



Universidade de Brasília - UnB

Faculdade de Economia, Administração, Contabilidade e Gestão de Pol. Públicas - FACE

Departamento de Administração - ADM

INTELIGÊNCIA ARTIFICIAL APLICADA À PREVISÃO DE JOGOS DE FUTEBOL

LAURA DEFRANCO FERREIRA PECONICK

Brasília

2018

INTELIGÊNCIA ARTIFICIAL APLICADA À PREVISÃO DE JOGOS DE FUTEBOL

Monografia apresentada ao Departamento de Administração - ADM como requisito parcial à obtenção do título de Bacharel em Administração.

Orientador: Prof. Dr. Vinicius Amorim Sobreiro.

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

DL377i Defranco Ferreira Peconick, Laura
Inteligência Artificial Aplicada à Previsão de Jogos de
Futebol / Laura Defranco Ferreira Peconick; orientador
Vinicius Amorim Sobreiro. -- Brasília, 2018.
89 p.

Monografia (Graduação - Administração) -- Universidade de
Brasília, 2018.

1. Inteligência Artificial. 2. Redes Neurais
Artificiais. 3. Futebol. 4. Previsão. I. Amorim Sobreiro,
Vinicius, orient. II. Título.

DEDICATÓRIA

Ao longo da minha vida, minha família sempre foi sinônimo de futebol. Seja aos domingos com os gritos (e xingamentos) durante os jogos do Cruzeiro ou nas intermináveis discussões por *WhatsApp* com o (único) primo atleticano, não dá pra falar de Peconick sem falar de futebol. E tudo começou com um jovem goleiro em Conselheiro Lafaiete (vice capital de Minas Gerais) chamado Planika. E é a ele que eu dedico este trabalho, José Andrade Peconick, Planika para os amigos, vovô para mim, que me ensinou a amar esse esporte que me trouxe tantas alegrias.

Com amor, LAURINHA.

AGRADECIMENTOS

Agradeço aos meus pais, pelo apoio incondicional nessa caminhada; à minha irmã, pelos conselhos e por estar sempre ao meu lado; ao Belo, pelos momentos de distração; ao meu orientador, Prof. Vinicius Sobreiro, pela paciência e esforço feito para que o trabalho chegasse à esta versão final; e ao César Augusto, pela imensurável ajuda com o Python, que tornou este TCC possível.

LAURA DEFRANCO FERREIRA PECONICK

EPÍGRAFE

“Acabou! Acabou!”

Galvão Bueno

Final da Copa do Mundo de 1994 contra a Itália.

RESUMO

O futebol é um dos esportes mais populares do mundo, e também um dos responsáveis pelas maiores movimentações financeiras do mundo esportivo. Um dos grandes problemas atuais estudados no futebol é qual a forma de predição de resultados tem o melhor desempenho. A maioria dos trabalhos que aborda esse tema utiliza técnicas estatísticas para realizar as previsões, mas ainda são poucos os que utilizam Inteligência Artificial (IA) para esse fim. Neste trabalho, busca-se usar uma ferramenta da IA para conseguir realizar previsões mais precisas do que as dadas pelo mercado de apostas. Para isso, utilizou-se um código em Python para criar uma Rede Neurais Artificiais (RNAs) para prever os vencedores dos jogos da Copa do Mundo 2018. Essa previsão foi feita em dois momentos, no primeiro sem que os dados fossem atualizados a cada rodada e no segundo com os dados sendo recalibrados, os resultados encontrados foram comparados com os resultados reais e os dados obtidos no mercado de apostas. Conclui-se que RNAs é uma boa ferramenta de previsão, conseguindo resultados semelhantes ao mercado de apostas com uma quantidade limitada de informação.

Palavras-chaves: Inteligência Artificial; Redes Neurais Artificiais; Futebol; Previsão.

ABSTRACT

Football is one of the most popular sports in the world, it is also responsible for the biggest money transfers in the sport's markets. Currently, one of the biggest issues studied in football is which results' prediction tool has the best performance. Most works related to this theme use statistical techniques to make predictions, there are still only a few that use Artificial Intelligence (AI). In this work, the main goal is to use one of the AI tools to make predictions that are more precise than the ones given by the betting market. To achieve this, a Python code was used to create an Artificial Neural Network (ANN) to predict the winner of each game of the 2018 World Cup. The prediction was made in two different scenarios, the first one with the data not being updated and the second one with the data being recalibrated each round, the results found were then compared with the real ones, and the data obtained from the betting market. It is concluded that ANNs are a good prediction tool, achieving results similar to the betting market with a limited amount of information.

Keywords: Artificial Intelligence; Artificial Neural Networks, Football; Soccer; Prediction.

LISTA DE FIGURAS

2.1	Média de gols marcados em copas do mundo pelas oito melhores seleções vs por todas as seleções participantes.	9
2.2	Exemplo de RNAs multicamada de alimentação direta.	17
3.1	Exemplo do gráfico de uma função sigmoide.	24
3.2	Chaveamento conforme a previsão feita pela RNAs “cega”.	27

LISTA DE TABELAS

3.1	Relação de nomes que foram alterados ao longo do tempo.	25
3.2	Exemplo extraído dos dados de pódio.	26
3.3	Exemplo extraído dos dados de resultados.	26
3.4	Pontuação atribuída à comparação entre realidade vs previsão.	28
3.5	Classificação de acordo com os diferentes tipos de previsão.	29
4.1	Resultados cegos da primeira rodada.	32
4.2	Estatísticas da primeira rodada.	32
4.3	Resultados da segunda rodada.	33
4.4	Estatísticas da segunda rodada.	33
4.5	Resultados da terceira rodada.	34
4.6	Estatísticas da terceira rodada.	34
4.7	Resultados das oitavas de final.	34
4.8	Estatísticas das oitavas de final.	35
4.9	Resultados das quartas de final.	35
4.10	Estatísticas das quartas de final.	35
4.11	Resultados das semi-finais.	36
4.12	Estatísticas das semi-finais.	36
4.13	Resultados das finais.	36
4.14	Estatísticas das finais.	36
4.15	Estatísticas globais.	37
1	Resultados das partidas.	54
2	Pódios ao longo dos anos.	54

LISTA DE SIGLAS

<i>CONMEBOL</i>	<i>Confederación Sudamericana de Fútbol.</i>
<i>DFB</i>	<i>Deutschen Fußball Bund.</i>
<i>FIFA</i>	<i>Fédération Internationale de Football Association.</i>
<i>ID3</i>	<i>Iterative Dichotomiser 3.</i>
<i>RF</i>	<i>Random Forests.</i>
<i>SVM</i>	<i>Support Vector Machines.</i>
<i>kNN</i>	<i>k-Nearest Neighbors.</i>
AM	Aprendizagem de Máquina.
CBF	Confederação Brasileira de Futebol.
IA	Inteligência Artificial.
RNAPs	Redes Neurais Artificiais Probabilísticas.
RNAs	Redes Neurais Artificiais.

SUMÁRIO

DEDICATÓRIA	iii
AGRADECIMENTOS	v
EPÍGRAFE	vii
RESUMO	ix
ABSTRACT	xi
LISTA DE FIGURAS	xiii
LISTA DE TABELAS	xv
LISTA DE SIGLAS	xvii
SUMÁRIO	xix
1 INTRODUÇÃO	1
2 REVISÃO DA LITERATURA	5
2.1 Futebol	5
2.1.1 Performance	5
2.1.2 História e impacto social	7
2.1.3 Previsão	10
2.2 Inteligência Artificial	13
2.2.1 Aprendizagem de máquina	15
2.2.2 Redes Neurais Artificiais	16
3 MÉTODO	21
3.1 O código	22
3.2 Os dados	25
3.2.1 Dados brutos	25
3.2.2 Ajustes	26
3.2.3 Geração de dados auxiliares	27
3.3 Análise dos resultados	28
4 RESULTADOS E DISCUSSÃO	31
4.1 Fase de grupos	31
4.2 Mata-mata	34
4.3 Resultados finais	36
5 CONCLUSÃO	39

REFERÊNCIAS

59

GLOSSÁRIO

67

CAPÍTULO I

INTRODUÇÃO

“Em futebol, o pior cego é o que só vê a bola.”

—Nelson Rodrigues.

Futebol é considerado o esporte mais popular do mundo. Evidências de jogos com bola são encontradas na história em diferentes culturas, desde a China antiga, onde era um treinamento militar, em 3000 a.C., passando pelo Japão em um jogo chamado *Kemari*, popular entre os membros da corte japonesa, e pela Grécia com o *Episkiros* no século I a.C., praticado por soldados e disputado entre dois times com 9 a 15 jogadores (Giulianotti, 2012, p. 1).

Já na Idade Média há indícios de um esporte similar ao futebol, mas no qual se empregava mais da violência. Chamado de *Soule* ou *Harpastum*, era praticado por militares que eram divididos em duas equipes: uma voltada ao ataque e uma à defesa. Esse jogo deu origem ao *gioco del calcio* na Itália, que no século XVII se espalhou para a Inglaterra, onde foi formalizado e recebeu pela primeira vez regras em 1863. A difusão internacional do esporte se deu durante o período 1870–1920, impulsionada pelas rotas comerciais e educacionais inglesas. O jogo foi especialmente popular nas novas concentrações urbanas da Europa e da América do Sul (Giulianotti, 2012, p. 1).

Charles Miller foi o precursor do futebol no Brasil, o paulistano foi para a Inglaterra estudar aos nove anos de idade e, quando voltou em 1894, estava familiarizado com o esporte e suas regras. Em 1895 aconteceu o primeiro jogo do país entre os funcionários ingleses da Companhia de Gás e da Companhia Ferroviária São Paulo *Railway*. Apesar de sua popularidade entre operários, o futebol no Brasil foi por muito tempo um esporte elitizado sendo inclusive proibida a participação de negros durante os jogos até 1900. Desde então, o futebol foi democratizado e eleito pelos brasileiros como o esporte nacional (Mazzoni, 1950, p. 5).

Um grande passo para a oficialização do futebol como um esporte foi a criação da *Fédération Internationale de Football Association (FIFA)*, que é o corpo governante do futebol, futsal e futebol de areia. A *FIFA* foi criada em 1904, em Paris, pelas associações nacionais da França, Bélgica, Dinamarca, Holanda, Espanha, Suécia e Suíça com a Alemanha se juntando logo em seguida. No Brasil, o órgão responsável pela organização do esporte é a Confederação Brasileira de Futebol (CBF), criada em 1914 e associada à *FIFA* desde 1923 (FIFA, 2018).

Durante a Primeira Guerra Mundial, a *FIFA* quase foi dissolvida graças às tensões entre os países membros. Apenas em 1930 ocorre a primeira Copa do Mundo entre 13 seleções: Argentina, Bélgica, Bolívia, Brasil, Chile, França, México, Paraguai, Peru, Romênia, Estados Unidos, Uruguai e Iugoslávia, tendo o Uruguai como campeão. Desde então, com exceção dos anos 1942 e 1946, a Copa do Mundo acontece de quatro em quatro anos, em um país sede escolhido previamente por votação dos países membros (FIFA, 2018, p. 1).

A importância da *FIFA* como organização responsável pela coordenação do futebol internacional é refletida nos altos valores econômicos relacionados à ela. Em 2017, a *FIFA* gerou uma receita de 734 milhões de dólares (FIFA, 2017, p. 1), valor baixo em comparação à receita gerada em anos em que Copas do Mundo são realizadas, em 2014, por exemplo, a receita foi de 5.718 milhões de dólares (FIFA, 2014, p. 1). Além dos valores obtidos pela própria organização, eventos como a Copa do Mundo ainda movimentam a economia dos países que sediam, em 2014 foi responsável pela injeção de 4,4 bilhões de reais na economia brasileira (Governo Federal, 2014).

Apesar da grande movimentação de dinheiro relacionada à *FIFA*, os clubes também são responsáveis por uma grande parcela do mercado esportivo. Entre patrocínios, vendas de produtos e ingressos, cotas de televisão e o crescente mercado de apostas, o futebol movimenta bilhões anualmente ao redor do mundo. No Brasil, os vinte¹ principais clubes tiveram uma receita total de 5.061 bilhões de reais² em 2017, já na escala global, os 20 principais times somaram juntos uma receita de 36.163,18 bilhões de reais (Dawson, 2018, p. 1). Já o mercado de apostas movimenta, por ano, entre 2.470 e 3.528 bilhões de reais globalmente (Keogh & Rose, 2013, p. 1).

