

# **Utilização de algoritmos de inteligência artificial na predição de partidas de basquetebol**

**Marcos Vinicius Fernandes Vital<sup>1</sup>**

**Resumo:** O basquetebol é um dos esportes mais assistidos é praticado no mundo com isso surgiu diversos interesses na predição de usas partidas. Por isso os sistemas computacionais que utiliza a predição passaram a ser usado, para auxiliar na formação de novas estratégias para aqueles que necessitam de dados mais preciso e confiáveis. Este projeto teve por objetivo utilizar algoritmos de inteligência artificial para a predição de partidas de basquete. Atualmente existe diverso algoritmos de predição para com diversas características, com isso nesse projeto ao se utilizar de diversos algoritmos poderá se saber quais os são os mais adequados para a predição de dados nas partidas de basquetebol.

**Palavras-chave:** Basquetebol. Predição. inteligencia artificial.

## **Use of Artificial Intelligence Algorithms in Predicting Basketball Games**

**Abstract:** Basketball and one of the most watched sports is practiced in the world with this has arisen several interests in the prediction of using matches. Therefore the computational systems that use the prediction have been used to assist in the formation of new strategies for those who need more accurate and reliable data. This project aimed to use artificial intelligence algorithms to predict basketball games. Currently there are several prediction algorithms for several characteristics, so in this project when using several algorithms you can know which are the most suitable for predicting data in basketball games.

**Keywords:** Basketball. Prediction. Artificial intelligence.

---

<sup>1</sup>Estudante de Ciência da Computação, IFTM, Campus Ituiutaba, vinicius.pm901@hotmail.com

# 1 INTRODUÇÃO

O basquete é um dos esportes mais populares do mundo, e também, uma esporte de equipe, em que o objetivo é atirar uma bola através de uma cesta posicionada horizontalmente, para marcar pontos, com um conjunto de regras. Normalmente, há duas equipes de cinco jogadores jogando em uma quadra retangular marcado, cada lado com uma cesta. Com sua popularidade surgiu o interesses na predição de suas partidas.

*National Basketball Associativo* (NBA) desde a sua origem tem mais de 60 anos. Durante esta organização, há 30 equipes formadas e divididas em Conferência Leste e Conferência Oeste. Para a temporada regular terá 82 jogos para cada equipe e pós temporada usando um esquema de melhor de sete séries. Portanto, uma estimativa que, haverá pelo menos cerca de 2.300 jogos gerados. Com uma massa de dados e gerada depois de cada jogo da NBA, esses dados existentes nos permitem descobrir dados valioso.

Por isso os sistemas computacionais que utiliza a predição passaram a ser usado, para auxiliar na formação de novas estratégias para aqueles que necessitam de dados mais preciso e confiáveis. A matemática tornaram-se uma parte importante do esporte e muito esforço é dedicado a prever os resultados de eventos esportivo.

Já tendo sido usado em vários processos de tomada de decisão, a predição pode fornecer para os treinadores uma visão sobre o desempenho de suas equipes durante um jogo. Assim podendo simular como uma partida é estudar cenários que podem surgir em diferentes circunstâncias na quadra. Assim ajudará os treinadores a entender como uma equipe pode aumentar suas chances de ganhar, como as habilidades em jogos individuais afetam o desempenho da equipe e qual desempenho pode ser esperado usando diferentes abordagens.

## 2 REFERENCIAL TEÓRICO

A análise computacional é uma maneira objetiva de registrar o desempenho, de modo que os eventos críticos nesse desempenho podem ser quantificados de maneira consistente e confiável. Essa análise permite que o treinador e o gerente avaliem objetivamente o desempenho competitivo e, portanto, melhorem-no (FRANKS, 2004).

A precisão e a velocidade das previsões dependerão da seleção manual ou automática adequada dos recursos mais significativos e altamente correlacionados. Kahn avaliou características primárias e empregou o método sugerido por (PURUCKER, 1996) para selecionar cinco carac-

terísticas finais para predição (KAHN, 2003).

As métricas da Associação de Pesquisadores de Basquete Profissional (ABPR) são semelhantes às da sabermetrics, que é uma das primeiras métricas para avaliar o desempenho dos jogadores de beisebol, mas as métricas da ABPR tentam visualizar as estatísticas em termos de desempenho de equipe e não de desempenho individual. Existem muitos fatores incertos para influenciar o resultado, no entanto, a mineração de dados ainda tem seu próprio valor na previsão do resultado na previsão de resultados de jogos de basquete. (SCHUMAKER; SOLIEMAN; CHEN, 2010)

Bernard, Earl e W (2009) fizeram uma pesquisa sobre a previsão de jogos da NBA usando redes neurais. Autores exploraram subconjuntos obtidos a partir de relações sinal-ruído e opiniões de especialistas para identificar um subconjunto de recursos de entrada para as redes neurais. Os resultados obtidos a partir dessas redes foram comparados com as previsões feitas por vários especialistas no campo do basquete. Após o experimento, o projeto teve 70,33% de precisão.

