

Análise de desempenho de jogadores da NFL para auxílio em contratações

Luan Carlos Klein
luanklein@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná
Curitiba, Paraná - Brasil

Abstract

O presente trabalho tem por objetivo analisar os dados referentes aos atletas da *National League Football* (NFL) ao longo das temporadas, visando fornecer informações úteis para os times avaliarem se a contratação de um determinado jogador é boa e se esse jogador tende a ter um bom desempenho na próxima temporada, além de auxiliar no tempo de contrato a ser oferecido. Para tal, tomou-se como foco apenas os jogadores da posição de *quarterback*. Foram realizadas algumas análises gerais, como a média de pontuação em cada faixa de idade, e análises mais específicas do jogador, como identificação de temporadas com desempenho fora do comum (*outliers*) e predição de qual será o desempenho nas próximas duas temporadas.

Keywords: NFL, Analytics, Predição de desempenho

1 Introdução

O processo de contratação de jogadores em todos os esportes é um dos momentos que exige maior atenção dos dirigentes de um clube esportivo. Visando suprir carências em seus times, os dirigentes buscam jogadores com características que possam suprir tais deficiências. Entretanto, diversos fatores auxiliam a determinar se um jogador é uma boa contratação ou não, variando desde a idade do mesmo, comportamentos extra-esportivos e características físicas e psicológicas do próprio jogador. Nesse trabalho, o esporte selecionado para estudo foi o futebol americano, com foco na liga principal, a NFL (National League Soccer)¹.

Dado esse amplo leque de variáveis, o presente trabalho tem por objetivo fornecer uma série de informações relevantes que podem auxiliar na tomada de decisão de se um jogador deve ser contratado ou não. Para tais análises e geração de resultados, o desempenho de cada jogador foi reduzido a uma série de variáveis (como por exemplo, quantidade de *touchdowns* marcados durante uma partida), e a cada uma dessas variáveis foram atribuídas pontuações, podendo assim gerar uma pontuação geral de desempenho a cada jogador.

As pontuações foram atribuídas de acordo com o jogo *Fantasy Football*², que consiste no principal jogo de pontuação e desempenho de jogadores, onde o objetivo é, em cada rodada do campeonato, escalar os jogadores que se acredita que obterão as maiores pontuações.

2 Descrição dos Dados Estudados

Para a construção desse trabalho, foram utilizados dados que foram disponibilizados no Kaggle³. Esses dados compreendem um longo período de tempo, abrangendo desde a primeira temporada oficial até o ano de 2016. Dentro desses dados há diversas informações referentes aos jogadores (como por exemplo idade, peso, local de nascimento, universidade que estudou), informações do desempenho do jogadores nas partidas (e dentro desse tópico, separam-se os jogadores por suas posições, pois a depender da posição, as ações realizadas durante o jogo serão distintas, e consequentemente, terão pontuações diferentes) e dados do time em geral, como os resultados das partidas.

Para dar mais foco ao trabalho, escolheu-se uma única posição para o estudo, que foi a de *quarterback*. Ele é considerado o jogador mais importante do jogo, pois, praticamente todas as jogadas passam pelas mãos dele. Dessa maneira, com o foco limitado, os dados que foram efetivamente utilizados nesse trabalho são os dados pessoais dos jogadores e o histórico de desempenho ao longo das partidas dos *quarterback*.

3 Metodologia

Para o desenvolvimento do trabalho, a primeira ação foi a transformação dos dados dos jogos para a pontuação. Para tal, as pontuações consideradas foram as presentes na *Fantasy Football*⁴. Para exemplificar esse processo, é possível utilizar as jardas passadas (ou seja, a quantidade de jardas que a bola percorreu no ar após um passe, que ao final foi bem sucedido) e o *touchdown* corrido. No caso de jardas passadas, cada 25 jardas representam 1 ponto, e um *touchdown* corrido vale 6 pontos. Por exemplo, se em uma partida o jogador passar 40

¹https://pt.wikipedia.org/wiki/National_Football_League

²<https://fantasy.nfl.com/>

³<https://www.kaggle.com/datasets/kendallgillies/nflstatistics>

⁴<https://www.golimsports.com.br/post/fantasy-football-o-que-%C3%A9-e-como-funciona-%3Atext=O%20restante%20das%20pontua%C3%A7%C3%B5es%20%C3%A9,Fumble%20perdido%20que%20ele%20sofreu.>

jardas e realizar 2 *touchdown*, ele irá obter 22 pontos ($40 \times 0.25 + 6 \times 2$).