O grande volume de dinheiro envolvido no mercado de apostas também atraiu pesquisadores que buscam investigar a previsibilidade dos jogos (Stekler, Sendor, & Verlander, 2010, p. 606). De acordo com Stekler et al. (2010, p. 606), as pesquisas nessa área são mais centradas na lucratividade do mercado ao invés da previsibilidade dos jogos em si, e os autores também apontam os métodos mais comumente usados pelos pesquisadores, elencando inclusive quais métodos apresentaram os melhores resultados. As pesquisas sobre previsibilidade em esportes geralmente se dividem em dois grandes grupos: os que utilizam métodos estatísticos para realizar as previsões (Baker & McHale, 2013; Boulier & Stekler, 1999; Constantinou & Fenton, 2017; Dyte & Clarke, 2000; Koopman & Lit, 2013; Min, Kim, Choe, Eom, & McKay, 2008) e o que utilizam métodos de IA ou, mais precisamente, Aprendizagem de Máquina (AM) (Condon, Golden, & Wasil, 1999; Fernandez, Medina, Gomez, Arias, & Gavalda, 2016; Hamilton et al., 2014; Leung & Joseph, 2014; Maszczyk et al., 2014).

Apesar de vários trabalhos já terem sido desenvolvidos sobre a previsão de resultados de jogos de futebol como, por exemplo, Andersson, Edman, e Ekman (2005); Constantinou, Fenton, e Neil (2013); Dixon e Coles (1997); Fernandez et al. (2016); Suzuki, Salasar, Leite, e Louzada-Neto (2010), o número de pesquisas que utilizam AM ainda é pequeno quando comparado com a quantidade de trabalhos que aplicam métodos estatísticos para fazer suas previsões. Mesmo entre os trabalhos de AM, não há uma grande diversidade em relação a essas ferramentas empregadas com esse fim, e é pequeno o número de artigos científicos publicados dedicados a estudar o uso de AM para prever resultados da Copa do Mundo (Groll, Ley, Schauburger, & Eetvelde, 2018, pp. 1–3).

Para melhor entendimento desse tema, é importante que as ferramentas utilizadas na literatura passada e neste trabalho, sejam compreendidas. A IA foi definida por Turing (1950, p.

¹Por ordem decrescente de receita: Flamengo, Palmeiras, São Paulo, Corinthians, Cruzeiro, Grêmio, Atlético Mineiro, Santos, Botafogo, Internacional, Fluminense, Vasco, Atlético Paranaense, Coritiba, Sport, Bahia, Chapecoense, Vitória, Ponte Preta e Goiás.

²Dados retirados dos sites oficiais dos times brasileiros cf. <https://www.cruzeiro.com.br/imagens/bancodearquivos/39418.pdf>.

436) como a capacidade de um computador de realizar tarefas além do que ele foi programado originalmente para realizar. A IA é uma área da ciência da computação que estuda os chamados agentes inteligentes, que são qualquer aparelho que seja capaz de perceber o ambiente onde está inserido e age de acordo a aumentar suas chances de alcançar seus objetivos³.

Uma área de estudo da IA muito desenvolvida por pesquisadores é a AM, que usa diversas técnicas para dar ao computador a capacidade de aprender, isso é, progressivamente melhorar seu desempenho em uma tarefa específica a partir de dados passados ao computador (Haykin, 1998, pp. 57–58). Das várias técnicas existentes em AM, a escolhida neste trabalho é a RNAs, por ser uma técnica bastante utilizada na previsão de eventos futuros baseadas em dados históricos (Haykin, 1998, pp. 56–57).

RNAs são sistemas computacionais inspirados no sistema nervoso, que funcionam com a alimentação de *inputs* que são processados e, então, é dado um *output* (Haykin, 1998, p. 25). RNAs são especialmente úteis para a previsão de eventos a partir de dados passados (ver Tam e Kiang (1992); Tsai e Wu (2008); Zhang, Hu, Patuwo, e Indro (1999)). Foram utilizadas aqui para prever o resultado de partidas de futebol entre seleções na Copa do Mundo com base nos resultados históricos de cada seleção.

Além da baixa utilização da IA na previsão de resultados, deve-se compreender que o futebol é um esporte complexo em que diversas variáveis podem influenciar o resultado de uma partida como, por exemplo, a localização do jogo, as condições climáticas, o estado de saúde dos jogadores, a importância do jogo e as condições financeiras de cada time. No caso da Copa do Mundo, ainda deve ser considerado que os jogadores não estão acostumados a jogar com o resto da equipe, e apenas um time tem o fator de jogar em casa ao seu favor, o que pode afetar a acurácia das previsões. Apesar da estimativa a partir de resultados passados ter se mostrado eficiente, deve-se considerar que os jogadores e os técnicos dos times mudam ao longo do tempo, e que isso afeta diretamente os resultados dos jogos.

Por isso, o desenvolvimento de um método capaz de prever jogos de futebol é tão complexo (Groll et al., 2018, p. 1). Na literatura, o padrão para a mensuração quanto a eficácia de uma ferramenta é o mercado de apostas, que é basicamente a opinião de especialistas, feita com antecedência e tem um índice de acerto de 62,5%, que nesse contexto é considerado satisfatório (Stekler et al., 2010, p. 615). Consequentemente, uma ferramenta de previsão de resultados deve ser ter uma taxa de acertos superior à de mercado para que tenha utilidade.

Ainda segundo Stekler et al. (2010, pp. 606–607), essa importância dada às comparações com o mercado de apostas se deve principalmente ao objetivo primário dos modelos de previsão de resultados, que é majoritariamente a lucratividade do modelo e não sua capacidade de previsão. Baker e McHale (2013); Goddard e Asimakopoulos (2004) e Constantinou et al. (2013) usam o mercado de apostas como medida de eficiência e nos três trabalhos, os modelos apresentados obtiveram índices de acertos superiores ou equivalentes ao mercado de apostas, o que significa que sua aplicação tem lucratividade no mínimo correspondente à do mercado de apostas, o que faz desses modelos opções viáveis para o mercado.

Com base no contexto previamente apresentado, o objetivo neste trabalho é apresentar um método de previsão de resultado de jogos de futebol, utilizando Redes Neurais Artificiais, que consiga prever os resultados da Copa do Mundo de futebol com uma taxa de acertos acima do mercado de apostas. Esse método pode ser adaptado para outros campeonatos de futebol, e como a ferramenta utiliza resultados anteriores dos times e não leva em consideração a forma de pontuação, pode ser utilizada também para outros esportes.

O cálculo da previsão dos resultados ocorreu após o fim da Copa do Mundo, sendo que os dados colhidos são referentes aos jogos que aconteceram até o dia 14 de junho, já que o objetivo é apresentar um método capaz de prever o resultado final de partidas, o time vitorioso ou o empate, de maneira a assertivamente determinar o chaveamento em competições, e que isso

³O tema será desenvolvido com maior propriedade na Seção 2.2.

seja alcançado com uma taxa de sucesso maior do que aquela apresentada pelos mercados de apostas. Para chegar a esse resultado, utiliza-se o seguinte processo dividido em quatro etapas:

- A realização de uma pesquisa bibliográfica sobre a literatura existente na base de artigos científicos *ScienceDirect* para a formulação da base teórica, essa base foi escolhida por se tratar da maior base de pesquisa científica no mundo com mais de 71 milhões de registros, contendo mais de 23.700 periódicos⁴;
- A adaptação de um código existente⁵ e programado em Python⁶, que foi utilizado por se tratar de uma linguagem simples, cujas ferramentas são em sua maioria de fácil e gratuito acesso na internet, com uma grande comunidade de usuários *online*. O autor do código, o Prof. João Pedro Fisanotti, faz a previsão dos resultados da Copa do Mundo de 2014, neste trabalho o código é implementado para a previsão dos resultados da Copa do Mundo de 2018⁷ a partir da utilização de RNAs, o código utilizando RNAs foi escolhido por ser um método de AM muito eficaz na predição de eventos a partir de dados de séries cronológicas, como é o caso de partidas passadas;
- A programação do chaveamento⁸ de acordo com as regras do torneio, para que os resultados obtidos sejam o mais próximos possíveis da realidade;
- A análise dos resultados, a partir da comparação com os dados reais e com as previsões fornecidas pelo mercado de apostas.

Os resultados de jogos de futebol são de grande interesse para os torcedores, por sua ligação emocional com o esporte, mas principalmente aos patrocinadores e emissoras de TV que televisionam as partidas, e investem grandes quantidades de dinheiro nos times na expectativa que tenham resultados positivos. Desse modo, com uma ferramenta capaz de prever de maneira eficaz os resultados dos jogos, patrocinadores e emissoras podem investir nos jogos em que eles acreditam que terão resultados que trarão um maior retorno monetário.

O presente trabalho é dividido como se segue: no próximo Capítulo foi feita uma revisão da literatura existente sobre futebol, englobando sua história, seus impactos sociais e os trabalhos sobre previsão de resultados, AM e análise do mercado de apostas. No Capítulo 3 foi abordado o método utilizado neste trabalho para a realização da previsão dos resultados finais dos jogos da Copa do Mundo; no Capítulo 4 os resultados encontrados foram apresentados e discutidos. Por fim, no Capítulo 5 as conclusões foram expostas, assim como as propostas para pesquisas futuras.

⁴Informação de 20/05/2018.

⁵O algoritmo utilizado neste trabalho foi retirado do endereço https://github.com/fisadev/world_cup_learning

⁶Python é uma linguagem de programação criada por Guido van Rossum e amplamente divulgada a partir de 1991, sua filosofia de criação buscava apresentar uma linguagem que fosse de fácil leitura para os usuários, através do uso significativo de espaços em branco (Kuhlman, 2012).

⁷O código utilizado está sob a “MIT license”, licença que permite a modificação, distribuição, uso comercial e privado do código.

⁸Em algumas competições de futebol, como é o caso da Copa do Mundo, os times são divididos em grupos, nos quais os dois primeiros colocados de cada grupo seguem para uma segunda fase em que são distribuídos no que é chamado de chave. Nessas chaves, o time vitorioso segue na competição até que se tenha um campeão e os derrotados são eliminados.

CAPÍTULO 2

REVISÃO DA LITERATURA

“Clássico é clássico e vice versa.”
—Mário Jardel Almeida Ribeiro.

Neste capítulo uma análise sobre a literatura existente sobre futebol e IA, focando especialmente em RNAs, é feita. No que diz respeito ao futebol, primeiro são apresentadas as áreas de estudo que envolvem o esporte, em seguida a história do jogo é apresentada para, por fim, analisar-se a literatura que envolve previsão de resultados e como ela evoluiu ao longo dos anos. Sobre IA faz-se em um primeiro momento, uma análise global sobre o que é, e quais as principais pesquisas sobre o assunto, então o tema RNAs é abordado.

2.1 FUTEBOL

Por se tratar de um esporte tão popular, é extensa a literatura que trata de futebol. Para facilitar a compreensão e trabalhar de maneira mais organizada, o tema foi dividido em três grandes grupos: o primeiro engloba os trabalhos que fazem um estudo sobre a performance, do jogador, técnico ou time; no segundo grupo estão os artigos que tratam da história do futebol e seus aspectos sociais na comunidade onde está inserido; no terceiro e último grupo estão os artigos sobre previsão de resultados e análise da eficiência do mercado de apostas e que estuda diferentes métodos de previsão.

2.1.1 PERFORMANCE

Os estudos sobre performance geralmente se concentram em uma das áreas: jogador, time, técnico ou campeonato. Szymanski e Smith (1997, p. 135) desenvolveram um modelo empírico da performance financeira dos times da liga inglesa durante o período entre 1974 e 1989 a partir da análise da indústria de futebol inglesa. Os autores tinham como objetivo responder porque forças de mercado geralmente falham a reestruturar indústrias bem estabelecidas, como

é o caso do futebol. Os obstáculos enfrentados nas tentativas de aquisição dos times levam os donos dos clubes a tomarem decisões não racionais e que pequenos grupos sejam capazes de vetar ações que seriam do interesse da empresa, o que explica a dificuldade em reestruturar a indústria. O modelo proposto pode ser usado na análise de impacto de reformas na organização do futebol inglês no futuro.

Já Hughes e Franks (2005, pp. 512–513) analisaram a performance dos times de acordo com sua capacidade de marcar gols. Para isso, os autores analisaram o número de passes que foram necessários para cada seleção conseguir marcar durante jogos de duas Copas do Mundo. A partir dessa análise, concluiu-se que o jogo de passes longos e ligação direta entre ataque e defesa é mais eficiente do que o chamado “jogo de posse”, onde o time favorece o tempo de posse de bola. Ainda é feita uma comparação com o trabalho de Reep e Benjamin (1968), que apresenta uma hipótese similar, de que passes longos de ligação direta são mais efetivos na marcação de gols, mas na visão dos autores, a análise feita é superficial.

Outra linha de pesquisa que trabalha com avaliação de performance é a que tem como foco o jogador. Williams e Hodges (2005, p. 648) notaram uma lacuna na pesquisa científica que trata da aquisição efetiva de habilidades no esporte, enquanto a maior parte das pesquisas tem como foco os aspectos psicológicos ligados ao desempenho dos jogadores. Os autores destacam que os treinadores se baseiam em tradição, intuição e na tentativa de repetir práticas que já viu serem bem sucedidas antes, ao invés de se basear em evidências empíricas. O objetivo com o artigo era fazer uma revisão do conhecimento existente que, a sua época, guiava o trabalho de instrução de jogadores de futebol. Os autores encontraram várias evidências que provam que a maior parte do conhecimento popular que guiava os treinadores não passava de mito. Os autores também buscaram enfatizar para cientistas esportivos a importância do papel de aquisição de habilidades para quem tem interesse no aperfeiçoamento de desempenho.

