football, while results from analysis of sepak takraw turned out inconclusive. However, it was demonstrated in the findings, that a number of performance indicators are essential in distinguishing between losing and winning. It was also concluded, that anthropometric indexes (such as standing and sitting

height, leg length, skinfold measurements and other) have an impact in performance of sepak takraw. Deep statistical analysis of the team and individuals performing in team sports have also shown correlation between behavioural characteristics and overall result of the team. Excellent players, who are hated in the dressing room, who are prone to injuries or homesick, are less likely to bring out the best in their colleagues (Tunaru & Viney, 2010). Another important characteristic of a successful talent development is motivation. Forsman et. al. (2016) identify motivation as one of the key reasons of successful talent development among football players in Finland. Motivation does not only help a player maximize his abilities, furthermore, it helps the team boost their confidence and rises morality. Authors also recognized agility, passing and centering skills as most important ones in becoming elite player. Pappalardo et. al. (2019) show that elite players do not always play excellence, they just achieve it more often than the other players. Murr et al. (2018) add, that elite players perform significantly better in dribbling tests and ball control, while shooting ability had little impact on future development success. Authors have also found significant positive correlation between motivation and future performance, which confirms Forsman et. al. assumptions. Players with higher self-determination were more likely to get selected to a higher performance level compared to the players with lower self-determination. The results are supported by Kelly et. al. (2020) analysis, which identified higher potentials had significantly better technical skills such as lob pass ability, possession reliability, pass completion and average total touches. Another important factor in talent identification at early age is relative age effect (RAE). Doncaster et. al. (2020) demonstrate a persistent bias toward selecting individuals born earlier in year in male football and basketball.

4 Conclusion

Based on this paper, the findings regarding talent development process and data analysis show significant correlation. This study systematically analysed literature review within connections between sports, talent development and use of data. It is suggested that an appropriate use of data can determine and identify future abilities and skills of young athletes. Moreover, it is required for professional sport teams to use modern machine learning processes in order to compete on a top level. However, it is important to understand the value of the ownership of the data. Furthermore, data need to be addressed professionally, as too much data might bring more confusion rather than positive effect if treated wrongly. Conclusions regarding

the importance of social factors show that parents, coaches and management of the club have an important role in talent development. Another important feature in JST is personality. Studies have shown that individuals with professional approach and teamwork mindset are more likely to influence positive on the other individuals within the squad. They are also more likely to advance faster in development process as they spend more leisure time for individual practice. Future research is required to explore other attributes and factors which play an important role in elite and sub-elite player assessment. Additionally, data driven decision-making in football is still relatively young process, which offers many opportunities to research and explore.

References

- Bennie A., O'Connor D. (2006). Athletic transition: An investigation of elite track and field participation in the post-high school years. *Change: Transformations in Education*. Volume 9.1, November 2006.
https://www.researchgate.net/publication/277786513_Athletic_transition_An_investigation_of_elite_track_and_field_participation_in_the_post-high_school_years. Accessed: 05/04/2020
- Chamorro J. L., Torregrosa M., Oliva Sánchez D., Calvo Garcíá T., León B. (2016). Future Achievements, Passion and Motivation in the Transition from Junior-to-Senior Sport in Spanish Young Elite Soccer Players. Cambridge University Press: 20 October 2016.
<https://www.ncbi.nlm.nih.gov/pubmed/27762186> Accessed: 10/04/2020.
- De Koning J. J. (2010). World Records: How Much Athlete? How Much Technology? *International Journal of Sports Physiology and Performance*, 2010, 5, 262-267.
https://www.researchgate.net/publication/45167594_World_Records_How_Much_Athlete_How_Much_Technology Accessed: 07/01/2021
- Doncaster G., Medina D., Drobnic F., Gomez-Diaz, A. J., Unnithan, V. (2020). Appreciating Factors Beyond the Physical in Talent Identification and Development: Insights From the FC Barcelona Sporting Model. *Front. Sports Act. Living* 2:91.
- Ekström, C., & Sundqvist, C. (2009). Karriärövergången från nybörjarstadiet till utvecklingsstadiet inom ishockey utifrån spelare, tränare och föräldrars perspektiv. (C-uppsats i psykologi inriktning idrott, 61-90p) Högskolan i Halmstad: Sektionen för Hälsa och Samhälle.
<https://www.diva-portal.org/smash/get/diva2:292741/FULLTEXT01.pdf>. Accessed: 26/03/2020
- Forsman H., Blomqvist M., Davids K., Liukkonen J., Kontinen N. (2016). Identifying technical, physiological, tactical and psychological characteristics that contribute to career progression in soccer. *International Journal of Sports Science & Coaching* 2016, Vol. 11(4) 505–513
- Franck A. (2018). THE JUNIOR-TO-SENIOR TRANSITION IN SWEDISH ATHLETES. A longitudinal study. Linnaeus University Dissertations Nr 305/2018. <https://www.diva-portal.org/smash/get/diva2:1168706/FULLTEXT01.pdf> Accessed: 10/04/2020
- Hawkings R.D., Hulse M. A., Wilkinson C., Hodson A., Gibson M. (2001). The association football medical research programme: an audit of injuries in professional football.
<https://bjsm.bmj.com/content/35/1/43.short> Accessed: 27/03/2020

