



08, 09, 10 e 11 de novembro de 2022  
ISSN 2177-3866

## **O futebol modelado por Machine Learning: uma revisão bibliométrica dos últimos 24 anos de pesquisa na área**

**FELIPE ALEXANDRE DE SOUZA FÉLIX NUNES**

UNIVERSIDADE FEDERAL DE MINAS GERAIS (UFMG)

**CARLOS ALBERTO GONÇALVES**

UNIVERSIDADE FEDERAL DE MINAS GERAIS (UFMG)

Agradecimento à órgão de fomento:

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

# **O futebol modelado por *Machine Learning*: uma revisão bibliométrica dos últimos 24 anos de pesquisa na área**

## **1. Introdução**

Em determinado momento o número de investigações científicas realizadas sobre a gestão de clubes de futebol cresceu, da mesma maneira que o esporte passou a ser tratado como um negócio bem mais que apenas recreação (Fajardo et al., 2017; Leoncini, 2001; Proni, 1998). Dentre as decisões tomadas pelos clubes de futebol está a alocação de recursos para a transferência de jogadores interclubes, cuja importância é justificada pelos investimentos que podem comprometer a estrutura financeira da organização (Payyappalli & Zhuang, 2019).

O futebol profissional é um mercado competitivo, é um sistema dentro da Teoria dos Jogos considerado tipicamente como Soma Zero, sendo preciso diminuir as margens de erro e aproximar-se da excelência, evitando que os custos da aquisição de novos atletas conduza à resultados desportivos ruins (Szymanski, 2017). Desta forma alguns clubes constituíram departamentos de análise de desempenho, ou scouting, e metodologias surgiram para identificar, explicar e recomendar os diversos elementos que compõem o desempenho de um jogador de futebol (Santos, 2012).

O *scouting* é uma ferramenta que municia os clubes com informações de equipes adversárias, da própria equipe, avaliações de treino, dos atletas de maneira individual e, também, de acompanhamento de jogadores cedidos pelo clube em regime de empréstimo (Ribeiro, 2009). É também usado para análise, detecção e prospecção de jovens atletas para os escalões mais baixos – em idades cada vez mais reduzidas – sendo um instrumento para ignorar menos e conhecer mais (Almeida, 2010).

Para codificar as ações do jogo em dados é preciso definir os indicadores de desempenho mais significantes para vencer as partidas (Zambom-Ferraresi et al., 2018). Como esporte de equipes integradas, atuando em faixas especializadas de competência, o futebol exige a combinação de desempenhos individuais aliados aos planos, modelos e estratégias gerais e operacionais (táticas) de jogo (Costa et al., 2009). Contudo o levantamento e a compreensão de indicadores de desempenho individual permitem aos treinadores avaliar a contribuição de cada um de seus atletas na equipe, nos tipos de eventos dentro do sistema de jogo proposto (Rein & Memmert, 2016).

É importante ressaltar que no contexto do esporte de alto rendimento, em determinado momento as investigações sobre os indicadores de desempenho passaram a buscar o “Santo Graal” do futebol (A. M. Williams & Ward, 2017). Alguns autores associam essa motivação ao conceito do *Moneyball*, trabalho de Billy Beane e Paul DePodesta no time de beisebol *Oakland Athletics* relatado no livro de mesmo nome (Lewis, 2003). O livro, romantizado posteriormente na visão do cineasta Bennett Miller, tratava a modelagem baseada em um método científico (chamado no livro de *Sabermetrics*) para identificar características essenciais em cada uma das posições para tornar mais eficiente e econômica as contratações naquele esporte (M. D. Hughes et al., 2012; Lewis, 2003; J. J. Williams, 2009).

Desta maneira, a busca por uma ferramenta que possa prever o impacto da contratação de um atleta nos resultados do clube de futebol adquiriu relevância. O conceito de recorrer aos dados estatísticos históricos para avaliar investimentos em jogadores de futebol mostra-se importante para minimizar riscos financeiros associados aos montantes gastos. A antecedente ineficiência das agremiações esportivas justifica o interesse na evolução de uma modelagem que promova menos prejuízo aos recursos dessas organizações.

O desenvolvimento tecnológico permitiu o surgimento de algoritmos de análise utilizando modelos de Aprendizado de Máquina e inteligência artificial. A aplicação destas técnicas na área da gestão esportiva vão recebendo atenção de cientistas no campo da engenharia, administração e do desporto (Almulla & Alam, 2020; Pariath et al., 2018). O objetivo deste artigo é avaliar as contribuições e benefícios da utilização do método de Aprendizagem de Máquina (*Machine Learning*) na avaliação de performance no futebol. Como método para essa análise utilizou-se a Revisão Bibliométrica. Essa técnica é correspondente à utilização de ferramentas matemáticas e estatísticas para estudar as publicações de um determinado campo de pesquisa (Diodato & Gellatly, 2013).