Por outro lado, Carling, Bloomfield, Nelsen, e Reilly (2008, p. 840) destacam a importância do estudo do movimento para performance de jogadores profissionais de futebol. A preparação física ótima se tornou parte indispensável do futebol profissional, principalmente quando se considera que a exigência física durante os jogos aumentou. Os métodos tradicionais de análise de movimento são extremamente trabalhosos e historicamente restritos a projetos acadêmicos em universidades. A evolução da tecnologia permitiu o desenvolvimento de sistemas complexos capazes de gravar e processar as contribuições físicas dos jogadores durante partidas inteiras que foram implementados no futebol profissional. Carling et al. (2008, p. 860) apontam que o objetivo do trabalho é conscientizar pesquisadores e praticantes dos diferentes sistemas de análise motora disponível, além de identificar quais implicações essa nova tecnologia tem no estudo do movimento e na avaliação de desempenho de atletas.

Uma terceira análise sobre performance de jogadores foi apresentada por Hennig e Sterzing (2010, p. 3), relacionada ao desenho das chuteiras utilizadas e como o desempenho dos jogadores é afetado por mudanças nos calçados dos jogadores. Em seu trabalho, Hennig e Sterzing (2010, p. 9) fazem um sumário do conhecimento existente e discutem para qual direção o *design* de chuteiras deve seguir, baseado em pesquisa biomecânica. Para obter o modelo ideal de calçado, os pesquisadores aplicaram questionários para compreender quais características eram mais importantes no conforto dos atletas, foram elencadas alta tração e boa estabilidade para acelerações, paradas e viradas repentinas. Os autores encontraram que o *design* da chuteira tem impactos significantes da velocidade de corrida, acurácia e velocidades do chute.

Giske, Benestad, Haraldstad, e Høigaard (2013) têm uma abordagem diferente que avalia o processo decisório de técnicos de times profissionais e amadores para medir o sucesso dos seus respectivos clubes. O estudo é feito sob a suposição básica de que o mais eficiente é dependente da qualidade das decisões tomadas por seus técnicos. A pesquisa, realizada com técnicos amadores e profissionais na Noruega mostra que quanto mais experiente é o técnico, maior é seu uso de decisões intuitivas. Esse achado confirma a teoria de que o número de decisões baseadas na

intuição cresce proporcionalmente ao nível de experiência. Também é observada uma diferença na tomada de decisão por intuição entre técnicos profissionais e amadores, com aqueles fazendo um uso maior da intuição do que estes. Para pesquisas futuras, os autores sugerem um aprofundamento na pesquisa sobre como o processo de decisão acontece em diferentes contextos e que estratégias que facilitem a tomada de decisão eficiente sejam clarificadas.

Com base no contexto anterior, os artigos que avaliam performance no futebol têm como objetivo comum indicar, dentro do que foi estudado, uma maneira que dará aos times maior chance de conquistarem seus objetivos. Porém deve-se considerar que os resultados encontrados nessas pesquisas avaliam apenas um fator que pode influenciar no resultado de um jogo e que futebol é um esporte dinâmico, está em constante evolução e que tem uma alta rotatividade de atletas. Ou seja, a forma como esses fatores influenciam o resultado podem variar de acordo com a geração de atletas e treinadores.

As pesquisas acadêmicas sobre as influências que diversos fatores têm no resultado de jogos de futebol são mais uma ferramenta para os *stakeholders* do mundo do futebol. Os times podem fazer uma melhor preparação para jogos futuros; empresas patrocinadoras podem investir em certos times com maior segurança de retorno; os próprios jogadores podem tomar melhores decisões quanto aos seus futuros, como, por exemplo, com qual time assinar em uma janela de transferência¹; e os torcedores, que podem utilizar dessas ferramentas para apostar em jogos.

2.1.2 HISTÓRIA E IMPACTO SOCIAL

O segundo grupo de estudos sobre futebol trata da história do esporte, seu contexto social e como ele influencia e é influenciado pela sociedade. Nesse contexto, o trabalho de Elsey e Pugliese (2017, p. 2) expõe 17 textos sobre as interdisciplinaridades presente no futebol, que analisam desde o poder da comunicação visual em diferentes contextos sociais até no poder de influência da política local no desenvolvimento do esporte em um país. Os textos são divididos nas seguintes disciplinas: Crítica de Arte; Psicologia Coletiva e Identidade de Grupo; Ciência Política e Relações Internacionais; Raça e Estudos Étnicos; Sexualidade e Gênero; o Estado e a Sociedade Civil; e Filosofia e Teoria Crítica. A justificativa para essa coletânea de textos é que os autores acreditam que o estudo da história do futebol fornece uma oportunidade única de compreensão da condição humana.

Em um desses textos, Kittleson (2017, p. 161) faz uma análise interessante de como o racismo no futebol brasileiro conviveu, e ainda convive, com a idolatria de jogadores negros. Nesse trabalho, o autor foca no primeiro grande ídolo do futebol brasileiro, Fausto dos Santos, que conviveu tanto com a glória graças à sua fama no esporte, quanto com o descaso com mulatos e negros no Brasil à época, refletido nas precárias condições de saúde que dispunha, tendo morrido de tuberculose aos 34 anos. O autor ressalta a dualidade do tratamento do brasileiro com Fausto, que também era conhecido como “Maravilha Negra” por seu estilo de jogo vigoroso e pela sua grande habilidade com a bola que eventualmente foi considerado “futebol mulato”. Enquanto dentro de campo, Fausto era reconhecido e exaltado pelas suas conquistas, mas no momento em que retornava para a “vida real”, sofria com os mesmos preconceitos que seus semelhantes.

Ainda no impacto social do futebol, Boyle (2016, p. 407) aborda como a história e a economia política podem ser ensinadas utilizando o futebol. Quatro argumentos são feitos pelo autor: não é difícil criar um currículo acadêmico rigoroso focado no ensino do esporte; Ensino sobre futebol permite que o professor experimente com sua técnica de ensino, principalmente no que tange envolvimento da comunidade; A imensa popularidade do esporte faz com que seja mais fácil para os alunos assimilarem o jogo de maneira crítica; e o quão gratificante é a experiência de lecionar sobre futebol. O autor defende seus argumentos a partir de sua própria

¹O período do ano em que os times estão autorizados pelas respectivas federações a adquirir novos jogadores.

experiência no ensino desde uma escola de arte liberal na Califórnia até uma prisão em Uganda.

O artigo de Dietze (2012) já trás uma abordagem diferente na análise do contexto social que está inserido no mundo do futebol. A autora investiga os efeitos de raça, gênero e sexualidade na seleção alemã de futebol. Em países onde futebol é o principal esporte, as seleções nacionais acabam ganhando *status* simbólico na narrativa nacional. Essa afirmativa é especialmente evidente em casos onde o país em questão foi desacreditado internacionalmente, como foi o caso da Alemanha pós-Segunda Guerra. A vitória alemã na Copa do Mundo de 1954 trouxe ao país um senso de orgulho e união há muito perdido, mas também foi responsável pelo estabelecimento do padrão de alemão ideal, representado pelos jogadores da seleção, sendo homem, branco e nascido na Alemanha. Em 1955, as alemãs foram banidas da *Deutschen Fußball Bund (DFB)*, órgão responsável pela organização do futebol na Alemanha, medida relacionada à rígida divisão de trabalho da época, ampliada pela glorificação por parte da mídia da mulher no papel de dona de casa (Dietze, 2012, p. 55).

Apenas nos anos 60, com o fortalecimento do movimento feminista ao redor do mundo, começaram a criticar o modelo da mulher ideal como submissa e dependente (Dietze, 2012, p. 54). Mesmo assim, foram anos até que a *DFB* suspendesse a proibição de mulheres e permitisse que times femininos jogassem em estádios de futebol profissional. Pelak (2009, pp. 63–64) fez um estudo similar, mas abordando a evolução do futebol feminino na África do Sul e como ele estava ligado à questões de gênero e do afastamento gradual da noção de feminilidade submissa. A autora divide em três as fases de evolução do futebol feminino sul-africano: seu surgimento e desenvolvimento inicial entre 1970 e 1990, seu crescimento e transição entre 1991 e 2000 e sua institucionalização desde 2001 até os dias atuais. Também são discutidas as principais contribuições de administradores, líderes políticos e atletas proeminentes na construção de uma cultura de futebol voltada ao público feminino (Dietze, 2012, p. 56).

O contexto da globalização e como a seleção nacional holandesa ajudaram os Países Baixos a construir uma identidade nacional foram estudados por Lechner (2007, p. 216). O autor ainda mostrou como o envolvimento da seleção holandesa em competições internacionais ajudou na construção de um mito de distinção do futebol nacional, mito esse reforçado pela cobertura midiática e que o transformou em um elemento básico da comunidade nacional. Esse ponto de vista que os holandeses têm de serem o país do futebol é compatível com o padrão global, e uma análise sociológica mais aprofundada é capaz de expor esse mito pelo que ele é. O artigo ainda apresenta uma crítica da imagem feita das seleções nacionais, que força a ideia de que os jogadores são heróis do país.

Seguindo uma linha similar, Kaelberer (2016, pp. 276–278) analisou os discursos de identidade nacional na Alemanha relacionados ao futebol. O autor deu enfoque aos quatro títulos de Copa do Mundo do país, e como esses títulos vieram em momentos críticos da história alemã no pós-guerra. De acordo com o autor, ao se observar essas quatro vitórias alemãs é possível traçar as mudanças na identidade nacional alemã que eram refletidas não só nos times vencedores, mas também no relacionamento da população geral com o time e o país. O artigo alcança seu objetivo de analisar estas mudanças ao fazer uma interpretação comparativa de um período de mais de seis décadas do time nacional.

Apesar de também estudar as relações entre futebol e a identidade nacional, Bunk (2015, p. 453) faz isso ao analisar o uso do futebol como ferramenta de manutenção da identidade nacional de um grupo em um país estrangeiro. Na década de 1920, espanhóis que imigraram para Nove Iorque usaram o futebol como ferramenta de criação e manutenção de suas identidades nacionais e regionais. O sucesso dentro de campo era considerado um símbolo da força e superioridade sob seus rivais. O futebol acabou criando redes de contatos entre espanhóis, formando um senso de unidade e comunidade fora de seu país. O autor nota que, ao contrário do que se esperava com o resultado desses jogos, isto é, que o futebol promovesse a integração entre os espanhóis e a comunidade onde se inseriram, o efeito foi oposto, criando pequenas comunidades

cada vez mais afastadas da sociedade onde estavam incorporadas.

No artigo de Sorek (2003, pp. 417–419) uma oportunidade parecida foi exposta, mas não utilizada: a do futebol como maneira de estabelecer um senso de identidade nacional de árabes em Israel. No artigo, o desenvolvimento histórico do futebol árabe-palestino durante o período do Mandato Britânico e depois como essas relações mudaram sob o Estado de Israel. Em grande parte, essa não formação de uma identidade nacional recai nos próprios palestinos, que preferem manter o futebol separado de suas lutas políticas, o tendo como um espaço seguro, enquanto a luta pela identidade e reconhecimento nacional segue forte em outras frentes.

Já Milanovic (2005) apontou as características do futebol que permitiram que o esporte chegasse aos altos níveis de globalização que ele mostra hoje em dia. Com o fim do limite do número de jogadores estrangeiros em várias ligas da Europa², a livre circulação de jogadores aumentou exponencialmente. De maneira oposta, os times nacionais aumentaram suas restrições a jogadores que não nasceram no país. O modelo apresentado pelo autor mostra que onde há livre circulação de trabalhadores, rendimentos de escala crescentes e endogeneidade de habilidades, pode ser observado um aumento na qualidade do jogo e na desigualdade dos resultados entre os clubes ao mesmo tempo que diminui a diferença entre as performances das seleções. Exemplos empíricos da liga europeia e da Copa do Mundo validam esse modelo, já que com o passar dos anos e a maior liberdade de movimento de jogadores, ambos campeonatos passaram a ser menos díspares, o que demonstra uma maior igualdade entre os competidores, explicada pelo autor graças ao aprimoramento geral de jogadores.

Na Figura 2.1 é bem ilustrada a teoria defendida por Milanovic (2005), mostrando que ao longo dos anos a diferença entre os gols marcados pelos principais oito times que disputam a Copa do Mundo e os demais diminuiu, até que eventualmente a relação se inverteu. O intercâmbio de jogadores entre países explica esse fenômeno, já que permite que os jogadores se desloquem para jogar em ligas melhores, aumentando o nível de competitividade entre as seleções.

