

# PREVENDO A PREMIAÇÃO DE MVP DA TEMPORADA REGULAR DA NBA

Cristopher Braatz Cardoso, Leandro Vilson Battisti e Lucas Amaral

FURB – Pós Graduação em Data Science  
10/10/2020

## 1. INTRODUÇÃO

A National Basketball Association (NBA) elege desde 1956 o jogador mais valioso todo ano (MVP). Para a mídia esportiva, é o principal prêmio individual de basquetebol na temporada. Desde a temporada 1980-1981 o prêmio é decidido por um grupo de jornalistas esportivos dos Estados Unidos e do Canadá. Cada jornalista escolhe os 5 jogadores mais valiosos da temporada, sendo que cada posição possui um peso na pontuação. O jogador com a maior pontuação vence o prêmio.

Várias estatísticas são utilizadas pelos jornalistas para chegar a uma votação final. Desde as mais simples, como pontos por jogo (PPG), até mais complexas como classificação de eficiência do jogador (PER). Além de estatísticas individuais, o desempenho do time é considerado na premiação, incluindo dados como classificação do time na temporada e contribuição individual do jogador nas vitórias do time.

Nas próximas seções, serão mostrados experimentos realizados para determinar as relações entre as variáveis obtidas com a premiação de MVP. No final será feito uma simulação para o resultado da temporada atual (2019-2020).

## 2. OBJETIVOS

A predição do ganhador do prêmio de MVP da temporada da NBA pode ser tratada tanto como um problema de classificação, quanto de regressão. A NBA calcula uma pontuação, chamada de award share, através da representatividade de votos recebida pelo jogador, em relação ao total de votos.

Neste estudo, foi escolhido prever se o jogador ganhou o prêmio, ao invés de prever o award share. Em resumo, foram utilizados métodos de classificação. Os métodos de classificação aplicados retornam uma probabilidade e com esses valores foi feito um ranqueamento por temporada, para determinar se o jogador venceu ou não o prêmio de MVP.

## 3. ANÁLISE E PREPARAÇÃO DOS DADOS

Seguindo o padrão de outros esportes americanos, a NBA possui muitas estatísticas para avaliar a performance dos times e jogadores. Portanto, existem sites como o Basketball Reference [1] que disponibilizam as estatísticas dos jogadores e times.

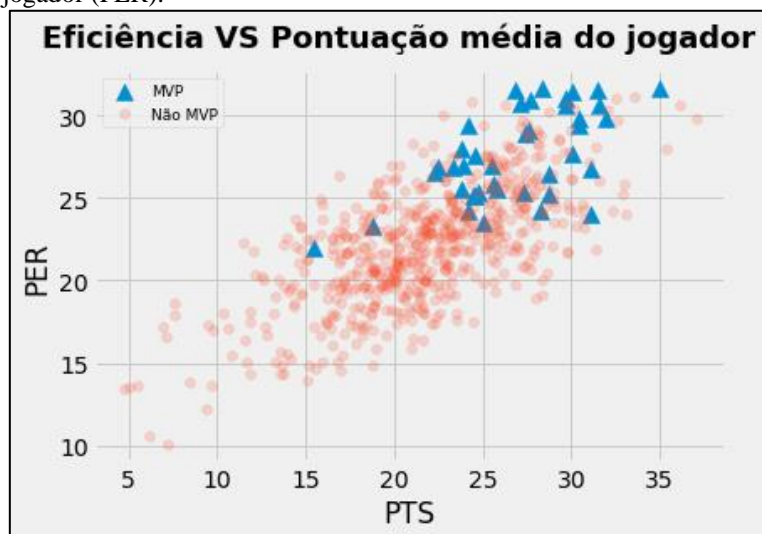
Para elaboração dos datasets de treino e produção, foi feito um scrapper que consulta as informações no site do Basketball Reference. Foi necessário remover algumas linhas do dataset de treino, para não considerar todos os jogadores de cada temporada. Foram mantidos apenas os jogadores que receberam pelo menos um voto em qualquer posição para MVP. Já no dataset de produção, foram removidos os jogadores que jogaram menos de 55 partidas na temporada que é de 82 jogos, ou seja, 2/3.

Analisando o dataset, foi possível identificar as correlações entre as variáveis independentes e a variável objetivo (Won MVP). A tabela abaixo descreve as principais variáveis e o percentual de correlação com a variável Won MVP.

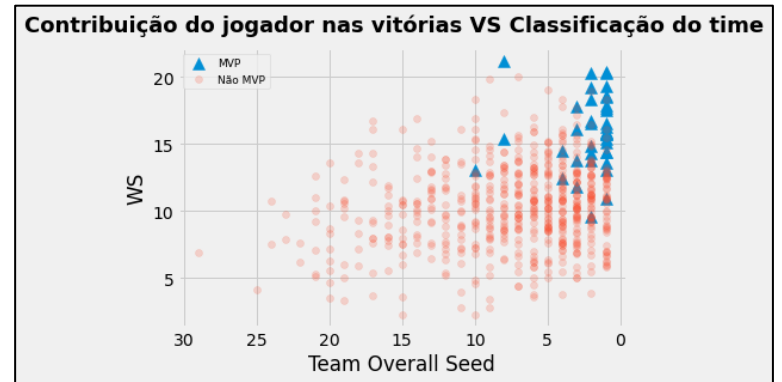
Variável	Descrição	% Correlação Won MVP
WS	Win Shares. Índice de impacto do jogador nas vitórias da equipe.	39%
VORP	Value over replacement player. Diferença da contribuição do jogador para o time em relação com um possível jogador substituto.	39%
BPM	Box Plus/Minus. Índice de contribuição em quadra.	37%
PER	Player Efficiency Rating. Índice de eficiência geral do jogador.	34%
Team Overall Seed	Classificação do time do jogador.	23%
PTS	Points. Média de pontos por jogo do jogador.	23%

As análises feitas para determinar o modelo desse trabalho envolvem basicamente as variáveis descritas acima. Além delas representarem o desempenho em quadra dos atletas, também são considerados desempenho da equipe na temporada regular.