## **2. *Machine Learning***

A área de *Machine Learning* trata-se, por definição, de uma programação orientada a compreender regras a partir de dados, adaptar-se a mudanças e aperfeiçoar o desempenho por meio do aprendizado da experiência (Blum, 2007). O conceito é um “sonho inicial” da área da Ciência da Computação, com uma de suas origens no problema “*Imitation Game*” do matemático Alan Turing (Muggleton, 2014; Turing, 1950).

Em seu artigo Turing sugeriu que o trabalho dedicado ao desenvolvimento da Inteligência Artificial (IA) deixaria o campo de natureza puramente filosófica para uma tarefa de engenharia motivada por três frentes distintas: IA por programação; IA por aprendizado de máquina *ab initio*; IA usando lógica, probabilidades, aprendizado e conhecimento prévio (Muggleton, 2014; Turing, 1950).

O aprendizado pessoal, *i.e.* como os humanos aprendem, são definidos em teorias distintas que envolvem campos como Behaviorismo, Cognitivismo, Construtivismo, Experiencialismo e Aprendizagem Social. As máquinas aprendem com a experiência, dependendo de dados para tratar os cenários e obter maior precisão. Portanto, o aprendizado de máquina (*Machine Learning*) é o evento em que um programa de computador aprende com a experiência em relação a uma tarefa e uma medida de desempenho (Alzubi et al., 2018; Samuel, 1959).

A Teoria do Aprendizado de Máquina busca entender os princípios fundamentais do aprendizado como um processo computacional. Isto é, em um nível computacional busca-se compreender com precisão matemática informações e dados para melhorar o desempenho. Para tal é seguido o Princípio da Parcimônia, chamado também de “Navalha de Ockham (Occam)”, em que as explicações mais simples devem ser preferidas às explicações mais complexas (Blum, 2007).

Destarte o *Machine Learning* envolve a criação de modelos matemáticos que detectam os principais aspectos do aprendizado da máquina, desenvolvimento de algoritmos de aprendizagem que atendam aos critérios desejados e a análise matemática de questões gerais (Blum, 2007). O aprendizado de máquina é um campo multidisciplinar, com ampla gama de domínios de pesquisa, como pode ser visto na Figura 1 (Alzubi et al., 2018).

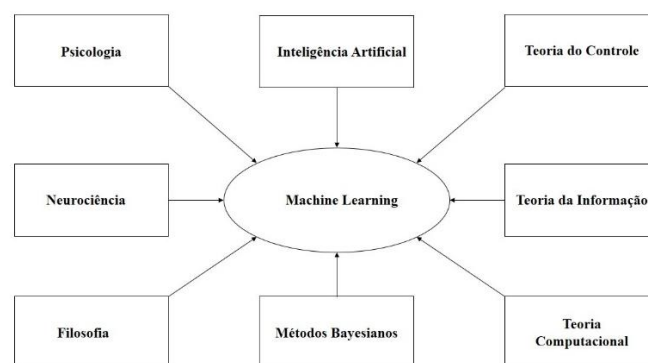


Figura 1 - *Machine Learning* multidisciplinar (Fonte: Alzubi et al., 2018)

### **3. Técnicas de análise matemática e futebol**

Identificar as características condizentes com a filosofia do clube para selecionar e desenvolver indivíduos talentosos é um importante aspecto do futebol atual como negócio. Nesta perspectiva tem-se definido que para um jogador atuar em alto desempenho não lhe basta apenas nascer com talento, mas uma intensa carga de estímulos e treinos, uma vez que diversos fatores vão impactar esse desenvolvimento (Macdonald, 2011). Também dentro desta ótica considera-se que um atleta é constituído principalmente de características do campo físico, técnico, tático, psicológico, clínico e administrativo (Paoli, 2007). Assim como aspectos sociais do jovem potencial, como estrutura famílias, ambiente em que está inserido e escola (Böhme, 2010).

O atleta, considerando benchmarking, de elite é quase visto como um sobrevivente, na visão mais Darwiniana do desporto, dado ao ambiente excessivamente competitivo em que apenas os melhores alcançam um domínio elevado de todas as componentes futebolísticas (Day, 2011). E identificar atletas com potencial demanda entendimento extenso das obrigações da performance, além de uma avaliação contundente das capacidades de um atleta em relação a estas exigências (Day, 2011).

Embora o futebol exija que o atleta não apenas se destaque por um aspecto, mas pela combinação entre eles, o elemento técnico é considerado fator fundamental da capacidade de um jogador atingir uma performance superior (M. D. Hughes et al., 2012; Meylan et al., 2010; Zhou et al., 2018). É por este motivo que algumas investigações se debruçam sobre o entendimento e a codificação desses indicadores sugerindo relações entre determinados índices e as vitórias ou derrotas de uma equipe (Yi et al., 2018; Zhou et al., 2018).

