

Análise dos Arremessos do Kobe Bryant

Gabriel Maron Machado Lima
*Pontifícia Universidade
Católica do Paraná*
Curitiba, Brasil
gabriel.maron@pucpr.edu.br

Gustavo Furini
*Pontifícia Universidade
Católica do Paraná*
Curitiba, Brasil
gustavo.furini@pucpr.edu.br

Henrique Monteiro Fernandes
*Pontifícia Universidade
Católica do Paraná*
Curitiba, Brasil
henrique.conceicao@pucpr.edu.br

Leonardo
*Pontifícia Universidade
Católica do Paraná*
Curitiba, Brasil
friedrich.leonardo@pucpr.edu.br

Lucca Libanori
*Pontifícia Universidade
Católica do Paraná*
Curitiba, Brasil
lucca.libanori@pucpr.edu.br

Vitoria Izabel Mendes Pinto
*Pontifícia Universidade
Católica do Paraná*
Curitiba, Brasil
vitoria.pinto@pucpr.edu.br

Abstract. Neste estudo, realizamos uma análise abrangente dos arremessos de Kobe Bryant na NBA, explorando diversas características de sua carreira, incluindo distância, eficiência, desempenho em playoffs, preferências de zonas na quadra e mudanças de estilo de jogo em diferentes contextos de tempo. Além disso, aplicamos técnicas de aprendizado de máquina, incluindo modelos de Random Forest, Árvore de Decisão e KNN, tanto antes quanto depois de balancear a base de dados. Esta análise revelou insights valiosos sobre o desempenho de Kobe Bryant e a complexidade do basquete profissional. Aprofundamos nossa compreensão sobre como fatores como distância, localização na quadra e tempo influenciaram suas escolhas e eficácia. Essa abordagem multidisciplinar destaca a importância da análise de dados no contexto esportivo e fornece uma visão abrangente para pesquisadores e entusiastas do basquete.

Key Words - Kobe Bryant, KNN, Random Forest, Árvore de Decisão, precisão, recall, análise, balanceamento, SMOTE, acurácia, classificação, insights, basquete.

1 Introdução

Kobe Bryant, uma lenda do basquete, deixou um legado na NBA com seu estilo de jogo único e habilidades excepcionais. Seu sucesso e reputação como um dos maiores jogadores de todos os tempos têm sido objeto de admiração e análise extensiva. Neste estudo, exploramos sua carreira por meio de uma análise abrangente dos arremessos realizados ao longo de sua trajetória na NBA. Investigamos a distribuição espacial de seus arremessos, sua eficiência, suas preferências táticas em diferentes zonas da quadra e como seu desempenho variava em jogos de playoffs e temporada regular.

Este estudo vai além da análise univariada, incorporando análises multivariadas complexas para descobrir insights ocultos. Examinamos como a distância dos arremessos impactava sua precisão, como seu posicionamento na quadra se relacionava com suas escolhas de arremessos e como ele adaptava seu estilo de jogo nos minutos finais das partidas. Essas análises revelam não apenas a maestria de Kobe no basquete, mas também ilustram a complexidade do esporte profissional.

Para aprofundar ainda mais nossa compreensão, aplicamos técnicas de aprendizado de máquina a essa base de dados. Avaliamos o desempenho de modelos de classificação, como Random Forest, Árvore de Decisão e K-Nearest Neighbors, tanto antes quanto depois de equilibrar a base de dados.

Este estudo oferece uma visão abrangente do desempenho de Kobe Bryant na NBA, desde uma análise exploratória dos seus arremessos até a aplicação de técnicas de aprendizado de máquina para aprimorar nossa compreensão. Ao fazê-lo, destacamos a importância da análise de dados no contexto esportivo e sua capacidade de fornecer insights valiosos que vão além das estatísticas tradicionais.

2 Trabalhos Relacionados

Outros estudos têm utilizado técnicas de aprendizado de máquina para prever resultados em diferentes contextos. Por exemplo, em [1] aplicaram técnicas de aprendizado de máquina para prever se Kobe Bryant faria cestas bem-sucedidas em jogos de basquete. O estudo utilizou uma rede neural feedforward para analisar dados de diferentes posições no campo e parâmetros básicos do basquete, como a distância da cesta, a posição do jogador e o tempo restante no jogo. Os resultados mostraram que as cestas de dois pontos foram o elemento mais importante na análise, e que a precisão da previsão aumentou quando foram considerados fatores como a distância da cesta e a posição do jogador.

