



Universidade de Brasília - UnB  
Instituto de Ciências Exatas - IE  
Departamento de Estatística - EST

# **Trabalho de Conclusão de Curso 1**

## **Predição de Resultados da NBA na temporada 2018-19**

Gustavo Pompeu da Silva

Orientador: Professor Eduardo Monteiro de Castro Gomes

Brasília

2018



## Introdução e Justificativa

Mineração de dados em esportes é um tópico que tem crescido rapidamente nos últimos anos. Jogadores de ligas de *fantasy* e entusiastas de esportes estão cada vez mais interessados em procurar uma vantagem nas apostas e previsões através de dados e números. Ferramentas e técnicas começaram a ser desenvolvidas para medir desempenho tanto de times quanto de atletas, e esses métodos vem chamando a atenção de grandes franquias esportivas.

Existe uma imensa quantidade de dados disponíveis sobre qualquer esporte. Esses dados podem ser de desempenho individual de jogadores ou da equipe, decisões da comissão técnica, eventos que acontecem nos jogos, entre outros. O problema não é como coletar esses dados, mas sim saber quais dados podem ser úteis e como fazer o melhor uso possível deles. Achando os meios para transformar esses dados em conhecimento, organizações esportivas tem o potencial de obter uma vantagem competitiva sobre seus oponentes. Não devemos analisar performance no sentido de marcar mais gols ou pontos do que o oponente, pois esse é o objetivo geral de qualquer esporte, o que é interessante é encontrar padrões em outras estatísticas que mostram tendências justamente para chegar às vitórias.

*Data Mining* envolve procedimentos para descobrir padrões escondidos e descobrir novas informações a partir de fontes de dados. A fundação científica de data mining pode ser dividida em três disciplinas: estatísticas, inteligência artificial e machine learning. *Data mining* então pode ser definido como a busca de conhecimento dentro dos dados. (SCHUMAKER; SOLIEMAN; CHEN, 2010)

A NBA (National Basketball Association) é a principal liga de basquete profissional do mundo. Atualmente, é composta por 30 times baseados em cidades da América do Norte (29 nos Estados Unidos e 1 no Canadá). É a liga onde jogam os melhores atletas de basquete do mundo, e com os maiores salários do esporte. Uma das vantagens de trabalharmos com o basquete e a NBA especificamente é a grande quantidade de dados, pois em uma temporada, cada time joga 82 jogos, ou seja, são 1230 jogos por temporada, isso nos permite ter muitas observações para trabalhar.

## Objetivos

Utilizar diversas técnicas estatísticas para prever os resultados da maioria dos jogos da temporada de 2018-19 da NBA utilizando e chegar em uma conclusão sobre qual é a melhor técnica para esse problema em específico, através de medidas como acurácia.

## Objetivo Geral

Utilizar métodos como regressão linear, regressão logística, entre outros, para prever resultados e obter probabilidades de vitória para os times em cada um dos jogos da temporada 2018-19 da NBA.

## Objetivos Específicos

- Ler os dados de resultados dos jogos das temporadas anteriores automaticamente com o R
- Utilizar os métodos acima descritos em parte dos dados para ajustar modelos de previsões
- Validar esses modelos com outra parte dos dados, para então utilizá-los efetivamente durante a temporada 2018-19
- Comparar os métodos e descobrir qual o melhor para essas previsões especificamente

## Metodologia

A linguagem R (R Core Team, 2018) será utilizada em todo o trabalho. Com o auxílio do pacote *rvest* (WICKHAM, 2016), serão extraídos os resultados dos jogos das temporadas anteriores direto da internet, no site Basketball Reference (BASKETBALL...), que é um dos maiores sites com dados estatísticos sobre a NBA e basquete em geral. A partir desses dados, podemos criar inúmeras variáveis, e então criar modelos e métodos de previsão. Durante a temporada de 2018-19, os dados dos jogos também irão sendo extraídos direto da internet.

A ideia é, utilizar vitória ou saldo de pontos na partida como variável resposta para os modelos, e selecionar as variáveis explicativas que são significativas.

Então criar modelos com métodos como regressão linear, regressão logística, máquina de vetores de suporte, entre outros, e utilizar os resultados de múltiplas temporadas em anos recentes para ajustar esses modelos, e posteriormente utilizar nas previsões para a temporada de 2018-19 que começará em outubro de 2018.