Evidências empíricas destacam a importância da análise sequencial de ações do jogo e a análise individual de times em uma competição para melhor entendimento da performance. Partindo da ideia que o futebol é um esporte dominado por aspectos estratégicos e táticos é razoável suspeitar que elementos contextuais tenham uma grande influência na performance de times e jogadores (Costa et al., 2009). Muitos estudos fundamentados na análise notacional dos jogos puderam se concentrar no número de atividades técnicas praticadas pelos jogadores em relação ao resultado da partida (vitória, empate, derrota). Alguns trabalhos detalharam ainda mais as ações técnicas e sua influência no resultado (Konefał et al., 2018). As análises baseadas nos resultados dos

jogos são uma ferramenta útil para explicar as diferenças que os impactos contextuais tiveram sobre a performance dos atletas.

Indicadores de performance no esporte são usados para selecionar e combinar variáveis que definem os aspectos do desempenho. Esses indicadores são usados para constituir um perfil ideal que possa ser usado para prever comportamento futuro em atividades esportivas. No contexto do futebol a evolução dos indicadores de performance provê a oportunidade de desenvolver a qualidade das partidas (Zhou et al., 2018). A performance de jogo no futebol é um resultado das interações das atividades físicas, técnicas e táticas apresentadas pelos jogadores durante uma partida. Cada atleta cumpre tarefas e funções específicas, de acordo com a sua posição no campo de jogo, na intenção de alcançar resultados positivos e, eventualmente, ganhar o jogo (Yi et al., 2018).

Autores como Hughes & Probert (2006) definiram a análise de desempenho no futebol sobre cinco principais categorias: (i) física; (ii) tática; (iii) técnica-defesa; (iv) técnica-ataque; e (v) psicológica. Outras publicações sugeriram características chave para determinadas posições no futebol, no entanto essas investigações são em sua maioria baseadas em depoimentos de treinadores, o que provocam diferenças nas definições (M. D. Hughes et al., 2012).

Esta revisão torna-se necessária para integrar e dar uma perspectiva ampla do desenvolvimento do *Machine Learning* até os tempos mais recentes, com seus diversos algoritmos de aprendizagem, as aplicações e os desafios. O contexto é a utilização de algoritmos de aprendizagem de máquina no campo de pesquisa relacionado ao futebol. Espera-se que os achados possam orientar novas pesquisas sobre o tema, a fim de apontar as relações já existentes entre os dois campos de tal maneira que se possa avançar na utilização desta ferramenta dentro das investigações no esporte.

#### **4. Método**

Objetivando analisar as publicações relacionadas ao tema de *Machine Learning* utilizado como ferramenta de análise no futebol foi realizada uma revisão bibliométrica, cujo objetivo é investigar a evolução dos estudos científicos em certos temas utilizando técnicas matemáticas e estatísticas (Diodato & Gellatly, 2013). Esta revisão, portanto, buscou publicações sobre o tópico “*Machine Learning*” no contexto do futebol. O levantamento foi realizado nas bases de dados *Web of Science* e *Scopus*. Utilizou-se o software RStudio e o algoritmo *bibliometrix* (Aria & Cuccurullo, 2017).

O pacote *bibliometrix* foi desenvolvido na linguagem R utilizado para executar análises bibliométricas objetivando analisar a produção científica e prover visualizações intuitivas, apontando para elementos como a frequência e a centralidade dos artigos (Aria & Cuccurullo, 2017). As visualizações representam um mapa científico e o resultado de análises distintas, como modelos geométricos, redes temáticas ou mapas de proximidade entre itens (Cobo et al., 2011; van Eck & Waltman, 2010).

Ademais, a análise temporal indica a evolução do conceito dentro do campo de pesquisa, reconhecendo padrões, tendências e sazonalidades. Somam-se às ferramentas visuais os clusters, agrupamentos de referências, palavras-chave ou periódicos, relacionando-as por similaridade (Upham & Small, 2010). Para esse levantamento situou-se um período de 24 anos, de 1997 à 2021, utilizando termos de busca nas bases *Web of Science* e *Scopus* “*Machine Learning*” e variações do termo usado para buscar o esporte futebol “*football*” e “*soccer*”.

Foram encontrados 468 títulos na *Web of Science* e 394 retornos na *Scopus*, esta primeira contagem foi filtrada para conter apenas artigos, resultando em 238 apontamentos na *Web of Science* e 135 na *Scopus*. O *bibliometrix* permite a fusão dos achados das duas bases restando 273 artigos únicos.

## **5. Revisão bibliométrica**

### **5.1. Volume de publicações e temporalidade**

Nos últimos 24 anos de pesquisa sobre o tema é possível observar um estágio inicial que perdurou até o ano de 2013 com um crescimento pequeno até 2018. A partir de 2017 e do trabalho seminal de Gudmundsson & Horton (2017), e portanto nos últimos quatro anos deste recorte temporal, há um crescimento anual mais acentuado de artigos divulgando achados de pesquisas na área do futebol e utilizando *Machine Learning* em algum momento (Figura 2).

