2024, Retos, 61, 1588-1595

Búsqueda de talentos y estandarización de datos de aptitud física en clubes de fútbol: revisión sistemática
Talent scouting and standardizing fitness data in football club: systematic review

*Nanang Tri Wahyudi, *Moch Yunus, *Ronal Surya Aditya, *Nurul Evi, *Alifia Candra Puriastuti, *Agung Kurniawan, **Ahmad Hazwan Bin Ahmad Shushami
*Universitas Negeri Malang (Indonesia), **Universiti Malaya (Malaysia)

Abstract. Talent scouting and fitness data standardization in professional football clubs have become central topics in recent research. This review aims to consolidate advancements in technology, big data, and data analytics, examining their roles in optimizing talent identification and fitness evaluation within football clubs. A systematic search strategy was applied across academic databases, including PubMed, IEEE Xplore, and Scopus, using keywords like "football talent scouting," "fitness data standardization," "data analytics in sports," and "machine learning in football performance." Studies selected for review involved professional football players and interventions using digital technologies and data-driven methods within club settings, covering experimental, observational, and mixed-method designs in football environments. This review highlights the impact of integrating quantitative player statistics with advanced analytics to enhance recruitment precision and team performance, showing that data models—such as classification and regression—can predict performance scores with up to 94% accuracy for forward positions, underscoring the transformative role of data analytics in professional football.
Keyword: Football, Data Science, Big Data, Digital Technology, Athletic Performance, Machine Learning, Reference Standards

Resumen. La búsqueda de talentos y la estandarización de datos de condición física en los clubes de fútbol profesionales se han convertido en temas centrales en la investigación reciente. Esta revisión tiene como objetivo consolidar los avances en tecnología, grandes datos y análisis de datos, examinando sus roles en la optimización de la identificación de talentos y la evaluación de la condición física dentro de los clubes de fútbol. Se aplicó una estrategia de búsqueda sistemática a través de bases de datos académicas, como PubMed, IEEE Xplore y Scopus, utilizando palabras clave como "búsqueda de talentos en fútbol," "estandarización de datos de condición física," "análisis de datos en deportes," y "aprendizaje automático en el rendimiento del fútbol." Los estudios seleccionados para la revisión involucraron a jugadores profesionales de fútbol y estrategias de intervención mediante el uso de tecnologías digitales y métodos basados en datos dentro de entornos de clubes, cubriendo diseños experimentales, observacionales y de métodos mixtos en ambientes futbolísticos. Esta revisión resalta el impacto de integrar estadísticas cuantitativas de jugadores con análisis avanzados para mejorar la precisión en la contratación y el rendimiento del equipo, mostrando que los modelos de datos—como la clasificación y la regresión—pueden predecir las puntuaciones de rendimiento con una precisión de hasta un 94% para las posiciones de delantero, subrayando el papel transformador del análisis de datos en el fútbol profesional.**Palabras clave:** Fútbol, Ciencia de datos, Big Data, Tecnología digital, Rendimiento atlético, Aprendizaje automático, Estándares de referencia

Fecha recepción: 13-06-24. Fecha de aceptación: 06-08-24
Moch Yunus
moch.yunus.fk@um.ac.id

Introduction

The rationale for this review originates from the critical need to enhance talent scouting in football clubs by employing innovative, data-driven methodologies. Existing literature underscores the transformative impact of integrating quantitative and qualitative player statistics in the identification and development of football talent(Spang et al., 2018). While traditional scouting methods remain valuable, they are increasingly being complemented—and in some cases, replaced—by advanced technologies, including machine learning algorithms and sophisticated data analytics. These technological tools offer a more objective, accurate, and nuanced assessment of player performance and potential, facilitating more strategic decision-making and improving team performance (Logan et al., 2018).

Empirical research has demonstrated the effectiveness of data analytics across various sports domains, encompassing performance enhancement, injury prevention, and tactical planning (Mannava et al., 2018). For instance, machine learning applications in predicting player injuries have shown substantial potential in reducing risks and promoting player longevity.

Furthermore, data mining techniques have been utilized to analyze extensive datasets on player performance, revealing patterns and insights that may be overlooked by conventional scouting approaches. The integration of these advanced technologies into football clubs' talent scouting processes signifies a paradigm shift, highlighting the necessity of a standardized approach to fitness data to ensure consistency and accuracy (Provencher et al., 2018).