Embora o treinamento de um Máquina de vetores de suporte (MVS) leve mais tempo comparado a outros métodos, acredita-se que o algoritmo tenha alta precisão devido à sua alta capacidade de construir limites de decisão complexos e não-lineares. Também é menos propenso a *overfitting* (HAN; KAMBER; PEI, 2017).

Cao empregou um MVS, um classificador logístico simples uma combinação de algoritmos cujo núcleo é a regressão logística e usa o *LogitBoost* como uma função de regressão simples (LANDWEHR; HALL; FRANK, 2005), e uma rede neural multicamada para prever resultados de basquete.

Witten et al. (2017), no momento no qual deseja-se estimar o valor de uma variável numérica e os atributos do conjunto de dados também são numéricos, a escolha pela técnica de regressão linear é natural, a mesma vem sendo utilizada por décadas na aplicação de problemas estatísticos, de modo que mesmo quando o conjunto de dados não apresenta uma dependência linear a aplicação do algoritmo serve como um ponto de partida para a utilização de outros algoritmos mais complexos.

## **2.1 OBJETIVO GERAL**

O objetivo deste trabalho é a análise e predição de partidas de basquete utilizando dados das partidas e dos jogadores.

## 2.2 OBJETIVOS ESPECÍFICOS

- comparar e demonstrar a eficácia para os classificadores utilizados no estado da arte de predição de partidas de basquete.
- comparar e demonstrar a eficácia das bases de dados existentes no estado da arte na predição de partidas de basquete.
- comparar e demonstrar a eficácia dos métodos seletores de características utilizados no estado da arte de predição de partidas de basquete.

## 3 DESENVOLVIMENTO

### 3.1 MATERIAIS

Os materiais usados no trabalho foram duas bases de dados ambas da NBA Advanced Stats<sup>1</sup> sendo uma da *season* de 2014 a 2018 com 9.840 jogos com os dados armazenados em um arquivo csv, a outra indo da *season* de 2007 a 2019 com 30.000 jogos com os dados armazenados em um banco de dados e sendo acessado através da nba-api PyPI.

A ferramenta utilizada para o desenvolvimento foi o *JupyterLab*, linguagem de programação python e o uso das bibliotecas *pandas*, *numpy*, *sklearn*, *seaborn*, *matplotlib*.

A NBA advances stats é um site patrocinado pela SAP com o propósito de manter um registro de toda a liga da NBA e facilitar o acesso a essas informações pelas equipes e organizações. A nba-api PyPI é uma API para acesso a [www.nba.com](http://www.nba.com), o principal objetivo é mapear e analisar o maior número possível de jogos.

O *jupyterlab* é um ambiente de desenvolvimento interativo baseado na *web* para *notebooks*. O *jupyterlab* é fácil de configurar e organizar, a interface do usuário suporta uma ampla variedade de fluxos de trabalho em ciência de dados, computação científica e aprendizado de máquina. O *jupyterlab* é extensível e modular e fácil de adicionar os *plug-ins*, que adicionam novos componentes e se integram aos já existentes (JUPYTER, 2019).

Python é uma linguagem de programação criada por Guido van Rossum em 1991. Os objetivos do projeto da linguagem eram produtividade e legibilidade, é uma linguagem de alto nível, multi-paradigma, suporta o paradigma orientado a objetos, imperativo, funcional e procedural. Possui tipagem dinâmica e uma de suas principais características é permitir a fácil

---

<sup>1</sup>National Basketball Associativo

leitura do código e exigir poucas linhas de código se comparado ao mesmo programa em outras linguagens(TECHNOLOGY, 2019).

O pandas é uma biblioteca de código aberto, licenciada por BSD <sup>2</sup>, que fornece estruturas de dados de alto desempenho e fáceis de usar e ferramentas de análise de dados para a linguagem de programação *python*. O pandas ajuda a preencher essa lacuna, permitindo que você execute todo o fluxo de trabalho de análise de dados no *python* sem precisar mudar para uma linguagem(PANDAS, 2019).

O numPy é uma biblioteca *python* que é usada para realizar cálculos em *arrays* multidimensionais. Fornecendo um grande conjunto de funções e operações que ajudam os programadores a executar facilmente cálculos numéricos.(SANTIAGO, 2019)

O *scikit learn* é uma biblioteca *python* que é usada para aprendizado de máquina. Ela possui uma variedade de algoritmos incluindo vários algoritmos de classificação, regressão e agrupamento incluindo máquinas de vetores de suporte, florestas aleatórias, *gradient boosting*, *k-means*(VAROQUAUX, 2013).