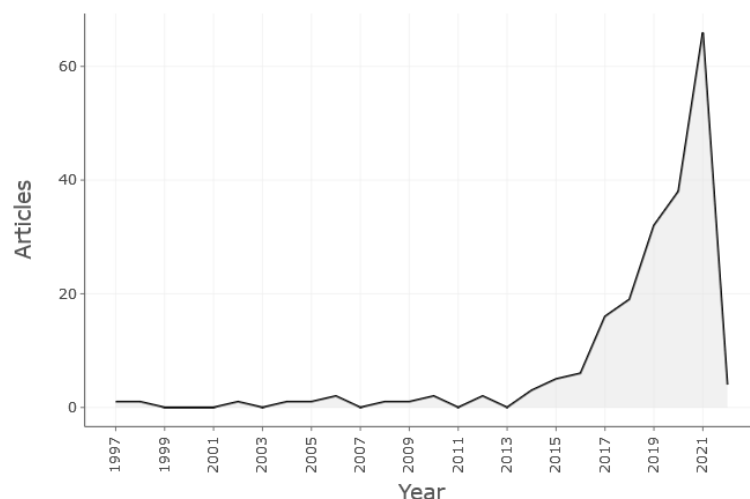


Figura 2 - Produção científica sobre *Machine Learning* e futebol

Entre os cinco trabalhos mais citados na busca dois se relacionam com o impacto do desempenho nos resultados, os trabalhos de Gudmundsson & Horton (2017) e o de Joseph et al. (2006), com dois dos títulos se relacionando ao entendimento e à prevenção de lesões dos atletas (Goswami et al., 2016; Rossi et al., 2018) e um trabalho relacionado ao estudo de torcedores de futebol nas redes sociais (Burnap et al., 2015). O resultado obtido na análise temporal das publicações também permite encontrar outros autores relevantes ao tema (Figura 3).

| <b>Autores</b>              | <b>Título</b>                                                               | <b>Ano</b> | <b>Citações</b> | <b>TC por ano</b> | <b>TC normalizada</b> |
|-----------------------------|-----------------------------------------------------------------------------|------------|-----------------|-------------------|-----------------------|
| Gudmundsson, J.; Horton, M. | Spatio-Temporal Analysis of Team Sports                                     | 2017       | 99              | 16,50             | 5,11                  |
| Burnap, P. et al.           | Detecting tension in online communities with computational Twitter analysis | 2015       | 79              | 9,88              | 2,84                  |
| Rossi, A. et al.            | Effective injury forecasting in soccer with GPS training                    | 2018       | 70              | 14,00             | 4,80                  |

|                    |                                                                                                                                       |      |    |      |      |
|--------------------|---------------------------------------------------------------------------------------------------------------------------------------|------|----|------|------|
|                    | data and machine learning                                                                                                             |      |    |      |      |
| Goswami, R. et al. | Frontotemporal correlates of impulsivity and machine learning in retired professional athletes with a history of multiple concussions | 2016 | 66 | 9,43 | 2,79 |
| Joseph, A. et al.  | Predicting football results using Bayesian nets and other machine learning techniques                                                 | 2006 | 62 | 3,65 | 1,49 |

Quadro 1 - Artigos mais citados

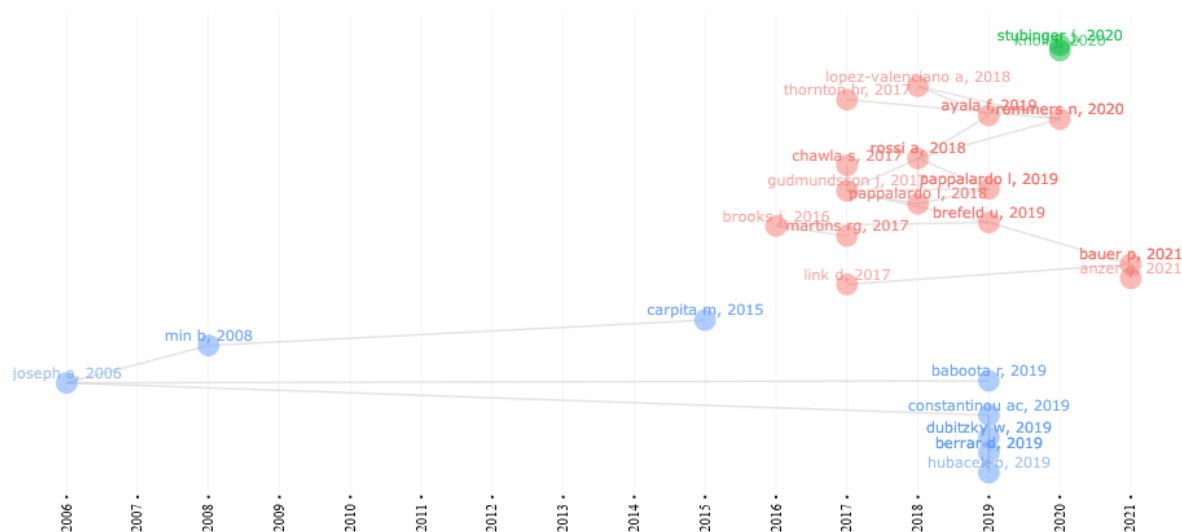


Figura 3 - Análise temporal da produção científica

## 5.2. Análise de clusters

A análise de clusters identificou três grandes relações entre os trabalhos da área (Figura 4), incluindo um polo englobando o trabalho de Carpita et al. (2015) derivado da linha proposta por Joseph et al. (2006), esse agrupamento dá especial destaque ao trabalho

recente de Galli et al. (2021). Uma outra linha derivada do trabalho de Joseph et al. (2006) culmina em um cluster que destaca trabalhos como Baboota & Kaur (2019) e Dubitzky et al. (2019). A linha de trabalho de Gudmundsson & Horton (2017) é acompanhada por títulos de Bauer & Anzer (2021), Pappalardo et al. (2019) e Pappalardo & Cintia (2018).