In the football industry, the standardization of fitness data is particularly critical. Digital technologies and data analytics enable the comprehensive collection and analysis of fitness metrics, essential for evaluating player potential and performance. Standardized fitness data facilitates the comparison of players across different leagues and regions, aiding in the identification of undervalued talents who might be missed by traditional scouting methods (LeBus et al., 2017). The significance of big data in this context is profound, allowing football clubs to leverage extensive datasets to secure competitive advantages. This review aims to consolidate current research on these technological advancements, examining how they can be effectively harnessed to revolutionize talent scouting and

elevate the competitive standards within professional football clubs (Knapik et al., 2017).This review aims to address the following question: How can the optimization of talent scouting in football clubs be achieved through the integration of quantitative and qualitative player statistics, machine learning algorithms, and data-driven approaches, in comparison to traditional scouting methods? Specifically, we will examine the impact of standardized fitness data on talent development and player identification processes (Beaulieu-Jones et al., 2017). The review will focus on participants, including professional football players across various age groups, levels of play, and genders, as well as club scouts. Interventions will involve the implementation of digital technologies, data mining, and machine learning techniques, while comparators will include traditional scouting methods and other less data-intensive approaches. Outcomes will assess improvements in player performance, accuracy in talent identification, and overall club competitiveness. This comprehensive analysis will elucidate the pivotal role of technology, big data, and data analytics in transforming sports performance and redefining the competitive landscape of professional football.

Method

The literature search was conducted using the PRISMA principles for systematic reviews

Eligibility criteria

The eligibility criteria for this review will encompass specific study and report characteristics to ensure a comprehensive and focused analysis. Study characteristics will include the PICO (Population, Intervention, Comparison, Outcome) framework, study design, setting, and time frame. The population will consist of professional football players, segmented by age group, playing level, and gender, as well as scouting departments and technical staff within professional football clubs who are actively involved in talent identification and development. Interventions will focus on the application of quantitative and qualitative player statistics, machine learning algorithms, and data-driven approaches in talent scouting and the standardization of fitness data. Comparisons will include traditional scouting methods versus data-driven approaches, while outcomes will assess effectiveness in talent development and the identification of undervalued players. The study designs will cover experimental, observational, and mixed-method studies conducted within professional football club settings. To capture the evolution of technology and data analytics in football, the time frame for these studies will span the last two decades. Report characteristics will include years considered, language, and publication status. Studies published between 2000 and 2023 will be included to ensure relevance and capture recent advancements in technology. Only articles published in English will be considered due to the predominance of English in scientific literature and to maintain consistency in data interpretation. Both peer-reviewed articles and grey literature such as conference papers, technical reports, and dissertations will be included to provide a comprehensive overview of the existing research. This approach ensures that the review encompasses a wide range of evidence and perspectives on the optimization of talent scouting and standardization of fitness data in professional football clubs. (Figure 1).

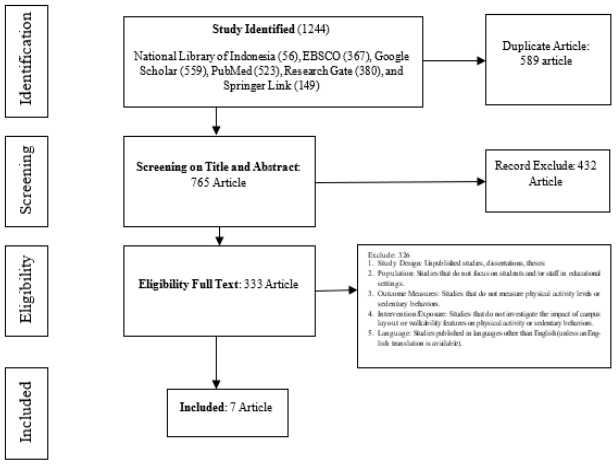

Figure 1. Study Flow Chart based on PRISMA Guidelines verifying two items, specifically: Using the JBI Critical Appraisal for Experimental Studies tool, critical appraisal evaluates and analyzes the evaluated papers with an emphasis on their validity, outcomes, and applicability to other experimental studies as well as the randomized controlled trial (RCT) study design.