O matplotlib é uma biblioteca de plotagem 2D do python, é uma biblioteca que tenta facilitar e facilitar a gerar gráficos, histogramas, espectros de potência, gráficos de barras, gráficos de erros, gráficos de dispersão etc(MATPLOTLIB, 2019).

O seaborn é uma biblioteca de visualização de dados Python baseada no matplotlib . Ele fornece uma interface de alto nível para desenhar gráficos estatísticos atraentes e informativos.(SEABORN, 2019)

## 3.2 METODOLOGIA

Os algoritmos usados para as predições são os de regressão linear, regressão logística, k Nearest Neighbors, árvore de decisão, floresta aleatória, Máquinas de Vetores de Suporte.

O algoritmo de regressão linear responsável por modelar uma associação entre uma ou mais variáveis de saída e entrada. O processo de regressão pode ser dividido em duas categorias, as paramétricas, no qual o relacionamento entre as variáveis é conhecido, e não paramétricas onde não existe conhecimento preexistente entre as variáveis. As técnicas de regressão linear procuram a relação entre duas variáveis por meio de uma equação de uma linha reta.

## 4 CONCLUSÃO

---

<sup>2</sup>Berkeley Software Distribution

## Referências

- BERNARD, L.; EARL, B.; W, B. K. Predicting nba games using neural networks. *Journal of Quantitative Analysis in Sports*, v. 5, n. 1, p. 1–17, 2009.
- FRANKS, I. M. *Notational Analysis of Sport*. Taylor & Francis Ltd, 2004. ISBN 0415290058. Disponível em: [https://www.ebook.de/de/product/3473295/notational\\_analysis\\_of\\_sport.html](https://www.ebook.de/de/product/3473295/notational_analysis_of_sport.html).
- HAN, J.; KAMBER, M.; PEI, J. *Data Mining: Concepts and Techniques*. Elsevier LTD, Oxford, 2017. ISBN 0123814790. Disponível em: [https://www.ebook.de/de/product/14641128/jiawei\\_han\\_micheline\\_kamber\\_jian\\_pei\\_data\\_mining\\_concepts\\_and\\_techniques.html](https://www.ebook.de/de/product/14641128/jiawei_han_micheline_kamber_jian_pei_data_mining_concepts_and_techniques.html).
- JUPYTER, P. *Jupyter*. 2019. <https://jupyter.org/>. [Online; accessed 19-Nov-2019].
- KAHN, J. Neural network prediction of nfl football games. *World Wide Web Electronic Publication*, 01 2003.
- LANDWEHR, N.; HALL, M.; FRANK, E. Logistic model trees. *Machine Learning*, v. 59, n. 1, p. 161–205, May 2005. Disponível em: <https://doi.org/10.1007/s10994-005-0466-3>.
- MATPLOTLIB. *Entendendo a biblioteca matplotlib*. 2019. <https://matplotlib.org/>. [Online; accessed 20-Nov-2019].
- PANDAS. *O projeto dos pandas*. 2019. <https://pandas.pydata.org/>. [Online; accessed 20-Nov-2019].
- PURUCKER, M. Neural network quarterbacking. *IEEE Potentials*, Institute of Electrical and Electronics Engineers (IEEE), v. 15, n. 3, p. 9–15, 1996.
- SANTIAGO, L. *Entendendo a biblioteca NumPy*. 2019. <https://medium.com/ensina-ai/entendendo-a-biblioteca-numpy-4858fde63355>. [Online; accessed 20-Nov-2019].
- SCHUMAKER, R. P.; SOLIEMAN, O. K.; CHEN, H. Sports knowledge management and data mining. *Annual Review of Information Science and Technology*, Wiley, v. 44, n. 1, p. 115–157, 2010.
- SEABORN. *Entendendo a biblioteca Seaborn*. 2019. <https://seaborn.pydata.org/>. [Online; accessed 20-Nov-2019].
- TECHNOLOGY, J. *Sobre o Python*. 2019. <https://www.python.org/about/>. [Online; accessed 19-Nov-2019].
- VAROQUAUX, L. B. e Gilles Louppe e Mathieu Blondel e Fabian Pedregosa e Andreas Mueller e Olivier Grisel e Vlad Niculae e Peter Prettenhofer e Alexandre Gramfort e Jaques Grobler e Robert Layton e Jake VanderPlas e Arnaud Joly e Brian Holt e Ga "e 1. In: *ECML PKDD Oficina : Línguas de dados Mining e Máquina de Aprendizagem*. [S.l.: s.n.], 2013.
- WITTEN, I. et al. *Data Mining*. Elsevier LTD, Oxford, 2017. ISBN 0128042915. Disponível em: [https://www.ebook.de/de/product/26440029/ian\\_witten\\_eibe\\_frank\\_mark\\_a\\_hall\\_christopher\\_j\\_pal\\_data\\_mining.html](https://www.ebook.de/de/product/26440029/ian_witten_eibe_frank_mark_a_hall_christopher_j_pal_data_mining.html).