| Linhas de ação e principais artigos                                   |                            |                                                                                                               |
|-----------------------------------------------------------------------|----------------------------|---------------------------------------------------------------------------------------------------------------|
| Linha de ação                                                         | Autores                    | Títulos                                                                                                       |
| Artigos referentes à relação entre os dados e os resultados dos jogos | Carpita et al. (2015)      | Discovering the Drivers of Football Match Outcomes with Data Mining                                           |
|                                                                       | Galli et al. (2021)        | Football: Discovering elapsing-time bias in the science of success                                            |
| Artigos referentes à modelagem utilizando <i>Machine Learning</i>     | Baboota & Kaur (2019)      | Predictive analysis and modelling football results using machine learning approach for English Premier League |
|                                                                       | Dubitzky et al. (2019)     | The Open International Soccer Database for machine learning                                                   |
| <i>Machine Learning</i> como suporte ao processo decisório.           | Bauer & Anzer (2021)       | Data-driven detection of counterpressing in professional football                                             |
|                                                                       | Pappalardo et al. (2019)   | PlayeRank: Data-driven Performance Evaluation and Player Ranking in Soccer via a Machine Learning Approach    |
|                                                                       | Pappalardo & Cintia (2018) | Quantifying the relation between performance and success in soccer                                            |

Quadro 2 - Linhas de ação e principais artigos

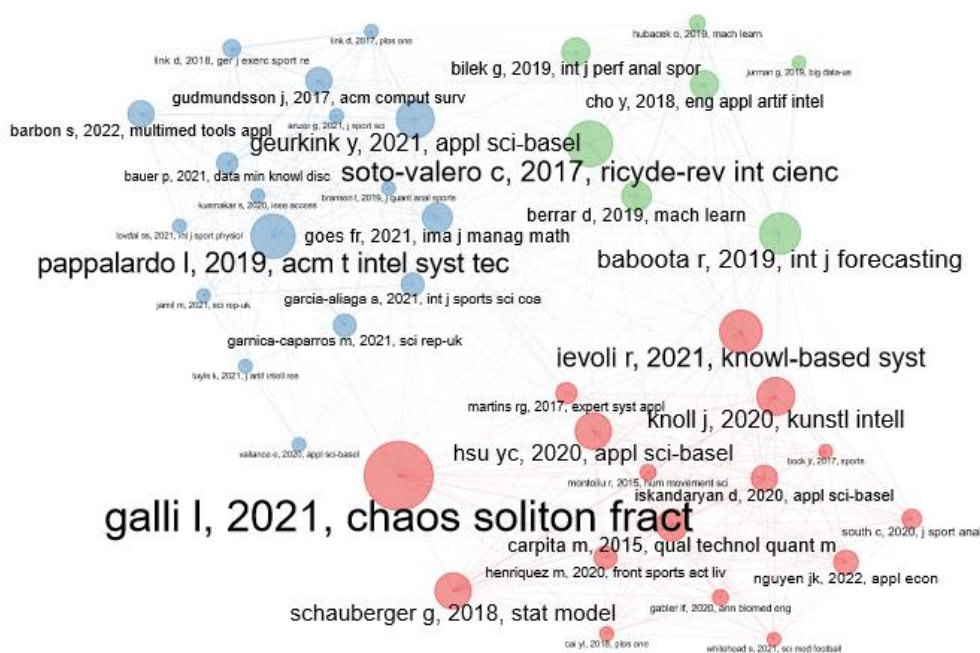


Figura 4 - Clusters dos principais artigos

### 5.3. Análise das palavras-chaves

As palavras-chaves são os principais tópicos dos quais se tratam as pesquisas sobre o tema, de uma maneira que o entendimento das conexões entre esses conceitos se torna relevante para uma pesquisa bibliométrica (Xiang et al., 2017). As palavras “*football*”, “*soccer*” e “*machine learning*” foram excluídas dessa análise por terem sido os termos de referência da busca nas bases científicas.

Desta maneira, entre as palavras-chaves utilizadas com frequência maior que 10 nas pesquisas levantadas, “*performance*” é o termo de maior frequência (16 vezes). Ela está ligada às palavras “*football*” e “*soccer*” e têm intercessões com as palavras “*model*” e “*scores*”. Essas relações podem ser interpretadas por trabalhos focados em estabelecer uma relação entre *Machine Learning* e futebol voltados para a modelagem da performance e dos gols (pontuações – “*scores*”). Em vista disso a palavra “*model*” é a segunda de maior frequência entre as palavras-chaves (12 vezes).