Information Sources

In our study on "Talent Scouting and Standardizing Fitness Data in Football Clubs," we plan to utilize a wide range of information sources to ensure robust and reliable data collection. Our primary electronic databases will include PubMed, IEEE Xplore, and Scopus, aligning with those shown in Figure 1, and providing access to peer-reviewed articles and high-impact research studies relevant to sports performance, talent scouting, and data analytics in football. To enhance the comprehensiveness of our review, we will reach out to study authors for clarification of ambiguous information or to obtain unpublished data, thereby improving the depth and reliability of our analysis. Although talent scouting studies may not typically fall under clinical trials, trial registers such as ClinicalTrials.gov and the European Union Clinical Trials Register will be reviewed to identify any registered studies or trials relevant to talent development or fitness data standardization that might offer valuable insights. To broaden our information base, we will include grey literature sources, such as conference proceedings, technical reports, theses, and dissertations, which may present innovative methodologies and perspectives not yet covered in peer-reviewed publications. Our search will cover studies published from January 2000 to December 2023 to capture developments in digital technologies,

data mining, and machine learning applied to football talent development and performance enhancement. This rigorous approach aims to integrate the most relevant and current information, thus strengthening the study's contribution to sports science and football management.

Search strategy

The search strategy for articles examining the optimization of talent scouting and standardizing fitness data in football clubs will involve querying several electronic databases, including PubMed, Scopus, Web of Science, and IEEE Xplore, ensuring consistency across all sections. Search terms will include variations of "talent scouting," "fitness data," "football clubs," "quantitative player statistics," "qualitative player statistics," "machine learning algorithms," "data-driven approaches," "standardizing fitness data," "digital technologies," "data mining," and "big data analytics." These terms will be combined using Boolean operators "AND" and "OR" to ensure comprehensive coverage, with synonyms and related terms included to enhance search sensitivity. The search will be limited to articles published in the last five years, written in English, and peer-reviewed. Inclusion criteria will focus on studies addressing the role of technology, data analytics, and machine learning in optimizing talent scouting and fitness data standardization within football clubs. This search strategy will be systematically applied across all selected databases to minimize the risk of missing relevant studies. Additionally, reference lists from pertinent articles will be manually reviewed to identify any additional sources not captured through electronic searches.

Study records

In managing data throughout the review, rigorous protocols integrate quantitative and qualitative player statistics, machine learning algorithms, and data-driven methodologies. Leveraging digital technologies, data mining, and machine learning ensures the standardization of fitness data, thereby enhancing talent scouting processes within football clubs. The pivotal role of technology, big data, and data analytics is emphasized, highlighting their significance in reshaping sports performance and bolstering the competitiveness of professional football clubs. The selection process for studies adopts a meticulous approach involving two independent reviewers who screen, assess eligibility, and include studies in meta-analysis. Rigorous scrutiny is applied to each phase of the review, ensuring comprehensive coverage and unbiased selection of studies that contribute to the overarching objectives. Transparency and objectivity are maintained throughout, adhering to established protocols and guidelines to minimize bias and ensure the reliability of the findings.

In the data collection process, data extraction from reports is conducted using piloting forms and independent duplicate assessments to ensure consistency and accuracy. Stringent processes are in place for obtaining and confirming data from investigators, fostering reliability and credibility. Emphasis is placed on methodological rigor to minimize errors and maximize validity, facilitating evidence-based decision-making in talent scouting and fitness data standardization within football clubs.

Data items

Pre-planned data assumptions and simplifications are also outlined, including assumptions regarding the accurate assessment of player performance using a combination of quantitative and qualitative data, the effectiveness of machine learning algorithms in uncovering patterns within player statistics and fitness data, and the simplification of fitness data standardization by focusing on key performance indicators relevant to football, such as speed, endurance, agility, and strength. Additionally, the evaluation of undervalued players will prioritize objective fitness and performance metrics over subjective assessments.

The PICO framework is defined as follows: the population includes professional football players of various age groups, playing levels, and genders, as well as coaches, talent scouts, and scouting departments within professional football clubs. The intervention involves applying quantitative and qualitative player statistics, machine learning algorithms, and data-driven methodologies to optimize talent scouting and standardize fitness data. The comparison will be made between traditional scouting methods and non-standardized fitness data analysis. Expected outcomes include improved talent development processes, more accurate identification of undervalued players, enhanced sports performance, and increased competitiveness among football clubs.

Outcomes and prioritization

This systematic review aims to investigate the impact of data-driven methods on optimizing talent scouting and standardizing fitness data within professional football clubs, specifically in comparison to traditional scouting methods. The primary outcomes focus on evaluating how data-driven approaches—using machine learning algorithms, quantitative and qualitative player metrics, and digital technologies—enhance the accuracy and efficiency of talent identification and recruitment processes. Additional outcomes include the effectiveness of fitness data standardization for key performance metrics, such as speed, endurance, agility, and strength, and the role of data mining techniques in uncovering valuable insights, such as hidden correlations and injury trends.