## Resultados Preliminares

Apenas com as informações de placar, data, público, e se o jogo foi para a prorrogação ou não, foram criadas mais de 50 variáveis diferentes. Entre elas temos, médias, mínimos e máximos de pontos marcados e sofridos, separados por jogos fora e dentro de

casa, tanto para o time mandante quanto para o visitante, além de quantos dias atrás foi o último jogo, quantidade de jogos totais e em casa, quantidade de vitórias totais e em casa, e o *strength of schedule*, que é uma medida para saber se os adversários do time até então foram mais fortes ou mais fracos.

Na temporada de 2017-18, os times com mais vitórias na temporada regular foram o Houston Rockets (65), Toronto Raptors (59) e Golden State Warriors (58). Os piores times foram o Phoenix Suns com 21 vitórias, o Memphis Grizzlies com 22 e o Atlanta Hawks com 24. O time com mais pontos marcados em média durante a temporada foi o Golden State Warriors, com 113.5 pontos por jogo, e o com a menor média foi o Sacramento Kings, com 98.8 pontos por jogo. A melhor defesa, ou seja, o time que levou menos pontos em média, foi o San Antonio Spurs, sofrendo apenas 99.8 pontos por jogo, enquanto que o pior time nesse quesito foi o Phoenix Suns, sofrendo 113.3 pontos por partida. O máximo de pontos marcados por um time em um jogo foram 149, pelo Miami Heat, em um jogo de duas prorrogações contra o Denver Nuggets no dia 19/03/2018. Já o mínimo de pontos foi 69, em duas ocasiões, pelo Chicago Bulls contra o Oklahoma City Thunder em 28/10/2017 e pelo Washington Wizards contra o Utah Jazz em 04/12/2017. Já nos *playoffs*, o Golden State Warriors se deu melhor, vencendo o Cleveland Cavaliers na *NBA Finals* e conquistando o seu sexto título na história, e o terceiro nos últimos 4 anos. (NBA. . . , )

Em uma análise inicial, apenas com os resultados da temporada anterior, de 2017-18, e fazendo uma regressão linear inicial com algumas variáveis que foram selecionadas testando o nível de significância, ajustando o modelo com os primeiros 1000 jogos, e fazendo as previsões para os últimos 230, foi obtida uma acurácia de quase 68% nas previsões. Ou seja, com uma seleção de variáveis melhor e dados de mais temporadas, além de utilização de outros tipos de modelos, há uma grande chance desse resultado preliminar aumentar bastante.

## Cronograma

Quadro 1 – Cronograma - 2/2018

Atividades	Jul	Ago	Set	Out	Nov	Dez
Escolha do tema a ser abordado						
Estudo do tema						
Estudo de técnicas						
Elaboração e entrega da proposta de Projeto Final						
Apresentação da proposta do projeto						
Elaboração de relatório parcial						
Entrega e correção do relatório parcial						

Quadro 2 – Cronograma - 1/2019

<b>Atividades</b>	<b>Fev</b>	<b>Mar</b>	<b>Abr</b>	<b>Mai</b>	<b>Jun</b>	<b>Jul</b>
Desenvolvimento dos modelos em estudo						
Análise dos Resultados						
Elaboração do relatório Final						
Entrega do relatório final ao prof. orientador						
Correção do relatório final						
Apresentação do relatório final						
Entrega do relatório final à banca						

# Referências

BASKETBALL Reference. <<https://www.basketball-reference.com/>>. Acessado em: 28/09/2018. 2

NBA Stats. <<https://stats.nba.com/>>. Acessado em: 28/09/2018. 3

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2018. Disponível em: <<https://www.R-project.org/>>. 2

SCHUMAKER, R. P.; SOLIEMAN, O. K.; CHEN, H. *Sports Data Mining*. 1st. ed. [S.l.]: Springer Publishing Company, Incorporated, 2010. ISBN 144196729X, 9781441967299. 1

UUDMAE, J. Predicting nba game outcomes. Acessado em: 28/09/2018. Disponível em: <<http://cs229.stanford.edu/proj2017/final-reports/5231214.pdf>>.

WICKHAM, H. *rvest: Easily Harvest (Scrape) Web Pages*. [S.l.], 2016. R package version 0.3.2. Disponível em: <<https://CRAN.R-project.org/package=rvest>>. 2