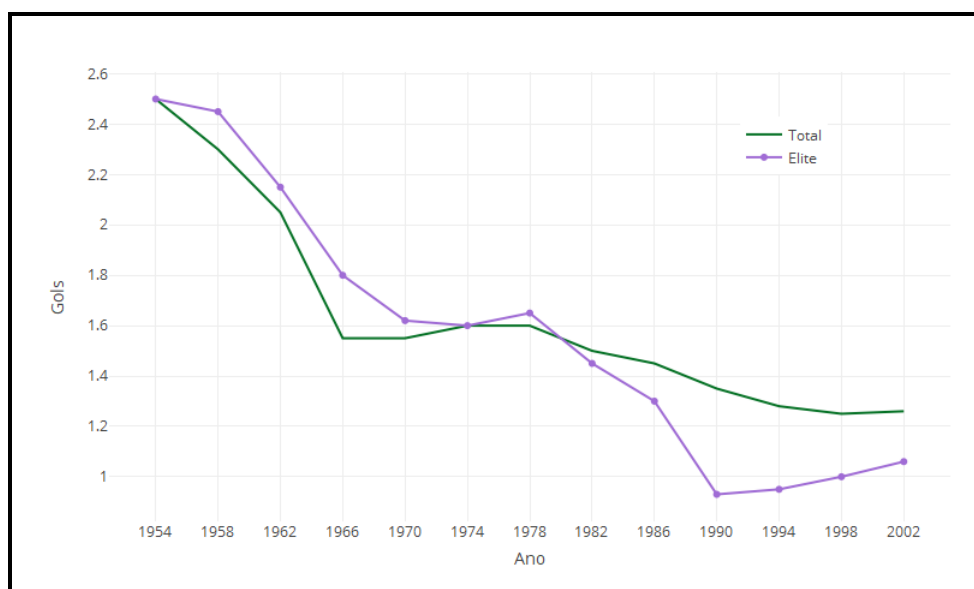


Figura 2.1: Média de gols marcados em copas do mundo pelas oito melhores seleções vs por todas as seleções participantes.

Fonte: Adaptado de Milanovic (2005, p. 843).

Considerando a grande popularidade do esporte, seu enorme impacto social e histórico em países onde ele é o principal esporte não é inesperado. A sua utilização como forma de

²Algumas ligas ainda mantêm o limite, como a inglesa – máximo 17 jogadores estrangeiros – e a espanhola – máximo de 5 jogadores estrangeiros.

controle da população por regimes autoritários (Else & Pugliese, 2017, p. 57) e como parte da construção de uma identidade nacional (Bunk, 2015; Kaelberer, 2016; Lechner, 2007) são testemunho da grande influência do esporte sobre diversas populações. Também é possível – em países em que o futebol está integrado com ao dia a dia de grande parte da população como, por exemplo, Brasil, Alemanha, Rússia, México – fazer uma análise sobre como o esporte teve influência em sua formação como país e no desenvolvimento cultural dessas regiões.

Não apenas o futebol afetou a formação cultural, mas o contrário também aconteceu. Culturas fortemente machistas, com predominância do patriarcado, principalmente na América do Sul, dificultaram ou impediram o acesso de mulheres ao mundo do futebol. Apenas recentemente a modalidade feminina do esporte vêm ganhando algum destaque, mas ainda não há comparação com sua contraparte masculina. Entre o praticamente inexistente tempo de televisão, raros patrocínios e baixo interesse geral do público, é difícil encontrar atletas que estejam dispostas a enfrentar essas dificuldades, o que cria um ciclo vicioso, onde não há interesse de patrocinadores e times em investir no futebol feminino, o que faz com que o esporte não evolua, levando à baixa relevância com o público o que faz com que patrocinadores e times não tenham interesse em investir.

Entretanto, há alguns anos vem crescendo o apoio das federações internacionais ao futebol feminino. A *FIFA* vêm aumentando seu investimento no futebol feminino (FIFA, 2014, 2017) e, em 2016 a *Confederación Sudamericana de Fútbol (CONMEBOL)* adicionou uma nova regra ao seu estatuto, que prevê a eliminação de times da Copa Libertadores, principal competição sul-americana, caso os clubes não tenham um time feminino em seu elenco (CONMEBOL, 2016).

2.1.3 PREVISÃO

O terceiro grande grupo de pesquisa abrange os estudos que buscam encontrar um método de previsão de jogos de futebol, esse grupo de ser dividido em dois subgrupos principais de acordo com a ferramenta utilizada para realizar a previsão: os que utilizaram métodos estatísticos e os que utilizaram métodos de AM. Será observado também qual o propósito dos trabalhos e se o objetivo é prever o resultado da partida (Vitória, empate ou derrota para o time A) ou o placar final (Quantos gols o time A marcou no time B, e vice-versa).

O primeiro autor a utilizar distribuição de Poisson³ para fazer previsões sobre o placar de jogos de futebol foi Maher (1982). Em seu estudo, o autor também inova ao ser o primeiro a fazer uma análise das forças ofensivas e defensivas dos times, e usou essas diferentes forças em seu modelo de previsão. Os dados de três temporadas (1973, 1974 e 1975) das quatro divisões do futebol inglês foram utilizados para comparação com os resultados esperados dados pelo modelo. Maher (1982, pp. 117–118) conclui que o modelo proposto é adequado e oferece bons resultados quando comparado com os resultados reais.

Dixon e Coles (1997, pp. 265–266) propuseram um modelo paramétrico⁴ aplicado aos dados liga inglesa entre 1992 e 1995, os autores buscam explorar possíveis ineficiências no mercado de apostas através das chances dos agentes de aposta de 1995 e 1996. Para isso é usada a distribuição de Poisson para prever o resultado das partidas de futebol de maneira a superar os agentes de aposta em eficiência. Os resultados encontrados foram satisfatórios e os autores notam que a simplicidade do modelo e da estratégia de apostas são os diferenciais do modelo apresentado.

³Distribuição de Poisson é uma distribuição de probabilidade de uma variável aleatória discreta que traduz a probabilidade de certo evento acontecer durante um determinado período de tempo, considerando que um evento é independente do seu antecessor (Haight, 1967, p. 30). A distribuição de Poisson é dada por: $f(k; \lambda = \frac{e^{-\lambda} \lambda^k}{k!})$, onde λ é o número esperado de ocorrências e k o número real de ocorrências.

⁴Modelo paramétrico é aquele que possui um número finito de parâmetros (Freedman, 2005, p. 18).

Em outro trabalho, apresentado por Dyte e Clarke (2000, p. 993), a distribuição de Poisson também é utilizada, mas o objetivo buscado era o de simular os placares de jogos da Copa do Mundo de 1998 e compará-los com os placares reais. Os autores trataram os gols como variáveis independentes de Poisson relacionadas aos rankings da *FIFA*⁵ de cada seleção e à localização de cada partida. Os autores fazem a ressalva de que, para o modelo sugerido apresente resultados mais fieis à realidade, são necessários ajustes manuais no ranqueamento das seleções.

Koopman e Lit (2013) desenvolveram um modelo estatístico capaz de avaliar e prever resultados de jogos de futebol, nesse modelo utiliza-se uma distribuição de Poisson bivariada cujos coeficientes de intensidade são modificados ao longo do tempo. Os autores se baseiam no tratamento eficiente conferido ao recolhimento de dados. Para realizarem suas previsões, foram coletados os resultados de nove campeonatos da primeira divisão inglesa. O número de gols marcados por um time pode depender de sua força de ataque, da força de defesa do seu oponente, se o time joga em seu estádio e como a partida se desenvolve (se algum jogador for expulso, por exemplo, isso muda a perspectiva do jogo). A novidade apresentada pelo artigo é a dinamicidade do modelo, que permite considerar várias influências no resultado do jogo. Os resultados obtidos foram suficientes para que fosse conseguido um retorno positivo contra os agentes de aposta.

Outra técnica estatística utilizada por pesquisadores na construção de modelos de previsão de resultados de jogos de futebol é a regressão logística⁶. Goddard e Asimakopoulos (2004, p. 51) apresentaram um modelo de regressão para prever os resultados dos jogos da primeira divisão inglesa a partir de dados de 10 anos. Além do resultado de partidas anteriores, o peso da partida para o resultado final do campeonato, o envolvimento dos times na competição e a distância entre as cidades sede dos dois times contribuem para o desempenho do modelo. As apostas feitas no final da temporada geraram retorno positivo contra os agentes de aposta.

Hvattum e Arntzen (2010, p. 460) também usaram regressão logística para realizar as previsões de resultados de jogos de futebol, mas em conjunto com o sistema de ranqueamento ELO⁷ que, baseado em resultados passados, atribui um valor ELO para cada time de acordo com as habilidades e performances dos seus jogadores. Os autores avaliam os resultados encontrados com esse tipo de previsão quando comparados a seis medidas econômicas, com o mercado de apostas, e outros métodos estatísticos. Após a contraposição entre o sistema apresentado pelos autores e os seis outros métodos usados como *benchmarks*, o sistema apresenta resultados piores que dois medidores econômicos, porém possui melhores resultados que os outros quatro. Conclui-se que o método que utiliza valores ELO é um previsor significativo dos resultados de partidas, e é útil para englobar informações passadas.

A técnica estatística mais utilizada por pesquisadores para previsão de resultados de jogos de futebol é a abordagem Bayesiana⁸. Suzuki et al. (2010, pp. 1530–1531) utilizaram essa técnica para prever os resultados das partidas da Copa do Mundo de 2006. Como informação básica, os autores utilizaram a opinião de especialistas e os rankings da *FIFA*. Os autores aplicaram seu método para calcular as probabilidades de vitória, empate e derrota para cada partida e também simularam a competição relacionada como um todo, tornando possível a estimativa de classificação na fase de grupos e a probabilidade de cada seleção sair vencedora do torneio. Os resultados encontrados foram satisfatórios, encontrando um índice de acerto 73,45% maior

⁵A *FIFA* faz um ranqueamento das seleções de acordo com os resultados das partidas oficiais entre os times. A pontuação por jogo varia de acordo com o resultado, importância do jogo, força do oponente e a força da confederação a qual a seleção pertence.

⁶Regressão logística é a técnica estatística que, a partir de um conjunto de dados, produz um modelo capaz de prever valores de uma variável binária dependendo de um conjunto de variáveis denominadas explicativas contínuas ou binárias (Hosmer & Lemeshow, 2000, p. 1).

⁷Sistema inventado por Arpad Elo, professor húngaro-americano, que originalmente era usado para classificar jogadores de xadrez (Elo, 1978, p. 1).

⁸Abordagem Bayesiana é aquela que avalia hipóteses pela máxima semelhança (Howson & Urbach, 2005, p. 5).

do que o esperado.

Constantinou, Fenton, e Neil (2012, p. 1) definiram uma rede bayesiana como um modelo probabilístico gráfico que representa dependências condicionais em variáveis incertas, que podem ser objetivas ou subjetivas. Os autores apresentam um modelo de rede bayesiana para prever os resultados de jogos de futebol, onde as variáveis subjetivas representam fatores que são importantes para a previsão, mas que os dados históricos não conseguem refletir. As previsões foram feitas a respeito dos resultados da primeira divisão inglesa e publicadas na internet antes do início dos jogos. Os autores mostram que o modelo apresentado é capaz de gerar lucro para quem aposta contra os agentes de apostas, mesmo quando se considera a margem de lucro destes.

Dando continuidade ao estudo iniciado em Constantinou et al. (2012), Constantinou et al. (2013, pp. 322–323) propuseram uma rede bayesiana mais simples do que a apresentada anteriormente que é capaz de gerar retornos ainda mais lucrativos. Os autores utilizam esse novo modelo para prever os resultados dos jogos da primeira divisão inglesa na temporada de 2011/2012. Para avaliar a eficácia do modelo em prever resultados, diferentes procedimentos de aposta foram considerados, demonstrando que: usando a formação do time no cálculo, apresentava mais sucesso do que ao considerar apenas as séries históricas; fadiga e motivação do time foram menos eficientes do que formação do time, mas ainda apresentaram resultados melhores do que aqueles de séries históricas; e o mercado de apostas é constantemente parcial à times de acordo com seu valor comercial e não pela sua performance de fato. Os resultados encontrados foram melhores do que aqueles apresentados pelo modelo anterior e o modelo em si é mais simples.

Em seu trabalho mais recente, Constantinou e Fenton (2017, pp. 91–94) fizeram uma crítica ao uso desmedido de Aprendizagem de Máquina nas tarefas de predição. Os autores defendem que é mais vantajoso trabalhar com um modelo que identifique quais informações são necessárias do que ter um modelo que processe todos os dados disponíveis. É aplicado então um conceito chamado “dado inteligente”, que é um método de processamento de dados e de conhecimento que coloca mais ênfase na aplicação de conhecimento de causa e fatos no processo de desenvolvimento de modelo, guiado pelos dados que o modelo precisa e não no que está disponível. A eficácia desse modelo é demonstrada pela sua aplicação na previsão da evolução da performance de times de futebol com dados limitados, onde consegue obter resultados com alto grau de assertividade.