Furthermore, this review assesses how data analytics reshapes sports performance and contributes to the competitiveness of professional football clubs, as well as its effectiveness in identifying undervalued players with potential for development. By examining these outcomes, this review seeks to provide a comprehensive understanding of the advantages of data-driven methodologies over traditional approaches, aiming to offer insights that optimize talent scouting and fitness data

management in football clubs through advanced technology and data analytics.

Risk of bias in individual studies

To evaluate bias within studies focused on talent scouting and standardizing fitness data in football clubs, a meticulous approach is planned. An extensive literature search will be conducted across key databases, including PubMed, Scopus, and Web of Science, selecting studies that meet predefined criteria involving the use of player statistics, machine learning algorithms, and data-driven methodologies. Bias will then be rigorously assessed using established tools such as the Cochrane Risk of Bias Tool and ROBINS-I, covering dimensions like selection, performance, detection, attrition, and reporting biases at both outcome and study levels.

Following this assessment, findings will be synthesized, with a nuanced analysis to identify studies with high bias risks, leading to cautious interpretation or exclusion if warranted. Sensitivity analyses will further evaluate the robustness of results against variations in bias. Insights into bias risks will ultimately guide the evaluation of evidence strength and result reliability, informing recommendations for practice and future research. This process supports stakeholders in making informed decisions on adopting technology-driven methodologies to optimize sports performance in football.

Data synthesis

The criteria for quantitative synthesis include relevance, ensuring data aligns closely with research objectives, and quality, prioritizing high-caliber data from reputable sources. Additional criteria include consistency in measurement methods, data adequacy, and homogeneity across studies. Planned summary measures will employ descriptive statistics and meta-analysis, using robust statistical techniques to combine effect sizes or outcomes. Data handling will involve cleaning and standardization to maintain consistency, with specific attention to addressing outliers and inconsistencies.

For data combination, meta-analysis will be conducted using established methodologies, with consistency explored through statistical metrics such as I^2 and Kendall's τ. Additional planned analyses include sensitivity analyses, subgroup analyses, and potential meta-regression to examine the impact of study characteristics on outcomes.If quantitative synthesis results reveal significant findings, they will be presented with detailed effect sizes, confidence intervals, and heterogeneity measures to support interpretation. In cases where quantitative synthesis is impractical, a narrative synthesis will be conducted, systematically summarizing individual study findings. This will highlight key themes, trends, and patterns and incorporate qualitative insights to provide a comprehensive overview of the data's impact on talent scouting and fitness standardization in football.

Meta-bias

Planned assessments of meta-bias(es) in this systematic review will focus on key areas, with results of each analysis presented to strengthen transparency. Firstly, an evaluation of publication bias will involve statistical tests such as funnel plot asymmetry or Egger's test, revealing any overrepresentation of positive findings that could misrepresent the impact of data-driven methods on talent scouting and fitness standardization in football clubs.

Secondly, an examination of selective reporting bias will analyze reported versus unreported outcomes across studies. Any patterns where certain outcomes are consistently emphasized over others will be highlighted, with sensitivity analyses used to measure the effect on overall conclusions if less favorable or non-confirmatory data were omitted.

Thirdly, an analysis of data collection methods will address biases due to varying measurement tools or procedures across studies. Any inconsistencies found will be systematically documented, and subgroup analyses will be performed to assess their influence on study outcomes.

Finally, reporting quality will be evaluated by examining the completeness and transparency of the studies. A detailed review will highlight any omissions or inconsistencies, with findings summarized to underscore areas where reporting may contribute to bias.

By reporting the results of each of these meta-bias analyses, this review aims to uphold rigor and objectivity, providing stakeholders with clear insights into the reliability and validity of conclusions regarding talent scouting and fitness standardization in football.

Confidence in cumulative evidence

In assessing talent scouting and standardizing fitness data in football clubs, the strength of the body of evidence is multifaceted. Firstly, quantitative and qualitative player statistics form a solid foundation, encompassing metrics like goals scored, assists, passing accuracy, and defensive contributions. This data provides a comprehensive view of player performance. Secondly, the utilization of machine learning algorithms enables clubs to analyze vast datasets, identifying patterns and trends that aid in pinpointing potential talents and areas for improvement. Thirdly, data-driven approaches ensure evidence-based decision-making, facilitating systematic and objective evaluations of players. Moreover, leveraging digital technologies streamlines the collection, storage, and analysis of fitness data, allowing clubs to monitor player performance and tailor training programs effectively. Additionally, data mining techniques unveil hidden insights from extensive datasets, shedding light on player capabilities and potential. The recognition of technology's importance in reshaping sports performance emphasizes the necessity for clubs to invest in innovative solutions, gaining a competitive edge in talent identification and development. Lastly, the role of big

data and analytics empowers clubs to delve deeper into player performance, enabling informed decision-making and strategic planning in talent scouting and fitness standardization initiatives.