- Hollings, S.C. (2013). The Transition from Elite Junior Track-and-Field Athlete to Successful Senior Athlete: Implications for Athletics High Performance Programmes. <https://core.ac.uk/download/pdf/56364504.pdf> Accessed: 06/04/2020
- Kelly, A., Wilson M. R., Jackson, D.T., Williams, C. A. (2020). Technical testing and match analysis statistics as part of the talent development process in an English football academy, *International Journal of Performance Analysis in Sport*, 20:6, 1035-1051.
- Kramer E.B. (2010). Football emergency medicine. *Emergency medicine in football is now regarded as a sub-specialty of emergency medicine. CME: Your SA Journal of CPD (Vol. 28, Issue 5)*.
- Krustrup P., Krustrup B.R. (2010). Football is medicine: it is time for patients to play! *British Journal of Sports Medicine* 2018;52:1412-1414. <https://bjsm.bmj.com/content/52/22/1412>. Accessed: 30/03/2020
- Link D., Lang S., Seidenschwarz P. (2016). Real Time Quantification of Dangerousity in Football Using Spatiotemporal Tracking Data. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5201291/>. Accessed: 25/03/2020
- Lorenzo, A., Borrás, P. J., Sánchez, J. M., Jiménez, S., Sampedro, J. (2009). CAREER TRANSITION FROM JUNIOR TO SENIOR IN BASKETBALL PLAYERS *Revista de Psicología del Deporte*, vol. 18, núm. 3, 2009, pp. 309-312. https://www.researchgate.net/publication/237030588_Career_transition_from_junior_to_senior_in_basketball_players. Accessed: 25/03/2020
- Murr, D., Feichtinger, P., Larkin P., O'Connor D., Höner, O. (2018). Psychological talent predictors in youth soccer: A systematic review of the prognostic relevance of psychomotor, perceptual-cognitive and personality-related factors. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0205337>. Accessed: 07/05/2021
- Musa, R. M., Majeed A. P. P. A., Kosni N. A., Abdullah M. R. (2020). Machine Learning in Team Sports Performance Analysis and Talent Identification in Beach Soccer & Sepak-takraw. *SPRINGER BRIEFS IN APPLIED SCIENCES AND TECHNOLOGY*.
- Pappalardo, L., Cintia P., Ferragina, P., Massucco, E., Pedreschi, D., Giannotti, F. (2019). PlayeRank: data-driven performance evaluation and player ranking in soccer via a machine learning approach. <https://arxiv.org/pdf/1802.04987.pdf>. Accessed: 07/05/2021
- Rein R., Memmert, D. (2016). Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *SpringerPlus* (2016) 5:1410
- Schütz N.S., Pedersen B.B., Rerup B. (2019). VALUE CREATION THROUGH DATA TRACKING TECHNOLOGIES IN THE FOOTBALL INDUSTRY OBSTACLES DYSFUNCTIONAL EFFECTS. *MCIS 2019 Proceedings*. 35. https://pure.au.dk/portal/files/173292768/VALUE_CREATION_THROUGH_DATA_TRACKING_TECHNOLOGIES_IN_THE_FOOTBALL.pdf. Accessed: 06/04/2020
- Stambulova N. (2010). Counseling Athletes in Career Transitions: The Five-Step Career Planning Strategy. https://www.researchgate.net/publication/232844112_Counseling_Athletes_in_Career_Transitions_The_Five-Step_Career_Planning_Strategy Accessed: 06/04/2020
- Stambulova N., Alfermann D., Statler T., Côté J. (2009). ISSP Position Stand: Career Development and Transitions of Athletes. *IJSEP*, 2009, 7, 395-412. https://www.issponline.org/images/isspdata/position_stands/Career_Development_and_Transitions_of_Athletes.pdf Accessed: 07/04/2020
- Tunaru R. S., Viney H. P. (2010). Valuations of Soccer Players from Statistical Performance Data. *Journal of Quantitative Analysis in Sports*. Volume 6, Issue 2, Article 10.
- Yokesh T.P., Kumar S. (2015). VARIOUS SPORTS TECHNOLOGIES ARE USED IN DEVELOPING THE SPORTS PERFORMANCE. ISSN 2231-3265 *International Journal of Health, Physical Education and Computer Science in Sports* Volume No.29, No.2 pp 5-7 https://www.researchgate.net/publication/339642888_Various_Sports_Technologies_are_used_in_Developing_the_Sports_Performance. Accessed: 07/01/2021