Os autores destacaram ainda fatores que são determinantes para a performance do time ao longo do campeonato: Gastos com salários e transferência de jogadores, times que gastam mais tendem a ter um desempenho melhor; envolvimento em competições europeias, times que participam de competições internacionais concomitantemente ao campeonato nacional tendem a ter desempenhos piores na disputa pátria; as trocas de técnicos não tem o impacto que se esperava; as lesões têm um impacto negativo significativo nos times, especialmente quando os jogadores muito habilidosos são poucos; os times que subiram de divisão mostram um desempenho muito superior àqueles que foram rebaixados em seu lugar (Constantinou & Fenton, 2017, p. 103).

Ainda há os estudos que focam em predição por meio da opinião de especialistas. Andersson et al. (2005, p. 565) fizeram uma comparação entre as previsões de especialistas e leigos para os resultados da primeira rodada da Copa do Mundo de 2002. Foi aplicado um questionário para 251 participantes em quatro diferentes níveis de conhecimento de futebol, desde especialistas a pessoas que não estavam familiarizadas com o esporte. Os autores encontraram que modelos básicos de previsão obtinham resultados muito superiores aos dos participantes, inclusive os que se declaravam especialistas. A única diferença entre especialistas e leigos estava baseada no nível de confiança de cada um, não na acurácia das previsões.

Mais recentemente, a AM passou a ser mais utilizada na literatura de futebol como ferramenta de previsão. Fernandez et al. (2016, p. 136), por exemplo, usam AM para prever performances dos jogadores. Com a aprovação por parte da *FIFA* de tecnologia de sistema de acom-

panhamento eletrônico durante partidas oficiais, uma grande quantidade de dados passou a ser disponibilizada sobre como a performance do jogador evoluía durante ao jogo. Os autores então estudaram o desempenho dos jogadores do Barcelona durante a temporada de 2015/2016 tanto durante treinos quanto jogos. Nessa pesquisa foi avaliado o quanto é possível prever em relação a performance dos jogadores e o quanto isso influencia no resultado de uma partida. Para a análise desses dados, os autores usaram tanto *Random Forests (RF)* quanto *Support Vector Machines (SVM)*⁹ e comprovaram que utilizando seus algoritmos nos dados, é possível prever a performance de jogadores e, conseqüentemente, do seu time.

Em seu estudo Martins et al. (2017, p. 79) utilizaram algoritmos polinomiais em dados históricos dos times para prever o resultado de jogos de futebol. Os autores usaram informações obtidas de olheiros durante as partidas para coletar dados sobre os times além do seu recorde histórico. O sistema proposto aplica um algoritmo polinomial para analisar e definir os resultados. A associação desses algoritmos polinomiais com a AM levou a uma porcentagem de acerto superior a 96%, depois de selecionadas os fatores importantes do período de teste. Os autores citam como desvantagem do método utilizado o crescimento exponencial do processamento de tempo do algoritmo com cada nova variável estudada e o fato dele não poder ser usado em tempo real, se adaptando a realidade da partida.

Por fim, o mais recente artigo sobre previsão de futebol utilizando IA foi publicado por Groll et al. (2018) que compara três métodos diferentes de previsão, distribuição de Poisson, *RF* e métodos de ranqueamento¹⁰ para prever o resultado da Copa do Mundo de 2018. O modelo utilizado é uma combinação entre *RF* e o sistema de ranqueamento, que através de teste, obtiveram os melhores resultados. Esse novo método então é aplicado nos dados entre as Copas de 2002 e 2018, o resultado encontrado é uma final entre Alemanha e Espanha, em que o modelo favorece a Espanha por uma pequena margem.

2.2 INTELIGÊNCIA ARTIFICIAL

Nesta Seção são abordados os temas mais técnicos relativos ao presente trabalho, primeiro a partir da revisão da literatura sobre IA, na qual procura-se apresentar conceitos e dar diferentes definições além de expor parte da literatura existente. De maneira mais pontual, é exposto posteriormente o referencial teórico sobre AM e RNAs.

Turing (1950, p. 433) trata em seu trabalho a possibilidade de desenvolvimento de uma máquina capaz de reproduzir o pensamento humano. O autor propõe um jogo, em que três pessoas participam, sendo um homem, uma mulher e um investigador que pode ser de qualquer gênero, o objetivo do jogo é que, através de perguntas o investigador seja capaz de deduzir quem é o homem e quem é a a mulher. Turing discute então o que aconteceria caso um dos investigados fosse substituído por um computador, quais sinais poderiam ser percebidos por essa troca e como a permuta deveria acontecer para que o investigador fosse “enganado”.

Alguns anos após a publicação do trabalho de Turing, Minsky (1961, p. 8) faz uma análise do caminho a ser percorrido para se alcançar a IA. O autor divide a questão de como é possível ensinar a uma máquina como resolver problemas em cinco etapas, em ordem: Busca, reconhecimento de padrões, aprendizado, planejamento e indução. Ainda é ressaltada a conexão entre as etapas, em que reconhecimento de padrões ajuda a melhorar a eficiência da busca e, em conjunto com o aprendizado, pode explorar generalizações baseadas em experiência acumulada. O planejamento, quando aplicado à busca, reduz significativamente a amplitude da investigação necessária para a resolução do problema proposto. A indução é necessária para que o computa-

⁹Para mais informações, cf. Breiman (2001) para *RF* e Cortes e Vapnik (1995) para *SVM*.

¹⁰Para esse método, estima-se parâmetros adequados que refletem a força atual dos times.

dor seja capaz de criar modelos de soluções para eventos similares e aplicar esses modelos às diferentes situações sem a necessidade de programação adicional.

Trabalhos em IA vão além da programação em computadores, Goldstein e Papert (1977, p. 85), por exemplo, fazem uma análise da IA e seu relacionamento com o estudo da linguagem e do processo de aquisição de conhecimento. Os autores defendem que

“Inteligência Artificial é baseada na capacidade de se utilizar grandes quantidades de diversos tipos de conhecimento de uma maneira sistemática, ao invés da posse de princípios gerais e uniformes.”

IA representa uma nova maneira de se encarar problemas de cognição, o que causa um grande impacto no estudo de linguagem (Goldstein & Papert, 1977, p. 84). Essas duas temáticas, estudo de linguagem e IA, são exploradas no artigo desses autores em sete seções distintas: a primeira trata da relação entre inteligência artificial com o estudo do processo cognitivo, a segunda estuda a teoria dos quadros como abordagem de representação de conhecimento, a terceira elenca os problemas na epistemologia computacional, a quarta aborda o auto conhecimento enquanto a quinta seção fala das estruturas de controle. A próxima seção trata da evolução da IA enquanto na sétima e última seção é feita uma análise do impacto da IA na educação.

Os trabalhos mais antigos sobre IA não estudam exclusivamente como essa nova ferramenta impacta certas áreas do conhecimento. Hart (1984), por exemplo, analisou como ocorreu a transição da IA de um área de estudo que interessava uma pequena quantidade de pesquisadores para uma pauta entre líderes mundiais e cidadãos comuns. O autor ainda estudou como a evolução do estudo e do desenvolvimento de produtos de IA ocorreu em seus primeiros vinte anos, período que ele dividiu em três partes: na primeira, durante a metade dos anos 1960, compreende o período de pesquisa pura, em que o que era produzido era material empírico sem utilidade para aplicação, mas importantíssimo para a formação da base teórica da técnica.

Durante os anos setenta esse cenário mudou, com os pesquisadores procurando aplicar os conhecimentos previamente adquiridos a problemas da vida real, e, apesar de terem alcançado seus objetivos, ainda era de forma prematura e não adequada para o mercado; já no período mais recente à época de publicação do trabalho de Hart, a primeira classe de produtos comerciáveis apareceu, entre eles, um dos mais comuns é o de identificação visual. Por fim, o autor destacou a velocidade com que a IA cresce como área de pesquisa, ele afirmou que o processo de evolução é imparável e ficará cada vez mais acelerado e finaliza levantando a questão se estamos ou não prontos para as mudanças que a IA irá proporcionar (Hart, 1984, p. 20).

Alguns anos mais tarde, em seu livro, Haykin (1998, p. 56) definiu os objetivos da IA como o desenvolvimento de paradigmas ou algoritmos que exigem que máquinas desempenhem tarefas cognitivas. Para esse autor, um sistema de IA deve ser capaz de realizar três funções: armazenar conhecimento, aplicar esse conhecimento na resolução de problemas e a adquirir novos conhecimentos pela experiência. O autor ainda destacou os três elementos básicos da IA: Representação, que é a utilização da estrutura de linguagem para representar dados sobre o problema de interesse ou dados sobre a solução do problema; Raciocínio, que é a capacidade de resolver problemas; e Aprendizagem, onde o ambiente oferece informações a um elemento de aprendizagem, esse elemento então usa essa informação para fazer melhorias na base de conhecimento e então o elemento de desempenho usa esse conhecimento para realizar a tarefa desejada.

No trabalho de Buchanan (2006, p. 53) foi apresentado um apanhado histórico da IA, destacando a interdisciplinaridade da área com outros campos de pesquisa, como a filosofia. O autor vai além, relacionando a IA com obras da literatura e do cinema que moldaram a perspectiva e as expectativas do público geral em relação ao que era a área e do que ela era capaz. O autor também citou alguns dos pesquisadores que foram responsáveis pelo desenvolvimento da área, destacando como o trabalho deles foi fundamental na evolução da área para que ela

chegasse até os dias atuais.

De maneira análoga ao feito no trabalho de Hart (1984) e Buchanan (2006), Brunette, Flemmer, e Flemmer (2009, p. 385) fazem uma revisão sobre a área de IA com o objetivo de mapear os principais trabalhos sobre o tema e dividi-los em subtemas mais comuns da área. O primeiro subtema é o de modelos de consciência, onde os autores destacam que apesar das várias tentativas de se replicar uma estrutura de consciência e controle, os sistemas criados para isso não foram implementados. O segundo subtema é o de linguagem computacional, onde os trabalhos em que foram desenvolvidas essas linguagens são revisados. O próximo subtema é o de modelos baseados em agentes, em que o comportamento do agente varia de acordo com seu estado emocional, que é calculado como a combinação de valores de quatro emoções distintas: raiva, felicidade, tristeza e medo. O quarto e o quinto subtema tratam de assuntos análogos, aquele fala das propostas para inteligência incorporada, ou seja, da utilização da inteligência em robôs físicos e este de experimentos reais feitos nessa área. Os autores destacam que não há um subtema que se sobressaia na IA e que as conquistas da área estão muito aquém do que era esperado.

A pesquisa e o desenvolvimento do campo de IA são discutidos no trabalho de Ning e Yan (2010, p. 110), que afirmaram que ela é a grande conquista da ciência da computação no século 21. Os autores discutiram ainda qual a direção está sendo tomada pelos pesquisadores atualmente, além de sumarizar sua história e apresentar qual o destino dessa área de pesquisa. É dada a seguinte definição para IA:

“É uma nova ciência tecnológica, que pesquisa e desenvolve a fim de simular, estender e expandir a inteligência humana em teoria, métodos, técnicas e aplicações. É um ramo da Ciência da Computação que procura entender a substância da inteligência e produzir uma nova máquina de inteligência que poderia produzir reações similares às de inteligência humana.”

Atualmente, a máquina capaz de empregar a IA é o computador, como consequência, a história do desenvolvimento da IA está conectada com a história do desenvolvimento da ciência da computação e da tecnologia (Ning & Yan, 2010, p. 110). Os autores ainda destacam a capacidade de interdisciplinaridade da IA, que envolve também teoria da informação, cibernética, automação, biologia, psicologia, lógica matemática, linguística, medicina e filosofia.

Com a disseminação da tecnologia e o crescente contato da população com a IA, uma preocupação que surgiu na sociedade é em relação a segurança das pessoas em contato com máquinas cada vez mais independentes (Kose & Vasant, 2017, p. 1). Essa criação gerou uma nova área de pesquisa relacionada à segurança e ao futuro da IA que foi explorada no trabalho de Kose e Vasant (2017). Os autores apresentam uma breve revisão do que já foi estudado na área, não só com os trabalhos já desenvolvidos, mas também com centros de pesquisa que se especializam no assunto. Os autores também propõem uma possível solução para o problema, com a implementação de tempo de vida para softwares de IA, esse tempo seria calculado de acordo com variáveis operacionais, e assim, quando o tempo se esgotasse, o sistema de inteligência se esgotaria, dando espaço para novos sistemas mais seguros.