Os dados que foram utilizados para considerar a pontuação são: *touchdown* passados (4 pontos), jardas passadas (1 ponto a cada 25), *fumbles* (-2 pontos), interceptações (-2 pontos) e *touchdown* corridos (6 pontos). Após isso, uma nova *feature* é criada, que é a soma de todas essas outras, que é a chamada pontuação total. Essa última variável, que é uma unificação do desempenho, é a variável que buscará ser prevista.

Após esse processo, realizou a unificação dos dados, calculando a média de pontuação por partida em cada uma das *features* por temporada.

A primeira análise a ser feita após essa unificação foi a análise de *outliers*. Através dessa, é possível observar quando um jogador atuou fora da "normalidade" (ou seja, ele pode ter tido um desempenho muito bom ou muito ruim em uma temporada). Essa comparação pode ser feita comparada com todos os jogadores ou um grupo específico (podendo aqui realizar uma filtragem em um período específico), ou realizar uma comparação com apenas as temporadas do próprio jogador (ou seja, vendo se em alguma temporada o jogador, comparado ao desempenho normal dele mesmo, teve um desempenho fora do comum).

Além disso, outras análises são geradas, como por exemplo a média de pontuação por idade (e o detalhamento dessa pontuação), e as médias também (por temporada) de acordo com o local de nascimento do jogador e da universidade que ele veio. Também foi gerada a média de pontuação (por ano) dos jogadores que vieram da mesma universidade que o atleta em específico (por exemplo, a média de cada temporada dos jogadores provenientes da universidade X). Por fim, utilizando esses dados, é treinado um modelo que busca prever (de acordo com o histórico dos jogadores e os demais dados gerados) a pontuação geral média por partida nas próxima temporada.

4 Resultados

O primeiro resultado a ser analisado é a média histórica das pontuações por idade dos jogadores ao longo de toda a história da competição. O gráfico dessa evolução é apresentado na Figura 1. Nesse gráfico, é possível observar que ao passar da idade, a pontuação média por partida tende a aumentar até um certo ponto, que são os 40 anos. Isso ocorre pois pouquíssimos jogadores ao longo da história jogaram acima dessa idade. Até 2018, apenas 15 jogadores haviam transpassado essa "barreira". Após essa idade, as médias são devidos aos poucos jogadores, o que pode causar distorções nos resultados (dado que esses jogadores podem jogar apenas partidas pontuais). Entretanto, é possível observar que jogadores jovens (perto dos 20 anos) tendem a correr mais. Outro ponto de destaque são as interceptações e *fumbles* sofridos. Essa média se mantém praticamente constante praticamente dos 21 aos 40 anos. Por fim, observa-se que a

pontuação total é elevada pois tanto as jardas passadas com os *touchdowns* passados aumentam ao longo da idade. Essas informações são bastante valiosas no momento da definição do tempo de contrato de um jogador. Por exemplo, para jogadores mais novos, terão a tendência de melhorar suas pontuações ao passar do tempo, porém ainda continuarão sofrendo, em média, a mesma quantidade de interceptações e *fumbles*. Já para jogadores mais velhos, que se apresentem perto dos 40 anos, propor contratos longos poderá ser um problema, devido ao fato de que, historicamente, poucos foram capazes de manter um bom nível de jogo com a idade elevada.

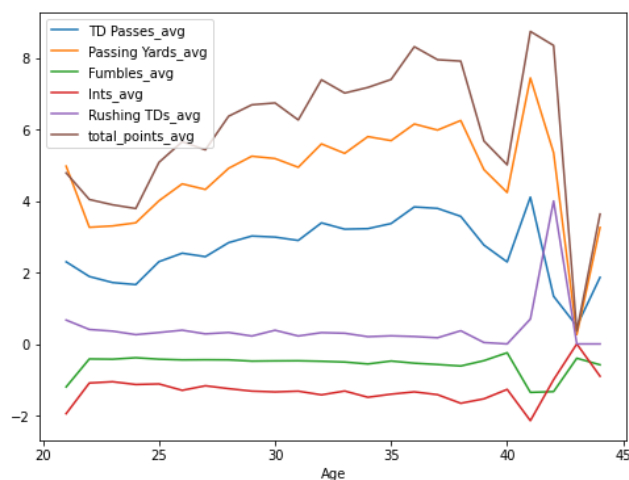


Figure 1. Evolução das médias das pontuações ao longo das idades

Outro resultado válido que auxilia na tomada de decisão para a contratação de jogadores a sua origem do jogador (ou seja, onde ele nasceu). Isso é considerado, pois, o lugar onde esses jogadores nasceram pode influenciar a maneira que aprenderam a jogar, e portanto, seu comportamento dentro de campo. No gráfico da Figura 2 observa-se que jogadores do estado do Mississippi apresentam uma média superior de mais de 3 pontos em relação a jogadores de Illinois. Ressalta-se que foram consideradas aqui apenas estados que tiveram pelo menos 50 jogadores ao longo da história.