Outras duas palavras entre aquelas com frequência maior que 10 são “*players*” e “*selection*” indicando uma linha de investigações dedicadas à modelagem utilizando *Machine Learning* para a seleção de jogadores de futebol. Complementando essa relação por meio da análise da aglomeração entre todas as palavras-chaves identificadas, percebe-se que uma linha de investigações segmentadas nessa associação. Essa associação ainda remete a uma linha de palavras-chaves referentes aos algoritmos mais comuns utilizados

nas modelagens dos artigos que utilizam *Machine Learning* para compreender algum aspecto do futebol.

| Palavras-chaves    | Frequência |
|--------------------|------------|
| <i>Performance</i> | 16         |
| <i>Model</i>       | 12         |
| <i>Players</i>     | 11         |
| <i>Selection</i>   | 11         |

Tabela 3 - Palavras-chaves com frequência igual ou maior que 10 vezes

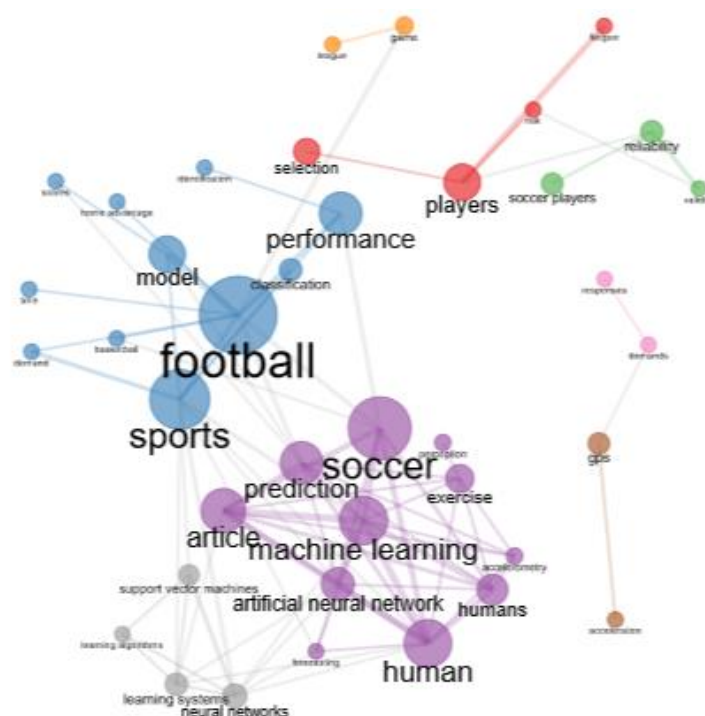


Figura 5 - Associação entre palavras-chaves

## 6. Conclusão

Este trabalho teve como objetivo compreender a utilização de uma técnica recente de modelagem estatística aplicada na análise de performance de jogadores de futebol. Utilizou-se o RStudio e o algoritmo *bibliometrix* para realizar uma revisão bibliométrica sobre o tema nos últimos 24 anos de publicações das bases *Web of Science* e *Scopus*, alcançando a marca de 273 publicações.

Observou-se com essa revisão que essa improvável mistura oriunda da Teoria Computacional com o campo da Ciência do Esporte vem adquirindo relevância à medida

que os trabalhos publicados vão se tornando mais frequentes. Ainda assim há uma potencial demanda por pesquisas que envolvam aspectos ainda não abarcados, tanto de uma das áreas (Computação) com de outra (Ciência do Esporte). Desta maneira, há uma vasta possibilidade para que novas pesquisas empíricas explorem os limites até agora alcançados para a expansão desses estudos.

Identificou-se uma potencial vertente para investigações acerca da utilização de indicadores no campo da performance, comprimindo palavras-chaves ausentes como (*technical* e *tactical* – com a primeira palavra-chave ausente do levantamento e a segunda com utilizada apenas duas vezes) dois aspectos frequentes nas pesquisas sobre análise de desempenho no futebol (Hughes & Probert, 2006; Paoli, 2007). Desta maneira, é preciso incorporar mais elementos referentes ao jogo de futebol nas modelagens computacionais.

O desenvolvimento de uma modelagem que identifique esses, e outros, fatores relevantes para o sucesso desportivo de uma equipe de futebol tem potencial de vantagem estratégica para os clubes de futebol. De maneira que esta revisão indica as tendências para futuras pesquisas cujos achados teóricos detêm potencial para evoluir estratégias de prospecção e contratação dos clubes de futebol.