2.2.1 APRENDIZAGEM DE MÁQUINA

Em 1959, Samuel publicou seu trabalho em que ele aplica AM para ensinar um computador a jogar damas. O objetivo do trabalho de Samuel (1959) era preencher uma lacuna no que diz respeito aos métodos específicos de resolução de problemas, já que o conhecimento sobre os princípios básicos das técnicas de IA ainda eram muito rudimentais. O autor buscava conseguir, através da AM, ensinar a um computador a jogar damas melhor do que um ser humano em menos de 10 horas. O autor concluiu que era possível desenhar esquemas de aprendizado que irão ter um desempenho muito superior ao do ser humano e que esses esquemas podem possivelmente

ser economicamente viáveis em sua aplicação a problemas da vida real.

Alpaydin (2004, p. 2) defendeu que a AM é a área que mais cresce na Ciência da Computação, não só pelo aumento desenfreado da quantidade de dados, mas também pela crescente necessidade de se processar esses dados para transformá-los em conhecimento. O autor também fez uma análise de como esses dados evoluíram ao longo do tempo, deixando de ser apenas sequências de caracteres para novos formatos em vídeo e áudio. Alpaydin (2004, p. 5) destacou que a AM é a programação de computadores para otimizar sua performance usando dados de exemplo ou experiência passada. Ele ressaltou ainda que a necessidade da AM ocorre quando não é possível desenvolver um código diretamente para resolver um problema, e a solução pode ser encontrada em dados passados ou pela experiência.

AM é a busca por algoritmos que produzem hipóteses gerais a partir de informações externas e então fazem previsões sobre acontecimentos futuros (Kotsiantis, 2007, p. 249). Ou seja, o objetivo da AM é construir um modelo conciso de distribuição de rótulos de classe em termos de características preditivas. Kotsiantis (2007, pp. 254–257) descreveu em seu trabalho diferentes tipos de AM e faz uma comparação entre os trabalhos e em como essas técnicas podem ser aplicadas. O autor salientou que a questão chave da classificação com AM não é a superioridade de um algoritmo em relação aos outros, mas sob quais condições um método particular pode ter um desempenho significativamente melhor que outros dado um certo problema de aplicação.

Dois grandes grupos podem ser definidos dentro da AM, os algoritmos supervisionados e os não supervisionados (Kotsiantis, 2007, p. 250), as principais características de cada um são:

- Supervisionado: É dado não somente o conjunto de dados a ser processado pelo algoritmo, mas também qual o resultado esperado após o processamento, são algoritmos usados quando já se tem conhecimento prévio de qual a relação entre as variáveis e qual informação pode ser retirada dessa relação. Entre os algoritmos supervisionados, tem-se dois tipos: regressão, usada em funções contínuas; e classificação, para funções discretas (Kotsiantis, 2007, p. 250);
- Não supervisionado: Para esse algoritmo, não há *feedback*, apenas são fornecidos os dados iniciais e espera-se que o algoritmo encontre a relação entre eles e quais resultados podem ser concluídos a partir de suas ligações. É utilizado quando não se tem conhecimento prévio da correlação entre os dados, mas busca-se investigar qual é o efeito dessas variáveis em conjunto (Kotsiantis, 2007, p. 250).

Langley (2011, p. 275) ao fazer uma análise da evolução das publicações sobre AM entre 1986 e 2011, destacou como a área evoluiu durante esse período e qual direção, na percepção do autor, os estudos dessa área estavam tomando. Ele destacou que no início desse período os estudos sobre AM abordavam avaliação de maneira informal, o que era contraproducente em uma área que buscava um aprimoramento de performance. Essa fase foi seguida pela representação simbólica de conhecimento adquirido como, por exemplo, árvores de decisão e réguas de produção. Langley (2011, p. 277) fez uma crítica à excessiva ênfase dada ao estudo de problemas de classificação e regressão ao invés de atacar problemas mais complexos como raciocínio e compreensão linguística. O autor concluiu que por maiores que tiverem sido as conquistas da área até então, é importante que se recupere a audácia dos primeiros estudos, que eram capazes de visualizar infinitas possibilidades para a AM.

2.2.2 REDES NEURAIS ARTIFICIAIS

O cérebro humano é um sistema de processamento de informação altamente complexo e não linear, funcionando de maneira completamente diferente de um computador comum, e foram essas diferenças que impulsionaram o grande número de estudos sobre RNAs (Haykin, 1998, p. 23). Tais técnicas são processadores paralelos massivos formados por unidades

de processamento simples, que tem uma propensão natural a armazenar conhecimento empírico e disponibiliza-lo ao uso. RNAs são semelhantes ao cérebro de duas formas: conhecimento é adquirido pela rede a partir de seu ambiente por meio de um processo de aprendizagem, e forças de conexão interneurais, chamados de pesos de sinapse, são usados para armazenar o conhecimento adquirido. Haykin (1998, pp. 24–26) elenca ainda as propriedades das RNAs, sendo elas:

- Não linearidade;
- Mapeamento de *input* e *output*, também chamado de amostras de treinamento, consistem de um sinal único de *input* e a resposta desejada correspondente, fazem parte da aprendizagem do sistema para o que o programador deseja obter do processo;
- adaptatividade, RNAs possuem a habilidade de adaptar seus pesos de sinapse de acordo com as mudanças no ambiente onde estão inseridas;
- Resposta empírica, quando em um contexto de classificação de padrões, uma RNAs pode ser desenhada para prover informações não apenas sobre qual padrão particular deve selecionar, mas também sobre quanta confiança pode ser colocada em cada decisão;
- Informação contextual, que é dada pela própria estrutura e estado de ativação de uma RNAs;
- Tolerância a falhas, uma RNAs, quando implementada na forma de *hardware* tem o potencial de apresentar tolerância a falhas, isto é, sua performance decai gradualmente quando operando sob condições adversas;
- uniformidade de análise e *design*, no sentido que os elementos das RNAs possuem a mesma notação em todos os domínios que envolvem a aplicação de RNAs;
- Analogia neurobiológica, o *design* de uma RNAs é motivada pela sua semelhança com o cérebro, o que é testamento que é um processamento rápido e poderoso.

Zhang, Patuwo, e Hu (1998, p. 35), em seu trabalho, destacam que RNAs são poderosas em sua capacidade de classificação e reconhecimento de padrões, já que são capazes de aprender e generalizar a partir da experiência. Os autores ainda definem RNAs como métodos auto adaptativos orientado por dados, já que precisam de poucas hipóteses anteriores a aplicação do modelo, capazes de generalizar, que são aproximadoras universais de função e são não lineares (Zhang et al., 1998, pp. 35–36). O trabalho de Zhang et al. (1998) é centrado no estudo de RNAs multicamada de alimentação direta, que é o mais utilizado e popular paradigma de RNAs em previsão. A Figura 2.2 abaixo é uma ilustração de uma RNAs desse tipo.

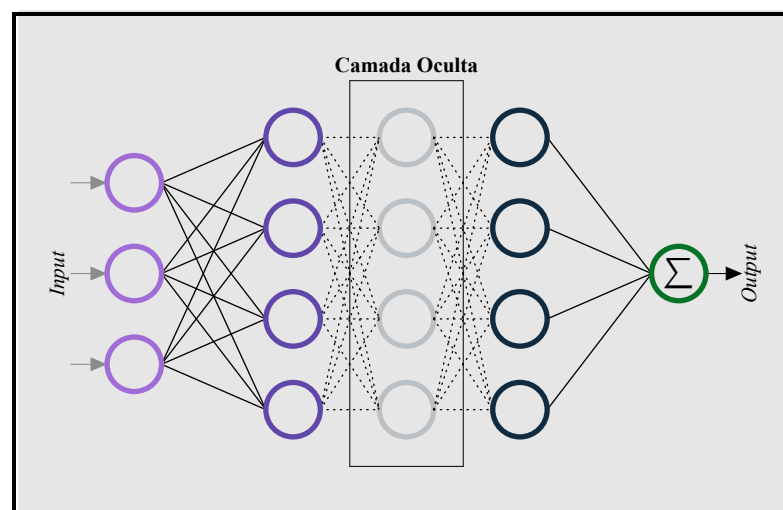


Figura 2.2: Exemplo de RNAs multicamada de alimentação direta.

Essas RNAs são compostas de várias camadas de nodos¹¹, a primeira camada é chamada

¹¹Em RNAs, nodos são os elementos de processamento mais simples da estrutura (Basheer & Hajmeer, 2000, p.

de “camada de *input*” é onde a informação é recebida, são os dados alimentados ao algoritmo; a última camada, chamada “camada de *output*” é onde a solução do problema é obtida (Zhang et al., 1998, p. 38). Entre a camada de *input* e a camada de *output* estão as camadas ocultas, onde de fato ocorre o processamento dos dados, no caso do exemplo acima, são três camadas ocultas para processamento. Ainda de acordo com Zhang et al. (1998, p. 38), a função que representa o relacionamento estimado pela RNAs pode ser dada por:

$$y = f(x_1, x_2, x_3, \dots, x_p) \quad (2.1)$$

onde $x_1, x_2, x_3, \dots, x_p$ são variáveis independentes (dados de *input*) e y é a variável dependente, sendo assim, pode-se fazer um paralelo entre RNAs e modelos de regressão não lineares. No caso de problemas de previsão de dados temporais, como é o utilizado neste trabalho, os *inputs* são geralmente observações passadas e o *output* é um valor do futuro. Nesse caso, a RNAs é representada pela seguinte função (Zhang et al., 1998, p. 38):

$$y_{t+1} = f(y_t, y_{t-1}, y_{t-2}, \dots, y_{t-p}) \quad (2.2)$$

Zhang et al. (1998, pp. 38–39) destacam que, antes de poder realizar qualquer tarefa, as RNAs devem ser treinadas. O treinamento dessas redes consiste na determinação dos pesos de cada arco de ligação entre os nodos e a parcialidade de cada nodo. É dessa maneira que o conhecimento adquirido pelo treinamento é armazenado na RNAs, que é então capaz de mapear diversas funções não lineares. O processo de treinamento segue a seguinte ordem: os exemplos são introduzidos nos nodos de *input*, a partir dos valores de ativação desses nodos são então determinados os pesos, sendo então acumulados em cada nodo; esse total é transformado em uma função de ativação no nodo, que em função disso se torna um *input* para a camada seguinte; assim sucessivamente, até que a última camada resulta no *output* desejado (Zhang et al., 1998).

Dentre as diversas aplicações de RNAs que serão exploradas mais a frente nessa subseção, uma bastante comum é a da previsão de falência de bancos (Tam & Kiang, 1992, p. 926). A RNAs representa uma função não linear discriminante como um padrão de conexões entre unidades de processamento. Em seu trabalho, Tam e Kiang (1992, p. 932) utilizaram dados bancários *default* para comparar RNAs com classificadores lineares, regressão logística, *k-Nearest Neighbors* (*kNN*) e *Iterative Dichotomiser 3* (*ID3*). É constatado que as RNAs representam um método promissor na avaliação de condições bancárias, em termos de acurácia de previsão, adaptabilidade e robustez. Os autores também discutiram as limitações desse método como ferramenta de modulação.

É interessante notar também a relação entre RNAs e métodos estatísticos, já que gradualmente essa técnica de AM vem substituindo os processos estatísticos (Warner & Misra, 1996, p. 284). Para explicar essa relação e fazer uma introdução sobre RNAs para pesquisadores, estatísticos, engenheiros e profissionais com noções de estatística, Warner e Misra (1996, pp. 286–289) fizeram um contraste entre modelos de regressão e a ferramenta de RNAs mais comum, a rede multicamada de alimentação direta treinada utilizando retropropagação. Os autores buscam explicar de maneira descomplicada as RNAs, ao mesmo tempo evitando se aprofundar nas complexidades da técnica e ressaltando a sua importância como ferramenta de análise de dados.

Outro trabalho que estuda as RNAs de rede multicamada de alimentação direta treinada utilizando a retropropagação é o desenvolvido por Yao e Liu (1997, p. 694). Nele, os autores propõem um sistema de evolução para RNAs, onde um algoritmo construtivo simples começa com uma rede mínima e que, de acordo com a necessidade durante o treinamento, adiciona novas camadas, nós e conexões, sendo o contrário também possível, com a eliminação de elementos que estavam sendo subutilizados. Para isso os autores sugerem um algoritmo de evolução de

RNAs que, quando aplicado a problemas comuns na literatura como, por exemplo, paridade, diagnóstico médico, a avaliação de crédito australiana, podem produzir RNAs compactas, com um bom poder de generalização quando comparado a outros algoritmos (Yao & Liu, 1997, pp. 710–712).

Ainda na relação entre RNAs e métodos estatísticos, Zhang et al. (1999, p. 16) ilustram, por meio de uma revisão compreensiva das aplicações de RNAs a relação entre elas e teoria de classificação Bayesiana. Os autores apresentam uma estrutura para auxiliar na compreensão do papel das RNAs nas previsões de falências bancárias. Zhang et al. (1999, p. 20–23) usam um método de validação cruzada para avaliar a variação dentro da amostra de RNAs para previsão de falência bancária. Tendo como ponto de partida uma amostra de 220 organizações, os autores concluem que RNAs são significativamente melhores que modelos de regressão logística na previsão e classificação de estimativas. Destaca-se ainda que, apesar de modelos estatísticos terem uma boa eficiência de maneira geral, eles falham miseravelmente quando os pressupostos estatísticos não são atendidos.