Além disso, o estudo analisou abordagens ofensivas e defensivas e concluiu que é crucial para vencer um jogo marcar pontos próximos à cesta. O estudo demonstrou a eficácia do uso de técnicas de aprendizado de máquina em análises de dados esportivos e pode ser aplicado em outros contextos para prever resultados com precisão.

Em outro artigo que abrangia mais os arremessos de todos os jogadores da NBA, intitulado "Predicting NBA Shots" por Brett Meehan da Universidade de Stanford [2], o sucesso dos arremessos na NBA foi previsto com precisão variando entre 54% e 68%. Brett usou uma variedade de modelos, como Naive Bayes, Random Forest, Boosting, Regressão Logística e redes neurais, na tentativa de encontrar o modelo mais adequado para prever o sucesso dos arremessos. Seu conjunto de dados foi obtido da NBA por meio de seu sistema de rastreamento SportVu.

Embora o conjunto de dados de Brett fosse ligeiramente diferente em termos de características e significativamente maior que o nosso, ele alegou alcançar uma precisão de até 68% com um modelo XGBoost ajustado. Como tal, estes estudos são relevantes para o presente trabalho, pois ambos utilizam técnicas de aprendizado de máquina para analisar dados e prever resultados em diferentes contextos.

3 Método

Neste estudo, utilizamos uma base de dados abrangente dos arremessos da carreira de Kobe Bryant, obtida no site Kaggle. O conjunto de dados inclui informações essenciais como o tipo de ação do arremesso, distância, período, área, zona, temporada e várias outras características relevantes. O objetivo principal foi explorar profundamente o desempenho de Kobe Bryant e extrair informações significativas relacionadas ao basquete profissional.

Iniciamos nossa análise realizando 20 análises univariadas para examinar a distribuição dos arremessos e identificar tendências notáveis. Esta etapa nos permitiu revelar informações cruciais, como a preferência de Kobe por arremessos de média distância e sua impressionante eficiência, com uma taxa de acerto geral de 44%.

Para aprofundar nossa compreensão, também conduzimos 10 análises multivariadas, explorando as correlações entre diferentes variáveis. Investigamos a relação entre a distância dos arremessos e a eficácia, a preferência de Kobe por zonas específicas da quadra, suas estratégias táticas em diferentes partes do campo e como

ele adaptava seu jogo em momentos cruciais.

Em seguida, aplicamos técnicas avançadas de aprendizado de máquina na base de dados. Testamos três modelos de classificação: Random Forest, Árvore de Decisão e K-Nearest Neighbors (KNN), para prever o sucesso dos arremessos. Essa etapa nos permitiu avaliar o desempenho dos modelos antes e depois do balanceamento da base de dados. Utilizamos o método holdout para treinar os classificadores. As métricas utilizadas para avaliar o desempenho dos modelos foram acurácia, precisão, recall e F1-Score. Essas métricas nos ajudaram a entender melhor o desempenho dos modelos em prever os resultados dos arremessos.

4 Resultados e Discussões

Seguindo a metodologia proposta, o conjunto de dados utilizado consiste em 30.697 linhas e 25 colunas de dados rotulados, abrangendo todos os arremessos da carreira de Kobe Bryant na NBA. Ao realizar análises univariadas, um dos principais insights revelou a notável habilidade de Kobe Bryant em arremessos de média distância, frequentemente a uma distância média de 3-4 metros da cesta. Sua capacidade de criar e converter arremessos difíceis nessa faixa de distância demonstra a maestria que ele possuía no jogo. Ele também mostrou uma preferência notável por duas zonas específicas da quadra, o "Center (C)" no centro e o "Right Side (R)" no lado direito. Essa preferência ressalta sua capacidade de encontrar oportunidades vantajosas nessas áreas e sua precisão ao executar arremessos. Esses são apenas alguns dos insights obtidos nas análises univariadas.

Não nos limitamos apenas a análises univariadas, mas também conduzimos análises multivariadas mais complexas que nos permitiram descobrir insights valiosos. Nossas análises multivariadas revelaram aspectos fundamentais do desempenho de Kobe e ao investigar a relação entre a distância dos arremessos e sua eficácia, identificamos sua capacidade notável em arremessos de média distância, contrastando com sua menor eficácia em arremessos de longa distância.

Além disso, ao correlacionar seu posicionamento na quadra com a escolha de arremessos, percebemos que Kobe estrategicamente preferia o centro da quadra para arremessos de curta distância e explorava as laterais para arremessos de longa distância, evidenciando sua adaptabilidade tática.

Uma descoberta adicional mostrou que, nos momentos finais das partidas, Kobe tinha uma tendência a arriscar arremessos de três pontos, demonstrando sua determinação e liderança em situações cruciais. Essas análises multivariadas proporcionam insights valiosos sobre a interação de fatores para uma análise aprofundada.