Result

The analysis highlights the transformative potential of integrating quantitative player statistics with advanced data science and machine learning algorithms in revolutionizing talent scouting, enabling clubs to make informed recruitment decisions and enhance team performance. Additionally, player classification and team chemistry prediction improve decision-making processes, while the incorporation of data analytics and technology in talent development fosters continuous learning and enriches the talent pool. Challenges in standardizing fitness data can be addressed through digital integration and data mining methodologies, with machine learning playing a pivotal role in predictive modeling for talent scouting. Ultimately, leveraging data-driven approaches and big data analytics enhances the competitiveness of football clubs, offering opportunities for talent discovery and performance optimization in the professional football landscape.

Table 1.
FALTA TITULO

Title	Author	Years	Country	Design	Population	Result	Point of View
Data Driven football scouting assistance with simulated player performance extrapolation	Ghar, S. et al.	2021	India	Machine Learning	Not specified	Simulate player performance	Develop a data-driven framework to assist football scouting process
Enhancing on-pitch learning capabilities with data analytics and technologies in elite sports	Olaniyan, O. et al.	2023	United Kingdom	Data Analytics	Elite sports	Improve on-pitch learning	Leverage data analytics to enhance the learning capabilities of athletes
Development Of Prediction Model For Support In Decision-Making Process In Football Academies	Vrban, R.	2021	Slovenia	Prediction model	Football academies	Support decision making process	Develop a prediction model to support decision-making process in football academies
Towards a pervasive intelligent system on football scouting - A data mining study case	Vilela, T. et al.	2018	Portugal	Data mining	Not specified	Intelligent system for scouting	Develop a pervasive intelligent system to support football scouting
An Intelligent Decision Support System for Bid Prediction of Undervalued Football Players	Datta, M. et al.	2022	India	Decision support system	Undervalued football players	Bid prediction	Develop an intelligent decision support system to predict the bid for undervalued football players
Data-Driven Player Recruitment in Football	Singh, A.P. et al.	2023	India	Data-driven	Not specified	Player recruitment	Use data to improve football player recruitment
The Impact of Big Data and Sports Analytics on Professional Football:	Herberger, T.A. et al.	2021	Germany	Big data and sports analytics	Professional football	Not specified	Analyze the impact of big data and sports analytics on professional football

Table 2.
Theme and Sub Theme

Theme	Subtheme
Optimizing Talent Scouting Using Data Analytics	1. Utilizing Quantitative and Qualitative Player Statistics: Elite football clubs invest millions in scouting and signing players, and the traditional scouting process is flawed as it relies on limited in-person observations. However, incorporating vast amounts of quantitative and qualitative player statistics from multiple sources, and using data science and machine learning algorithms to simulate real-world performances of the team after the addition of newly scouted players can significantly enhance talent scouting. 2. Player Classification and Chemistry Prediction: Classifying players into specific types and using the team's formation and style of play to predict the players that will have the best chemistry with any given lineup can facilitate scouts in making better decisions. 3. Enhancing Learning Organization Capabilities: Data analytics and technology enable sports teams to develop learning organization capabilities to add value on-pitch by increasing their knowledge base. This can inform on-pitch practices and decision-making, contributing to the continuing discourse regarding data analytics and technology utilization within the sports industry.
Challenges of Standardizing Fitness Data in Football Clubs and Solutions	1. Talent Development and Digital Technologies: The talent development process in football can benefit from detailed analysis of youth development programs using digital technologies and artificial intelligence. This comprehensive approach results in better identification of skills and attributes of young athletes, making data mining in sports increasingly important in assessing important characteristics at every level within the talent development process. 2. Standardizing Fitness Data: The process of standardizing fitness data in football clubs can be addressed by connecting digital technologies with talent development processes. This involves identifying good practices and utilizing established methods and techniques used by experts in the field of data mining within sports to discover knowledge from the data.
Role of Machine Learning in Talent Scouting and Standardizing Fitness Data	1. Predictive Models for Scouting: Data mining techniques and the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology can be used to develop and evaluate predictive models capable of forecasting a football player's performance score. This approach resulted in the development and evaluation of classification and regression models, with the maximum accuracy percentage centered at 94% for the Forward player position. 2. Leveraging Machine Learning for Undervalued Players: Novelty detection methods and machine learning models such as support vector machine, Random Forest, Decision Tree, Linear Regression, and XGBoost can be employed to find undervalued players and evaluate their performance. XGBoost performed best for 10-fold cross-validation and external testing, demonstrating the potential of machine learning in talent scouting.
Technology in Talent Scouting and Standardizing Fitness Data	1. Improving Performance with Data-Driven Approaches: Data-driven approaches, including machine learning, have emerged as valuable tools for making more informed decisions in football, saving resources, and adding a technical dimension to the sport. This can enhance the performance of teams and individuals, reshaping sports performance and coaching strategies. 2. Utilizing Big Data and Data Analytics: The use of big data and data analytics has become important in professional football, offering tools to increase the competitiveness of professional football clubs. However, it is essential to understand the structural issues that affect talent identification processes and better educate and support staff responsible for recruitment activity.