Esses dois primeiros resultados se referem a dados gerais, de todos os jogadores. Os próximos resultados irão se referir ao desempenho exclusivo de um jogador apenas. O primeiro é a análise de *outliers* ao longo das temporadas do jogador. Um exemplo de demonstração desse resultado é com o atleta Tom Brady, que é considerado por muitos, o maior jogador da história da NFL. Ele iniciou sua carreira no ano de 2000 e continua em atividade até hoje. Ao longo desse tempo, duas temporadas foram identificadas como sendo de *outliers*. A primeira delas é em 2000, quando ele ainda era calouro na NFL. Ou seja, um resultado esperado, dado que a transição

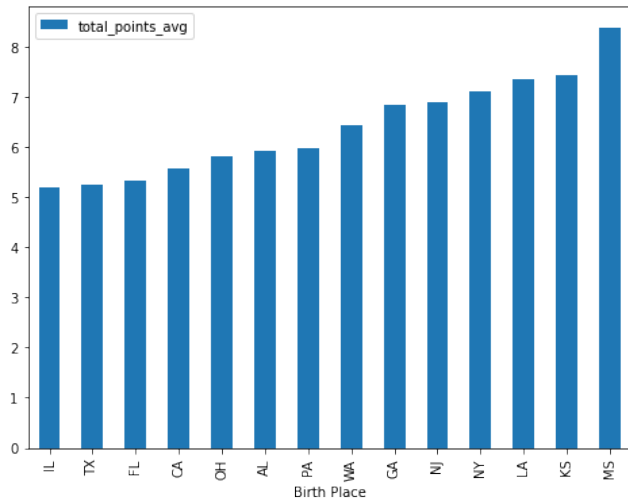


Figure 2. Média de pontuação do local de nascimento dos jogadores

entre as ligas universitárias e as profissionais apresentam grandes diferenças na forma de jogo e na qualidade técnica. A segunda temporada considerada *outlier* foi em 2008, quando ele sofreu uma grave lesão que o tirou dos gramados pela maior parte da temporada. Portanto, diante disso, é possível assumir que em condições normais (sem lesões), esse jogador tende a ter um desempenho parecido com sua média histórica. Isso é especialmente útil para analisar casos em que um jogador apresentou um desempenho espetacular em uma temporada, porém que não corresponde ao padrão dele, indicando que ele pode ter tido apenas uma "fase boa", e portanto, a contratação dele possivelmente não trará os benefícios que se espera.

Por fim, o último resultado a ser apresentado é a predição de pontuação para as próximas duas temporadas do jogador. Utilizando os dados explicados na metodologia, e utilizando uma regressão linear múltipla, construí-se um modelo (ou seja, para cada jogador é gerado um novo modelo) para a predição. Um exemplo dessa predição é novamente com o atleta Tom Brady. Utilizando os dados de 2004 a 2014, previu-se que a sua pontuação em 2015 seria 16,60 pontos e em 2016 de 17,34, enquanto que na prática o que ocorreu foi 13,38 em 2015 e 16,48 em 2018. Apesar de apresentarem diferenças, que foram inferiores a 20%, os resultados se mostraram satisfatórios, já que tem por função fornecer uma noção de qual será o desempenho daquele jogador na próxima temporada. Fazendo uma análise de todos os jogadores, a média do erro quadrático foi de 3,18 pontos por jogo, ou seja, inferior a 1 *touchdown* por partida.

5 Conclusão

Diante dos resultados é possível concluir que o trabalho conseguiu apresentar uma visão geral de qual será o desempenho

de um jogador na próxima temporada. Observa-se que as técnicas utilizadas, apesar da simplicidade, foram eficientes e forneceram resultados satisfatórios. Esse fato ressalta a necessidade de não apenas utilizar técnicas, mas sim aplicá-las em cenários compatíveis. Além disso, através desse trabalho, observou-se a necessidade de ter um escopo de trabalho e um objetivo bem definido, dado ao grande número de possibilidades que se abrem ao longo das análises, principalmente em temas tão amplos como o desse trabalho. Por fim, como trabalho futuro, é possível expandir os dados utilizados para a predição, incluindo a qualidade do time (ou seja, dos demais jogadores) como um dos fatores que afeta o desempenho individual. Além disso, outro passo é analisar as outras posições (considerando os jogadores já presentes no time), e através de uma análise de proximidade entre eles, indicar jogadores que teriam sinergia com os jogadores já presentes no time.