Para a aplicação do algoritmos de classificação foram excluídas as linhas que continham valores faltantes na coluna 'shot_made_flag' e algumas colunas que não utilizamos no modelo. também foi transformada as colunas categóricas em colunas numéricas,, assim ficando com 25.697 linhas e 89 colunas.

Neste estudo, avaliamos o desempenho de três modelos de classificação - Random Forest, K-Nearest Neighbors (KNN), e Árvore de Decisão - antes e depois do balanceamento da base de dados. A seguir, apresentamos os principais resultados em forma de tabelas e discutimos as conclusões.

Tabela 1: Resultado dos Classificadores Não Balanceados

Modelo de Classificação	Acurácia	Precisão	Recall	F1-Score
Random Forest	64%	62%	49%	54%
KNN	55%	49%	44%	46%
Árvore de Decisão	58%	52%	53%	52%

No cenário não balanceado, o modelo Random Forest demonstrou uma acurácia mais

elevada, o que significa que, em termos gerais, o modelo era capaz de fazer previsões corretas em uma maior porcentagem de casos. No entanto, ao observarmos as métricas de precisão e recall, notamos um desequilíbrio entre essas duas métricas. Nesse contexto, uma previsão positiva seria classificar um arremesso como bem-sucedido. Portanto, uma alta precisão indica que o modelo faz poucas previsões falsas positivas, ou seja, classifica arremessos como bem-sucedidos com precisão, por outro lado, o recall mede a proporção de casos verdadeiramente positivos (arremessos bem-sucedidos) que o modelo foi capaz de identificar corretamente. Portanto, uma alta pontuação de recall indica que o modelo é capaz de identificar a maioria dos arremessos bem-sucedidos.

O desequilíbrio entre precisão e recall em modelos não balanceados é uma questão comum. No contexto deste estudo, o modelo Random Forest obteve uma alta precisão, o que significa que estava fazendo previsões bem-sucedidas com precisão.

Na Tabela 2 é apresentado os valores após o balanceamento usando SMOTE:

Tabela 2: Resultado dos Classificadores Balanceados com SMOTE

Modelo de Classificação	Acurácia	Precisão	Recall	F1-Score
Random Forest	68%	71%	61%	65%
KNN	58%	58%	59%	59%
Árvore de Decisão	61%	61%	60%	61%

Após o balanceamento da base, o Random Forest manteve sua liderança em acurácia e obteve a melhor precisão e F1-Score, sugerindo sua eficácia na previsão de arremessos de Kobe Bryant. Além disso, o balanceamento resultou em modelos mais equilibrados em termos de precisão e recall.

Os resultados refletem a importância do balanceamento da base de dados para melhorar o desempenho dos modelos. Após a utilização do método SMOTE, todos os modelos demonstraram maior equilíbrio entre as métricas de precisão e recall, tornando suas previsões mais confiáveis.

5 Conclusão

Este estudo transcende o mero exame do desempenho de Kobe Bryant no contexto do basquete profissional. Ele destaca o impacto abrangente que a análise de dados e as técnicas de aprendizado de máquina podem ter no campo dos esportes. A aplicação destas ferramentas proporciona uma visão mais profunda e valiosa das nuances do desempenho esportivo, muitas vezes não perceptíveis a olho nu.

As análises univariadas permitiram uma compreensão detalhada dos padrões de arremessos de Kobe, evidenciando sua habilidade excepcional em arremessos de média distância. As análises multivariadas, por outro lado, revelaram como fatores complexos, como posição na quadra e tempo de jogo, interagem para moldar as estratégias táticas do jogador. Essas análises juntas destacam a interconexão de diversas variáveis e demonstram como a análise de dados pode melhorar nossa compreensão do jogo. O verdadeiro potencial dessas técnicas se estende além do basquete e pode ser aplicado a uma variedade de esportes. Os insights gerados não apenas beneficiam atletas individuais, treinadores e equipes, mas também têm o poder de influenciar as estratégias de jogo em esportes coletivos.

À medida que celebramos o legado de Kobe Bryant e sua incrível carreira, também celebramos o progresso contínuo da análise de dados no esporte.

Referências

- [1] Shahzad, Taimur & Rehman, Zahoor & Batool, Eliza & Ullah, Zain. (2021). Kobe Bryant Shot Prediction using Machine Learning. Vol 2 No 1 (2021). 10-20. 10.33897/fujeas.v2i1.420.
- [2] Meehan, Brett. Predicting NBA Shots, 2017, cs229.stanford.edu/proj2017/final-reports/5132133.pdf