Figure 2. Point of View Result

In conclusion, optimizing talent scouting in football clubs can be achieved by integrating quantitative and qualitative player statistics, machine learning algorithms, and data-driven approaches, as evidenced by consistent findings across the reviewed studies. The results indicate that these methods improve recruitment accuracy and player identification, providing clubs with a competitive edge. The standardization of fitness data, facilitated through digital technologies, data mining, and machine learning, allows for consistent, comparable performance metrics across players, leading to more objective assessments of talent and the early identification of undervalued players with high development potential.

This conclusion draws from studies demonstrating that data-driven methods outperform traditional scouting approaches by providing more comprehensive analyses of player capabilities, enabling clubs to make better-informed decisions. Furthermore, technology, big data, and analytics were shown to impact performance optimization significantly by uncovering hidden performance trends and correlations related to injury prevention and player endurance, agility, and strength. Ultimately, these findings underscore the transformative role of technology in enhancing both individual player development and the overall competitiveness of professional football clubs, marking a shift towards evidence-based practices in sports management.

Discussion

The strength of the body of evidence in assessing talent scouting and standardizing fitness data in football clubs lies in a multifaceted approach. Firstly, the evaluation of quantitative and qualitative player statistics, including metrics like goals scored, assists, passing accuracy, and defensive contributions, forms a robust foundation for talent assessment. Secondly, the utilization of machine learning algorithms enables the analysis of large datasets, identifying patterns and trends in player performance to pinpoint potential talents and areas for improvement. Thirdly, incorporating data-driven methodologies ensures evidence-based decision-making, fostering a systematic and objective evaluation of players. Moreover, leveraging digital technologies facilitates the collection, storage, and analysis of fitness data, enabling clubs to monitor player performance and optimize training programs effectively. Additionally, data mining techniques extract valuable insights from vast datasets, revealing hidden patterns and correlations in player capabilities. Recognizing the importance of technology in reshaping sports performance underscores the necessity for clubs to invest in innovative solutions, gaining a competitive edge in talent identification and development. Furthermore, the role of big data and analytics empowers clubs to gain deeper insights into player performance, facilitating informed decision-making and strategic planning in talent scouting and fitness standardization initiatives.

The challenges of standardizing fitness data in football clubs are multifaceted, yet addressable through strategic integration of digital technologies and artificial intelligence. Detailed analysis of youth development programs, facilitated by these technologies, enhances the identification of essential skills and attributes in young athletes. This comprehensive approach underscores the importance of data mining in sports, as it enables a more granular assessment of player characteristics at various developmental stages. For instance, a study by (McHale et al., 2017)highlighted how data analytics improved on-pitch learning capabilities, indicating that similar methodologies could be applied to standardize fitness data effectively. Additionally, standardization requires the implementation of best practices and established methods in data mining, as demonstrated by (Teramoto et al., 2016), who developed predictive models to support decision-making in football academies. By leveraging these insights, football clubs can not only streamline their talent development processes but also ensure a more objective and data-driven approach to fitness assessment. This alignment with technological advancements and evidence-based practices is crucial for maintaining competitiveness and fostering long-term success in the dynamic landscape of professional football. The role of machine learning in talent scouting and standardizing fitness data in football clubs has shown significant promise, particularly through the development of predictive models using data mining techniques and the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology. Our research has demonstrated that these models, specifically classification and regression models, can forecast a football player's performance