Também sobre a aplicação de RNAs, Rutkowski (2004, p. 811–812) sugeriu uma nova classe de Redes Neurais Artificiais Probabilísticas (RNAPs) capazes de trabalhar em ambientes dinâmicos. Na construção dessa classe, o autor trabalhou em quatro fases, sendo elas: a formulação do problema de classificação de padrão em ambientes dinâmicos como um problema de previsão; a apresentação inédita da definição de RNAPs; a investigação da velocidade de convergência das redes construídas; e o *design* da rede para utilização. O autor ainda destaca que os pressupostos sobre as propriedades das funções devem ser mais “fracos” para que a rede funcione de maneira satisfatória.

Uma das aplicações mais comuns para RNAs é a de previsão de falência e *scoring* de crédito (Tsai & Wu, 2008, p. 2639). Em seu estudo, Tsai e Wu destacaram ainda que técnicas de AM e IA são comuns na tentativa de solucionar esse problema. Um algoritmo de rede de multi-camadas treinada pela retropropagação é a técnica mais utilizada para problemas de tomadas de decisão financeira. Tsai e Wu (2008, pp. 2643–2646) investigaram o desempenho de um classificador de previsões de falência quando comparado a modelos estatísticos. Apesar de procurarem uma resposta definitiva sobre a melhor ferramenta para se prever com relativa precisão a técnica superior em termos de eficiência para casos de previsão, a inconsistência nos resultados não permitiu que se chegasse a uma conclusão final em relação a isso.

CAPÍTULO 3

MÉTODOS

“A regra é clara.”

—Arnaldo César Coelho.

Para a realização das previsões dos resultados dos jogos da Copa do Mundo foi utilizado um código escrito na linguagem Python, a preferência por essa linguagem em detrimento de outras, como, por exemplo, o R, se deu por três motivos principais: a simplicidade da linguagem, a velocidade de processamento e a disponibilidade de pacotes pré-existentes com funções que simplifiquem o código.

O código usado neste trabalho foi adaptado a partir da publicação de Juan Pedro Fisanotti, engenheiro da computação argentino que atualmente trabalha como desenvolvedor de Python como professor na Universidad Católica de Santiago del Estero¹. As adaptações feitas no código se resumem à atualização de comandos para que correspondessem às novas versões dos pacotes empregados e nas modificações dos arquivos de dados utilizados.

No código, o processo de previsão de resultados pode ser dividido em três grandes fases. Na primeira fase os dados foram manipulados, com as pontuações de pódio foram distribuídas, e os resultados dos jogos normalizados. A segunda fase ocorreu a definição da RNAs utilizada e realizou-se também seu treinamento. Na terceira e última fase, a RNAs foi utilizada para realizar as previsões.

RNAs foram escolhidas por serem uma ferramenta consolidada no que diz respeito a previsão de resultados futuros a partir de dados temporais (Haykin, 1998; Zhang et al., 1998, p. 59, pp. 16–17). Além disso, a capacidade das RNAs em modelar relacionamentos não lineares, de generalização e de não restringir as variáveis de *input* tornam essa técnica especialmente apropriada para a previsão de resultados como, por exemplo, em jogos de futebol.

Para a análise de resultados, o processamento da RNAs foi realizado em duas etapas. Na primeira etapa, o código foi processado sem a inclusão de novos dados a cada rodada de jogos da Copa do Mundo, ou seja, foi feita uma previsão, que foi referida como “cega” ao longo deste trabalho, onde a RNAs não aprendeu com os erros antes de fazer as próximas previsões. Na segunda etapa, os dados foram atualizados a cada rodada, esse tipo de previsão foi chamado de “recalibrada”, com a RNAs constantemente aprendendo com os resultados das fases

¹O código foi retirado da página https://github.com/fisadev/world_cup_learning, em 09/05/2018.

anteriores, permitindo que a RNAs faça as previsões conhecendo os erros feitos anteriormente, similarmente ao que ocorre em RNAs supervisionadas.

As previsões feitas nos dois casos foram então comparadas com os resultados reais e com as previsões feitas pelo mercado de apostas e três estatísticas foram extraídas dos resultados, sendo elas:

- A porcentagem de acertos pura (com todos os jogos considerados);
- A porcentagem de acertos líquida (em que os jogos cujo resultado foram empates ou que o código não tinha dados suficientes para realizar a previsão são retirados da contagem);
- A pontuação com critério de conversão, nesse último caso, um acerto conta como +1, um erro como -1 e o empate vale 0, como será melhor explicado na Seção 3.3.

Este capítulo foi organizado da seguinte forma: Na Seção 3.1, o código utilizado para realizar as previsões foi descrito detalhadamente, na Seção 3.2 os dados utilizados para suprir a RNAs foram discutidos e, por fim, na Seção 3.3 foi discutido como a análise desses dados foi realizada.

3.1 O CÓDIGO

Segundo Zobel (2014, p. 147), pseudocódigos são um formalismo utilizado para representar um algoritmo computacional. No pseudocódigo, os algoritmos são apresentados em uma linguagem estruturada em blocos, com todas as linhas numeradas. A vantagem no uso desse tipo específico de formalismo consiste em sua estrutura, que facilita o entendimento do código por pessoas que não estejam familiarizadas com a linguagem de programação, a desvantagem está na perda do detalhamento e nuances do código. Para uma melhor visualização e compreensão do código utilizado neste trabalho e suas partes, o pseudocódigo 1 foi utilizado para representar as principais operações realizadas ao longo do processamento. O código completo pode ser encontrado no Anexo ao final do texto.

Na primeira parte do código são definidos os pacotes de Python que foram utilizados, tanto para a manipulação de dados, quanto para a definição da RNAs e para a marcação de tempo de processamento. É nessa parte também que os dados de jogos passados foram carregados, sendo eles os resultados de todos os jogos de Copa do Mundo a partir de 1950², os pódios de todas as Copas do Mundo no mesmo período e um arquivo com a renomeação das seleções Alemã, Russa e Iraniana³.

Antes de se iniciar o código de fato foi definida uma variável para marcar o momento em que o processamento do código inicia, ao final do código foi então definida uma segunda variável de tempo para capturar o momento em que o código finaliza o processamento. A duração foi então dada ao final do código a partir da subtração entre o momento final e o inicial de processamento. No caso do código utilizado esse tempo varia entre 15 e 20 segundos, o que pode ser considerado ínfimo quando comparado aos anos de experiência que são levados em consideração por uma pessoa ao se tomar uma decisão (Ariely & Zakay, 2001; Klapproth, 2008).

No primeiro bloco de código foram determinados os *inputs* da RNAs o ano da partida, a porcentagem de partidas vencidas pelo time “A”, a pontuação média de pódio do time “A”, a porcentagem de partidas vencidas pelo time “B” e a pontuação média de pódio do time “B”.

²Os dados das Copas do Mundo anteriores a 1950 foram desconsiderados já que o formato da competição era diferente.

³No código foram utilizados os nomes dos países para representar as respectivas seleções, fazendo com que esse ajuste fosse necessário devido às mudanças de denominações pelas quais esses países passaram ao longo dos anos por diversas questões políticas.

Algoritmo I: Pseudocódigo.

Input : Ano da partida, Porcentagem de partidas ganhas - Time A, Média de pontuação de pódio - Time A, Porcentagem de partidas ganhas - Time B, Média de pontuação de pódio - Time B;

Output: Vencedor da partida entre os times A e B;

- 1 Excluir a possibilidade de empate como resultado;
- 2 Duplicar os resultados invertendo as posições dos times;
- 3 Calcular a porcentagem de vitórias de cada time;
- 4 Definir os *inputs* e *outputs* da RNAs;
- 5 Normalizar os dados de *input*;
- 6 Estabelecer uma função para o resultado da RNAs chamada *n*;
- 7 **def** *Resultado da RNAs*:
- 8 | **if** *Output de n for maior ou igual a 0,5 then*
- 9 | Retornar 1
- 10 | **else**
- 11 | Retornar 2
- 12 | **end**
- 13 **def** *Teste da RNAs*:
- 14 | Retornar(Resultados dos acertos do treino da RNAs, Resultados dos acertos dos testes da RNAs)
- 15 Criar um grupo de treinamento para a RNAs;
- 16 Treinar a RNAs dado um número de interações;
- 17 Retornar o *Teste da RNAs* para as interações;
- 18 **def** *Função predict(Ano, Time A, Time B)*:
- 19 | **def** *Função resultado*:
- 20 | **if** *Resultado da RNAs for 1 then*
- 21 | Retornar Time A
- 22 | **else**
- 23 | Retornar Time B
- 24 | **end**
- 25 **predict(Ano, Time A, Time B)**

O vencedor da partida foi determinado como *output*. Para que a RNAs não identifique uma relação entre qual time estava na posição “A” ou “B” e o vencedor da partida, todos os dados foram duplicados com as posições invertidas⁴.

Algumas estatísticas foram então calculadas a partir do conjunto de dados disponibilizados, como porcentagem de vitórias, média do diferencial de gols de cada partida, número de partidas jogadas por cada seleção, número de participações de cada seleção e número de copas vencidas por cada seleção em um código a parte, que também foi exposto no Anexo. além das estatísticas que foram utilizadas de fato como *inputs*, porcentagem de vitórias de cada time e pontuação média de pódio de cada. Essas estatísticas foram importantes para que se possa visualizar alguns dados com mais clareza e, possivelmente, correlacionar esses dados com os resultados encontrados como, por exemplo, se uma seleção com um maior número de participações tem propensão a vencer mais jogos.

Na última parte do preparo dos dados a função *normalize* foi utilizada para normalizar os mesmos. A normalização é o processo de reestruturação dos dados a fim de reduzir a

⁴No futebol, o time na posição “A” é o mandante do jogo, ou seja, é o time que joga em casa, o que pode ter um efeito no resultado das partidas na maioria das competições. Já na Copa do Mundo, a seleção na posição “A” é aquele que pode escolher qual uniforme usar na partida, cujo o impacto não foi estudado neste trabalho.

redundância e melhorar sua integridade (Codd, 1970, p. 377). No caso do código utilizado neste trabalho, a normalização teve como principal função diminuir o impacto de resultados fora do padrão de modo que estes sejam atribuídos pesos menores nas previsões.

Para melhor compreensão do que a normalização fez nesse caso, tem-se como exemplos os times hipotéticos “C”, “D” e “E”. Ambos os times “D” e “E” venceram o time “C” uma vez no passado, mas, enquanto o time “D” tem um saldo positivo na competição (mais vitórias do que derrotas), o time “E” tem um saldo negativo. Nesse caso, a vitória do time “D” tem um maior peso em um futuro jogo contra “C” do que a vitória do time “E”.

Terminada a manipulação de dados, iniciou-se o trabalho com as RNAs. Para isso, primeiro definiu-se a grandeza dos dados que serão utilizados para treinar a RNAs posteriormente, neste caso, 10% dos *inputs* serão aleatoriamente selecionados para o aprendizado da RNAs. Antes porém, uma restrição foi aplicada, empates não puderam ser um dos *outputs* da RNAs. Essa restrição se deve ao uso de uma função sigmoide que é utilizada para a realização da previsão.

A função sigmoide é uma função matemática que tem como característica sua curva em formato de “S”, conforme ilustrado na Figura 3.1, que varia entre os valores 0 e 1, e, por isso, não se pode a partir dela prever um evento em que os *outputs* podem receber mais de dois valores. São geralmente aplicadas quando um modelo matemático específico ao problema não é possível. Existem vários exemplos de função sigmoide, os mais comuns são a função logística, tangente hiperbólica e arco tangente, como apresentadas nas Equações 3.1, 3.2 e 3.3, respectivamente.

$$f(x) = \frac{1}{1 + \exp^{-x}} \quad (3.1)$$

$$f(x) = \tanh(x) = \frac{\exp^x - \exp^{-x}}{\exp^x + \exp^{-x}} \quad (3.2)$$

$$f(x) = \arctan(x) \quad (3.3)$$

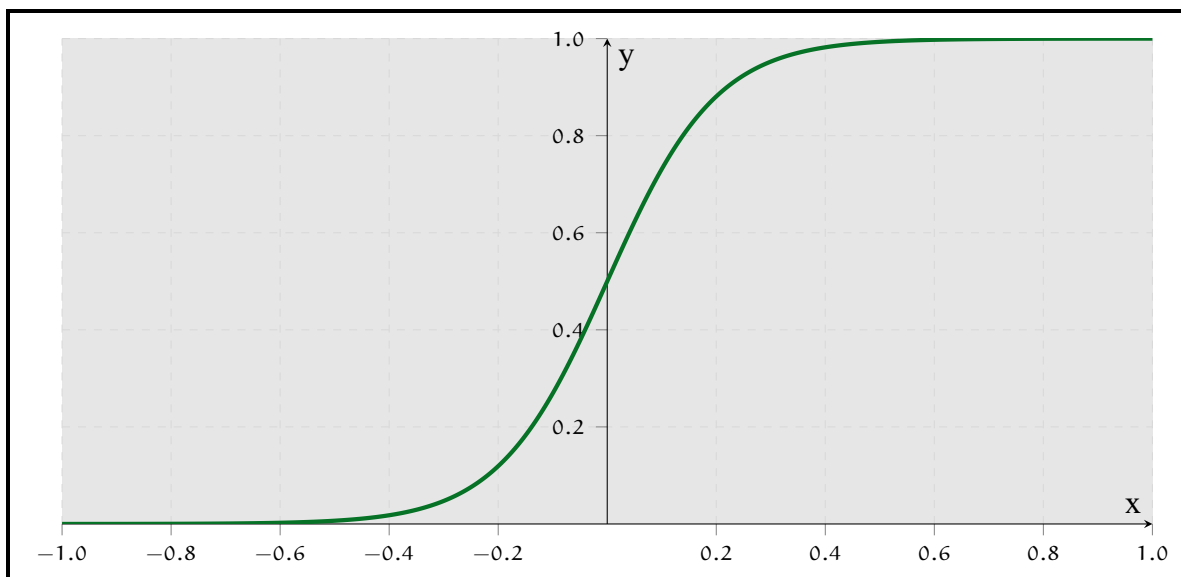


Figura 3.1: Exemplo do gráfico de uma função sigmoide.