## Referências

- Almeida, L. M. de O. M. de. (2010). *Perfil psicológico de prestação, orientações cognitivas e negativismo do futebolista português* [Dissertação de Mestrado]. Universidade de Trás-os-Montes e Alto Douro.
- Almulla, J., & Alam, T. (2020). Machine Learning Models Reveal Key Performance Metrics of Football Players to Win Matches in Qatar Stars League. *IEEE Access*, 8, 213695–213705. <https://doi.org/10.1109/ACCESS.2020.3038601>
- Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine Learning from Theory to Algorithms: An Overview. *Journal of Physics: Conference Series*, 1142, 012012. <https://doi.org/10.1088/1742-6596/1142/1/012012>
- Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Baboota, R., & Kaur, H. (2019). Predictive analysis and modelling football results using machine learning approach for English Premier League. *International Journal of Forecasting*, 35(2), 741–755. <https://doi.org/10.1016/j.ijforecast.2018.01.003>

- Bauer, P., & Anzer, G. (2021). Data-driven detection of counterpressing in professional football. *Data Mining and Knowledge Discovery*, 35(5), 2009–2049. <https://doi.org/10.1007/s10618-021-00763-7>
- Blum, A. (2007). *Machine learning theory* (26<sup>o</sup> ed). Carnegie Melon Universit, School of Computer Science.
- Böhme, M. T. S. (2010). Treinamento a longo prazo e o processo de detecção, seleção e promoção de talentos esportivos. *Revista Brasileira de Ciências do Esporte*, 21(2).
- Burnap, P., Rana, O. F., Avis, N., Williams, M., Housley, W., Edwards, A., Morgan, J., & Sloan, L. (2015). Detecting tension in online communities with computational Twitter analysis. *Technological Forecasting and Social Change*, 95, 96–108. <https://doi.org/10.1016/j.techfore.2013.04.013>
- Carpita, M., Sandri, M., Simonetto, A., & Zuccolotto, P. (2015). Discovering the Drivers of Football Match Outcomes with Data Mining. *Quality Technology & Quantitative Management*, 12(4), 561–577. <https://doi.org/10.1080/16843703.2015.11673436>
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, 5(1), 146–166. <https://doi.org/10.1016/j.joi.2010.10.002>
- Costa, I. T. da, Silva, J. M. G. da, Greco, P. J., & Mesquita, I. (2009). Princípios Táticos do Jogo de Futebol: Conceitos e aplicação. *Revista Motriz*, 15(3), 657–668.
- Day, D. (2011). Craft Coaching and the ‘Discerning Eye’ of the Coach. *International Journal of Sports Science & Coaching*, 6(1), 179–195. <https://doi.org/10.1260/1747-9541.6.1.179>
- Diodato, V. P., & Gellatly, P. (2013). *Dictionary of Bibliometrics*. Routledge. <https://doi.org/10.4324/9780203714133>
- Dubitzky, W., Lopes, P., Davis, J., & Berrar, D. (2019). The Open International Soccer Database for machine learning. *Machine Learning*, 108(1), 9–28. <https://doi.org/10.1007/s10994-018-5726-0>
- Fajardo, L., Werneck, F. Z., Coelho, E. F., & Matta, M. de O. (2017). A vantagem de jogar em casa em relação às séries do campeonato brasileiro de futebol. *Revista Brasileira de Futebol (The Brazilian Journal of Soccer Science)*, 10(2), 25–34.

- Galli, L., Galvan, G., Levato, T., Liti, C., Piccialli, V., & Sciandrone, M. (2021). Football: Discovering elapsing-time bias in the science of success. *Chaos, Solitons & Fractals*, 152, 111370. <https://doi.org/10.1016/j.chaos.2021.111370>
- Goswami, R., Dufort, P., Tartaglia, M. C., Green, R. E., Crawley, A., Tator, C. H., Wennberg, R., Mikulis, D. J., Keightley, M., & Davis, K. D. (2016). Frontotemporal correlates of impulsivity and machine learning in retired professional athletes with a history of multiple concussions. *Brain Structure and Function*, 221(4), 1911–1925. <https://doi.org/10.1007/s00429-015-1012-0>
- Gudmundsson, J., & Horton, M. (2017). Spatio-Temporal Analysis of Team Sports. *ACM Computing Surveys*, 50(2), 22:1-22:34. <https://doi.org/10.1145/3054132>
- Hughes, M. D., Caudrelier, T., James, N., Redwood-Brown, A., Donnelly, I., Kirkbride, A., & Dushesne, C. (2012). *Moneyball and soccer—An analysis of the key performance indicators of elite male soccer players by position*. <https://doi.org/10.4100/jhse.2012.72.06>
- Hughes, M., & Probert, G. (2006). A technical analysis of elite male soccer players by position and success. Em *Notational Analysis of Sport-VII* (p. 76–91). UWIC.
- Joseph, A., Fenton, N. E., & Neil, M. (2006). Predicting football results using Bayesian nets and other machine learning techniques. *Knowledge-Based Systems*, 19(7), 544–553. <https://doi.org/10.1016/j.knosys.2006.04.011>
- Konefał, M., Chmura, P., Zacharko, M., Chmura, J., Rokita, A., & Andrzejewski, M. (2018). Match outcome vs match status and frequency of selected technical activities of soccer players during UEFA Euro 2016. *International Journal of Performance Analysis in Sport*, 18(4), 568–581. <https://doi.org/10.1080/24748668.2018.1501991>
- Leoncini, M. P. (2001). *Entendendo o Negócio Futebol: Um estudo sobre a transformação do modelo de gestão estratégica nos clubes de futebol* [Tese de Doutorado]. Universidade de São Paulo.
- Lewis, M. (2003). *Moneyball: The Art of Winning an Unfair Game*.
- Macdonald, R. D. (2011). *The struggle for the body, mind and soul of AFL Footballers* [Book Review] (p. 123). Australian Society for Sports History Publications.
- Meylan, C., Cronin, J., Oliver, J., & Hughes, M. (2010). Talent Identification in Soccer: The Role of Maturity Status on Physical, Physiological and Technical Characteristics. *International Journal of Sports Science & Coaching*, 5(4), 571–592. <https://doi.org/10.1260/1747-9541.5.4.571>