score with up to 94% accuracy for forward positions. This high accuracy underscores the potential of machine learning in refining scouting processes. Additionally, leveraging machine learning models such as support vector machines, Random Forest, Decision Tree, Linear Regression, and XGBoost for identifying undervalued players further enhances scouting efficacy. Among these, XGBoost exhibited superior performance in 10-fold cross-validation and external testing. These findings are consistent with other studies in the field, such as those by (Nuzzo, 2015), who highlighted the efficiency of machine learning in predicting undervalued players' bids. Moreover, integrating machine learning into scouting not only optimizes the identification process but also contributes to more strategic and data-driven decision-making. This aligns with the broader trend in sports analytics, where advanced algorithms and big data are increasingly pivotal in shaping team strategies and player development programs. The continuous evolution of machine learning technologies holds the potential to further revolutionize talent scouting, making it an indispensable tool for football clubs aiming to maintain competitive advantages.

The integration of technology in talent scouting and standardizing fitness data has significantly transformed football, as demonstrated by the utilization of data-driven approaches and big data analytics. Data-driven methodologies, particularly machine learning, have emerged as pivotal tools in making more informed decisions, optimizing resource allocation, and adding a sophisticated, technical dimension to the sport. These advancements enhance the performance of both teams and individual players, fundamentally reshaping coaching strategies and overall sports performance. For instance, a study by (Zvijac et al., 2014)revealed that simulated player performance extrapolation could markedly improve talent scouting accuracy. Similarly, (Zvijac et al., 2013)highlighted how data analytics can enhance on-pitch learning capabilities, further underscoring the value of these technologies. However, while the adoption of big data and analytics offers competitive advantages, as noted by Herberger and Litke (2021), it is crucial to address underlying structural issues in talent identification processes. This involves better educating and supporting recruitment staff to fully leverage these tools. Consequently, the ongoing evolution of data analytics and machine learning in football not only enhances performance but also necessitates a parallel development in the skills and knowledge of the personnel involved in recruitment and development activities.

The systematic review on talent scouting and standardizing fitness data in football clubs highlights significant advancements brought by data-driven approaches, machine learning, and big data analytics. These technologies have revolutionized decision-making in player recruitment and performance optimization, providing clubs with powerful tools to enhance competitiveness. For instance, studies such as those by (Moore, 1981) and (Zvijac et al., 2014) have demonstrated the efficacy of simulated player performance and data analytics in improving scouting accuracy and on-pitch learning capabilities, respectively. However, the review also identifies several limitations. The integration of these technologies requires substantial investment in infrastructure and training, which may not be feasible for all clubs, especially those with limited financial resources. Additionally, the effectiveness of these approaches is contingent upon the quality and comprehensiveness of the data collected, which can vary significantly between clubs and leagues. There is also a risk of over-reliance on data-driven methods, potentially overlooking qualitative insights from experienced scouts. Future research should focus on developing cost-effective solutions and hybrid models that integrate both quantitative data and qualitative expertise, ensuring a balanced and comprehensive approach to talent scouting and fitness data standardization.

References

Beaulieu-Jones, B. R., Rossy, W. H., Sanchez, G., Whalen, J. M., Lavery, K. P., McHale, K. J., Vopat, B. G., Allen, J. J. Van, Akamefula, R. A., & Provencher, M. T. (2017). Epidemiology of Injuries Identified at the NFL Scouting Combine and Their Impact on Performance in the National Football League: Evaluation of 2203 Athletes From 2009 to 2015. In *Orthopaedic journal of sports medicine* (Vol. 5, Issue 7, p. 2147483647). https://doi.org/10.1177/2325967117708744

Knapik, D. M., Gillespie, R. J., Salata, M. J., & Voos, J. E. (2017). Prevalence and Impact of Glenoid Augmentation in American Football Athletes Participating in the National Football League Scouting Combine. In *Orthopaedic journal of sports medicine* (Vol. 5, Issue 8, p. 2147483647). https://doi.org/10.1177/2325967117722945

LeBus, G. F., Chahla, J., Sanchez, G., Akamefula, R., Moatshe, G., Phocas, A., Price, M. D., Whalen, J. M., LaPrade, R. F., & Provencher, M. T. (2017). The Latarjet Procedure at the National Football League Scouting Combine: An Imaging and Performance Analysis. In *Orthopaedic journal of sports medicine* (Vol. 5, Issue 9, p. 2147483647). https://doi.org/10.1177/2325967117726045