Começando pelas análises de desempenho em quadra, é apresentado o gráfico a seguir, relacionando a quantidade média de pontos do jogador por jogo (PTS) e o índice de eficiência do geral do jogador (PER).



Os MVPs, em sua maioria, possuem um índice de eficiência geral alto e uma alta média de pontos por jogo. Isso também reflete na participação do jogador nas vitórias do time. No gráfico a seguir, é relacionado o índice de impacto do jogador nas vitórias do time (WS) com a classificação da equipe na temporada (Team Overall Seed).



O gráfico anterior demonstra como o prêmio de MVP está relacionado com a performance coletiva das equipes. Em apenas dois casos, a classificação da equipe do MVP ficou abaixo da 5ª posição na temporada. Nesses casos outras estatísticas foram mais relevantes, como a média de 35 pontos de Michael Jordan na temporada 1987-1988 e o recorde de triplo-duplos em uma temporada quebrada por Russell Westbrook em 2017-2018.

## 4. MODELAGEM

Para construção do modelo, foi utilizado o algoritmo de regressão logística utilizando cross-validation. As variáveis apresentadas na seção anterior foram utilizadas para treinar o modelo, em conjunto com outras variáveis presentes no dataset.

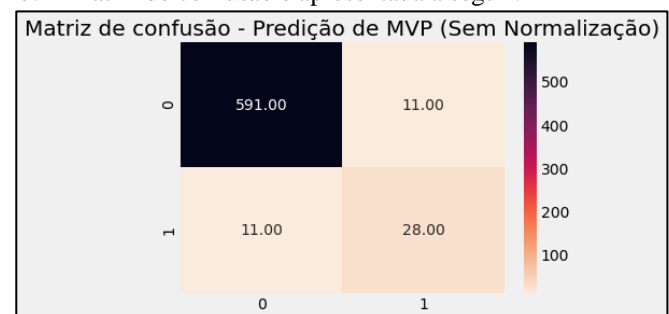
Para treino do modelo utilizando cross-validation, foi utilizado um algoritmo que itera por todos os anos presentes no dataset (1980 até 2019) e para cada ano, ele gera um dataset de treino e um de teste. O dataset de treino são todos os registros diferentes do ano corrente. Enquanto o dataset de teste são todos os registros do ano corrente. Então por exemplo, considerando o ano de 2010, o dataset de treino utiliza todos os anos menos o ano 2010 e o dataset de teste utiliza apenas o ano de 2010.

Após dividir os datasets, o próximo passo é treinar o modelo. É armazenado a probabilidade de o jogador vencer o prêmio e no final é feito um ranqueamento por temporada, do maior ao menor. Portanto, o jogador com maior probabilidade, é ranqueado em primeiro e consequentemente fica com o prêmio de MVP da temporada.

O modelo aplicado obteve os seguintes resultados:

	Treino	Teste
Acurácia	97%	96%
Precisão	76%	72%
Cobertura	76%	72%
Medida F	76%	72%

Com uma precisão de 72% no dataset de teste, foi possível prever com sucesso a nomeação de MVP de 28 jogadores. Mas o modelo também previu errado que 11 jogadores ganharam o prêmio. A matriz de confusão é apresentada a seguir.



Com o modelo criado e treinado, o próximo passo é utilizá-lo em produção. Os resultados previstos para a temporada 2019-2020 do prêmio MVP são apresentados na tabela abaixo.

Jogador	Probabilidade
Giannis Antetokounmpo	26%
James Harden	19%
LeBron James	14%
Anthony Davis	9%
Kawhi Leonard	7%

Como o prêmio de MVP da temporada 2019-2020 já foi entregue, é possível comparar o predito com o valor real. O modelo acertou o vencedor e chegou muito próximo na classificação do MVP.

Jogador	Votos 1º Lugar	Pontos ganhos	Award Share
Giannis Antetokounmpo	85	962	0.952
Lebron James	16	753	0.746
James Harden	0	367	0.363
Luka Dončić	0	200	0.198
Kawhi Leonard	0	168	0.166

## 5. CONCLUSÃO

Ao final desse estudo, pode-se concluir que a maioria dos MVPs eleitos possuem características semelhantes, como alta eficiência em quadra, alta contribuição para as vitórias do time e estão nos melhores times das temporadas. Mas em alguns anos, quebras de recordes, influência da mídia e outros fatores mais complexos que estatísticas de jogo influenciaram muito nos resultados.

Também se demonstrou um problema complexo de ser trabalho com aprendizado supervisionado, devido a pequena quantidade de dados e desbalanceamento da variável MVP won. Sendo poucos jogadores indicados ao prêmio, considerando em média 10 jogadores de 500 por ano, com apenas um ganhador.

Como extensão para esse estudo, poderia ser aplicado um modelo semelhante para determinar outros prêmios da NBA, como novato do ano (ROY). Outro exemplo de aplicação poderia ser no estudo de comportamentos de casas de apostas de esportes.

## 6. REFERÊNCIAS

[1] Sports Reference (org.). Basketball Reference. 2020. Disponível em: <https://www.basketball-reference.com/>. Acesso em: 06 out. 2020.