- Muggleton, S. (2014). Alan Turing and the development of Artificial Intelligence. *AI Communications*, 27(1), 3–10. <https://doi.org/10.3233/AIC-130579>
- Paoli, P. B. (2007). *Os estilos de futebol e os processos de seleção e detecção de talentos* [Tese de Doutorado]. Universidade Gama Filho.
- Pappalardo, L., & Cintia, P. (2018). Quantifying the relation between performance and success in soccer. *Advances in Complex Systems*, 21(03n04), 1750014. <https://doi.org/10.1142/S021952591750014X>
- 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. *ACM Transactions on Intelligent Systems and Technology*, 10(5), 59:1-59:27. <https://doi.org/10.1145/3343172>
- Pariath, R., Shah, S., Surve, A., & Mittal, J. (2018). Player Performance Prediction in Football Game. *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 1148–1153. <https://doi.org/10.1109/ICECA.2018.8474750>
- Payyappalli, V. M., & Zhuang, J. (2019). A data-driven integer programming model for soccer clubs' decision making on player transfers. *Environment Systems and Decisions*, 39(4), 466–481. <https://doi.org/10.1007/s10669-019-09721-7>
- Proni, M. W. (1998). *Esporte espetaculo e futebol-empresa* [Tese de Doutorado]. Universidade Estadual de Campinas.
- Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science. *SpringerPlus*, 5(1), 1410. <https://doi.org/10.1186/s40064-016-3108-2>
- Ribeiro, P. (2009). *A observação como ponto de partida para uma análise pormenorizada das características das equipas adversárias*. [Relatório de Estágio]. Universidade do Porto.
- Rossi, A., Pappalardo, L., Cintia, P., Iaia, F. M., Fernàndez, J., & Medina, D. (2018). Effective injury forecasting in soccer with GPS training data and machine learning. *PLOS ONE*, 13(7), e0201264. <https://doi.org/10.1371/journal.pone.0201264>
- Samuel, A. L. (1959). Machine learning. *The Technology Review*, 62(1), 42–45.
- Santos, P. M. R. (2012). *O modus operandi de um Departamento de Scouting de Futebol Estágio Profissionalizante realizado na Futebol Clube do Porto—Futebol, SAD* [Relatório de Estágio]. Universidade do Porto.

- Szymanski, S. (2017). Entry into exit: Insolvency in English professional football. *Scottish Journal of Political Economy*, 64(4), 419–444. <https://doi.org/10.1111/sjpe.12134>
- Turing, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- Upham, S. P., & Small, H. (2010). Emerging research fronts in science and technology: Patterns of new knowledge development. *Scientometrics*, 83(1), 15–38. <https://doi.org/10.1007/s11192-009-0051-9>
- van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
- Williams, A. M., & Ward, P. (2017). Searching for the Holy Grail: Can there ever be such a thing as a “Grand Unified Theory of sports performance”? *Human Movement Science*, 56(Pt A), 181–183. <https://doi.org/10.1016/j.humov.2017.04.012>
- Williams, J. J. (2009, janeiro). An investigation into operational definitions used within performance analysis. *Presentation at the International Workshop in Performance Analysis*.
- Xiang, C., Wang, Y., & Liu, H. (2017). A scientometrics review on nonpoint source pollution research. *Ecological Engineering*, 99, 400–408. <https://doi.org/10.1016/j.ecoleng.2016.11.028>
- Yi, Q., Jia, H., Liu, H., & Gómez, M. Á. (2018). Technical demands of different playing positions in the UEFA Champions League. *International Journal of Performance Analysis in Sport*, 18(6), 926–937. <https://doi.org/10.1080/24748668.2018.1528524>
- Zambom-Ferraresi, F., Rios, V., & Lera-López, F. (2018). Determinants of sport performance in European football: What can we learn from the data? *Decision Support Systems*, 114, 18–28. <https://doi.org/10.1016/j.dss.2018.08.006>
- Zhou, C., Zhang, S., Calvo, A. L., & Cui, Y. (2018). Chinese soccer association super league, 2012–2017: Key performance indicators in balance games. *International Journal of Performance Analysis in Sport*, 18(4), 645–656. <https://doi.org/10.1080/24748668.2018.1509254>