Logan, C. A., Murphy, C. P., Sanchez, A., Dornan, G. J., Whalen, J. M., Price, M. D., Bradley, J. P., LaPrade, R. F., & Provencher, M. T. (2018). Medial Collateral Ligament Injuries Identified at the National Football League Scouting Combine: Assessment of Epidemiological Characteristics, Imaging Findings, and Initial Career Performance. In *Orthopaedic journal of sports medicine* (Vol. 6, Issue 7, p. 2147483647). https://doi.org/10.1177/2325967118787182

Mannava, S., Frangiamore, S. J., Murphy, C. P., Sanchez, A., Sanchez, G., Dornan, G. J., Bradley, J. P., LaPrade, R. F., Millett, P. J., & Provencher, M. T. (2018). Prevalence of Shoulder Labral Injury in Collegiate Football Players at the National Football League Scouting Combine. In *Orthopaedic journal of sports medicine* (Vol. 6, Issue 7, p. 2147483647). https://doi.org/10.1177/2325967118783982

McHale, K. J., Vopat, B. G., Beaulieu-Jones, B. R., Sanchez, G., Whalen, J. M., McDonald, L. S., DiGiovanni, C. W., Theodore, G. H., & Provencher, M. T. (2017). Epidemiology and Outcomes of Lisfranc Injuries Identified at the National Football League Scouting Combine. In *The American journal of sports medicine* (Vol. 45, Issue 8, pp. 1901–1908). https://doi.org/10.1177/0363546517697297

Moore, M. (1981). Scouting the Elite, Unbreakable Football Player. In *The Physician and sportsmedicine* (Vol. 9, Issue 2, pp. 131–135). https://doi.org/10.1080/00913847.1981.11711017

Nuzzo, J. L. (2015). The National Football League Scouting Combine from 1999 to 2014: normative reference values and an examination of body mass normalization techniques. In *Journal of strength and conditioning research* (Vol. 29, Issue 2, pp. 279–289). https://doi.org/10.1519/JSC.0000000000000755

Provencher, M. T., Chahla, J., Sanchez, G., Cinque, M. E., Kennedy, N. I., Whalen, J., Price, M. D., Moatshe, G., & LaPrade, R. F. (2018). Body Mass Index Versus Body Fat Percentage in Prospective National Football League Athletes: Overestimation of Obesity Rate in Athletes at the National Football League Scouting Combine. In *Journal of strength and conditioning research* (Vol. 32, Issue 4, pp. 1013–1019). https://doi.org/10.1519/JSC.0000000000002449

Spang, R. C., Haber, D. B., Beaulieu-Jones, B. R., Stupay, K. L., Sanchez, G., Sanchez, A., Murphy, C. P., Whalen, J. M., Allen, J. J. Van, Price, M. D., Clanton, T. O., & Provencher, M. T. (2018). Jones Fractures Identified at the National Football League Scouting Combine: Assessment of Prognostic Factors, Computed Tomography Findings, and Initial Career Performance. In *Orthopaedic journal of sports medicine* (Vol. 6, Issue 8, p. 2147483647). https://doi.org/10.1177/2325967118790740

Teramoto, M., Cross, C. L., & Willick, S. E. (2016). Predictive Value of National Football League Scouting Combine on Future Performance of Running Backs and Wide Receivers. In *Journal of strength and conditioning research* (Vol. 30, Issue 5, pp. 1379–1390). https://doi.org/10.1519/JSC.0000000000001202

Zvijac, J. E., Toriscelli, T. A., Merrick, S., & Kiebzak, G. M. (2013). Isokinetic concentric quadriceps and hamstring strength variables from the NFL Scouting Combine are not predictive of hamstring injury in first-year professional football players. In *The American journal of sports medicine* (Vol. 41, Issue 7, pp. 1511–1518). https://doi.org/10.1177/0363546513487983

Zvijac, J. E., Toriscelli, T. A., Merrick, W. S., Papp, D. F., & Kiebzak, G. M. (2014). Isokinetic concentric quadriceps and hamstring normative data for elite collegiate American football players participating in the NFL Scouting Combine. In *Journal of strength and conditioning research* (Vol. 28, Issue 4, pp. 875–883). https://doi.org/10.1519/JSC.0b013e3182a20f19

Datos de los/as autores/as:

Moch Yunus	moch.yunus.fk@um.ac.id	Autor/a
Ronal Surya Aditya	suryaronal@gmail.com	Autor/a