Após serem feitas as delimitações dos parâmetros de treinamento, foram definidas as funções de resultado, que definem o *output* final do vencedor do jogo, e de teste dos erros da RNAs, calculando a porcentagem das discrepâncias encontradas no código em relação ao resultado real. A partir dessas funções, cria-se um grupo de teste para treinar a RNAs e mostra-se a acurácia do teste realizado.

Esse passo foi seguido pelo treinamento de fato, em que o usuário pode definir qual o número de interações⁵. Essas interações têm dois *outputs*, a acurácia da RNAs pós-treinamento (acima de 90%) e a acurácia da RNAs pré-treinamento (aproximadamente 30%).

Por fim, foi definida a função que será usada para fazer as previsões, os *inputs* dessa função foram os dados já normalizados e o *output* foi a seleção que, de acordo com os dados passados, tem a maior probabilidade de vencer o jogo. O formato da função de predição é `predict(Ano do jogo, 'Seleção A', 'Seleção B')`, já o *output* é dado como 'Seleção A/B'.

3.2 OS DADOS

Nesta Seção é apresentado como os dados utilizados neste trabalho foram obtidos, tanto com os resultados dos jogos quanto as previsões do mercado de aposta. Também são detalhados os ajustes que foram necessários para o processamento adequado do código, assim como foi feita a geração de dados auxiliares, ambos em um arquivo suplementar ao código principal. Por fim, tratou-se sobre os tipos de simulação realizados para a análise dos dados.

3.2.1 DADOS BRUTOS

Os dados iniciais utilizados para a previsão de resultados foram retirados do mesmo endereço eletrônico do código inicial. Três conjuntos de dados foram utilizados para se realizar a previsão, os resultados das partidas anteriores, as pontuações de pódio e as equivalências entre os diferentes nomes dos países ao longo dos anos. Tanto os arquivos de pontuação de pódio, quanto o com a relação dos nomes dos países não foram alterados durante o processo, enquanto o arquivo de resultados foi atualizado a cada rodada para a análise “recalibrada”.

O arquivo com a relação dos nomes dos países é simples, consistindo em apenas uma pequena tabela com duas colunas, sendo a primeira preenchida com o nome anterior e a segunda com o nome atual do país. Pode-se destacar aqui que a Iugoslávia não aparece entre esses países, como sua separação resultou na criação de diversos Estados diferentes, e vários deles já têm suas conquistas na Copa do Mundo, os dados da Iugoslávia permaneceram sendo considerados da Iugoslávia, sem atualização. Os dados referentes às mudanças de nome são mostrados na Tabela 3.1.

Nome	
Anterior	Atual
União Soviética	Rússia
Alemanha Ocidenta	Alemanha
Alemanha Oriental	Alemanha
República Democrática Alemã	Alemanha
República Federal Alemã	Alemanha
República Islâmica do Irã	Irã
Tchecoslováquia	República Tcheca

Tabela 3.1: Relação de nomes que foram alterados ao longo do tempo.

Já os dados referentes ao pódio foram organizados em uma tabela de três colunas, a primeira diz respeito ao ano, a segunda à posição – até a quarta colocação – e a terceira ao nome do país. Para exemplificar, os dados referentes a Copa do Mundo de 2002 podem ser observados

⁵No desenvolvimento deste trabalho, foram utilizadas 20 interações, número suficiente para conseguir uma acurácia que varia entre 95% e 100% mantendo o tempo médio de processamento do código.

na Tabela 3.2. Na RNAs, utiliza-se a pontuação inversamente à classificação, com o primeiro colocado recebendo quatro pontos, o segundo três e assim sucessivamente.

Ano	Posição	País
1950	1	Uruguai
1950	2	Brasil
⋮	⋮	⋮
2014	3	Holanda
2014	4	Brasil

Tabela 3.2: Exemplo extraído dos dados de pódio.

Por fim, os dados pertinentes aos resultados da partida estão organizados em seis colunas, a primeira é o número do jogo em relação à Copa do Mundo. A segunda coluna é o número de identificação do jogo; A terceira é o ano em que a partida aconteceu; A quarta o nome do time “A”; A quinta o número de gols do time “A”; a sexta o número de gols do time “B” e a sétima o nome do time “B”. Um exemplo desses dados pode ser observado na Tabela 3.3

Número do jogo	Ano	Time A	Gols Time A	Gols Time B	Time B
53	1950	Brasil	4	0	México
54	1950	Inglaterra	2	0	Chile
55	1950	Espanha	3	1	EUA
⋮	⋮	⋮	⋮	⋮	⋮
849	2014	Holanda	0	0	Argentina
850	2014	Brasil	0	3	Holanda
851	2014	Alemanha	1	0	Argentina

Tabela 3.3: Exemplo extraído dos dados de resultados.

Os dados brutos completos que foram utilizados para a realização deste trabalho estão no Anexo. Apesar dos dados iniciais terem sido retirados de um endereço eletrônico já em planilha, os dados da Copa do Mundo de 2018 utilizados para a recalibração da RNAs foram retirados do sítio eletrônico oficial da *FIFA* e, manualmente, adicionados aos originais na planilha.

3.2.2 AJUSTES

Devido à quantidade considerável de dados obtidos, foram necessários dois ajustes principais para que o processamento do código ocorresse de maneira adequada: a normalização dos dados e a retirada do empate não apenas das opções de *output*, mas também das estatísticas calculadas para a previsão.

Para a normalização, foi definida uma função `normalize` que diminui o peso atribuído aos resultados que são considerados fora do comum. Isso foi feito a partir dos dados auxiliares, que serão melhor explicados na subseção seguinte, em que foram calculadas diversas estatísticas dos times, como a porcentagem de vitórias e de participação de cada time. Assim foi possível determinar quais resultados estavam fora do padrão e distribuir os pesos à cada vitória adequadamente.

Essa manutenção dos dados foi necessária para que esses resultados não afetassem de forma desproporcional a previsão dos resultados dos jogos futuros. O resultado de cada jogo são igualmente importantes no cálculo de porcentagem de vitórias de cada time, não havendo diferenciação de peso para a realização desse cálculo. O peso é importante para a comparação de confrontos diretos entre os times.

No arquivo auxiliar “*utils*”, que também pode ser encontrado no Anexo, foram calculadas diversas estatísticas que dão suporte à RNAs para a realização das previsões. São calculados a porcentagem de vitórias de cada time, que é o cálculo bruto de da quantidade de vitórias que um time teve em relação à quantidade de partidas jogadas por ele; a pontuação de pódio de cada seleção, que é a média de pontos que uma seleção tem em cada edição de Copa do Mundo; e a média de vezes em que cada seleção saiu vencedora em relação ao número de Copas em que participou. Esses foram os dados que a RNAs utilizou como *input* para a realização das suas previsões.

Além desses dados, também foram utilizados para a avaliação e análise dos resultados obtidos, dados do mercado de apostas retirados do endereço eletrônico <https://www.oddsportal.com/soccer/world/world-cup-2018/results/#/>. Assim como foi feito com a RNAs, o mercado de apostas faz a previsão “recalibrada”, com os resultados das rodadas anteriores sendo considerados para a previsão da próxima rodada. O número de acertos dessas duas previsões foram comparadas com os acertos da RNAs. Além disso, antes do início da Copa do Mundo, o *site* faz a previsão “cega” de quem será o campeão da Copa do Mundo, nesta edição os favoritos foram Alemanha, Brasil e Espanha.

Para a pontuação de acertos dada à RNAs ou ao mercado de apostas, somente são consideradas as previsões feitas em relação aos jogos reais. Porém, como as diferenças nos resultados dos jogos causam mudanças em como esses times são classificados para as fases seguintes⁶ e mudam o cenário da competição, foi feita também a previsão de como seria a evolução da competição caso as previsões feitas pela RNAs se concretizassem, esse cenário pode ser visto na Figura 3.2.

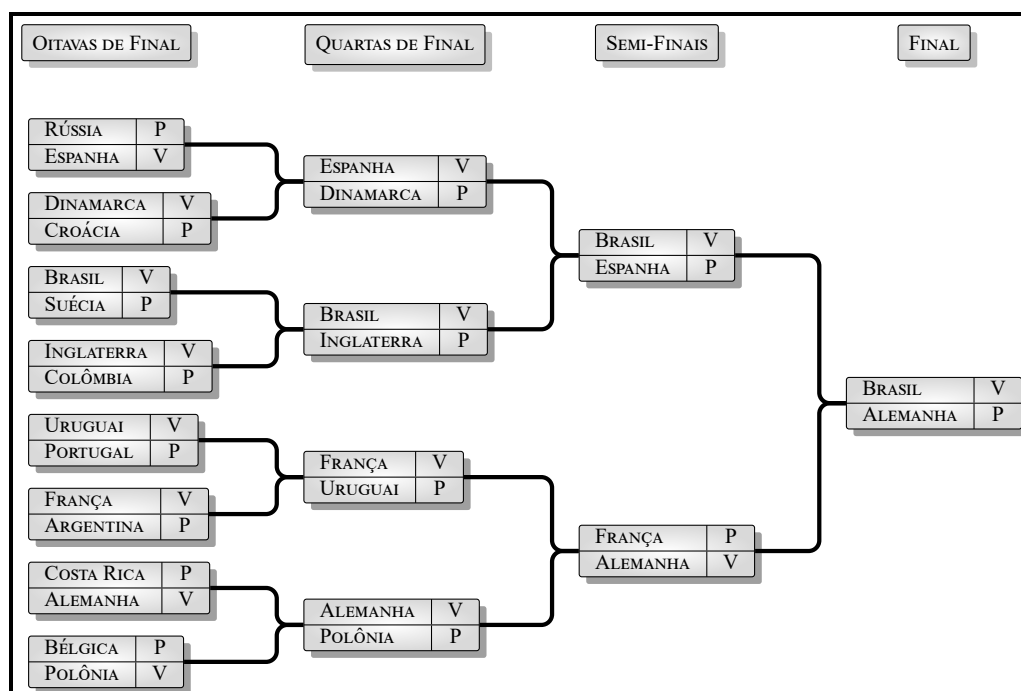


Figura 3.2: Chaveamento conforme a previsão feita pela RNAs “cega”.

Nota: V - Vencedor da partida e P - Perdedor da partida.

⁶Na Copa do Mundo os jogos são divididos em fases, na primeira fase cada vitória/derrota/empate atribui uma pontuação para cada seleção, os dois times com maior pontuação em cada grupo são então classificados para as próximas fases, na qual o time que perde é eliminado, essa fase é chamada de “mata-mata”. A classificação do time em cada fase determina quais serão seus próximos jogos no que é chamado de “chaveamento”.

A análise dos resultados foi realizada em duas etapas, mantendo-se os mesmos passos. Na primeira análise o conjunto de dados utilizado para a previsão comportou os jogos das Copas do Mundo entre 1950 e 2014, mas não foi atualizado no decorrer da competição de 2018. As informações obtidas com essa simulação têm sua importância por destacar qual o nível de acertos do código antes do início do torneio e tentar prever o campeão com antecedência, essa análise será chamada ao longo deste trabalho de “dados à cega” ou dados “cegos”.

Na segunda etapa em que o código foi testado, a cada rodada os dados foram atualizados com os resultados, permitindo assim que a RNAs aprendesse com os erros que cometeu na rodada anterior, e se ajustasse de acordo. Apesar da quantidade de dados ser pequena, principalmente quando comparado ao número de dados que se tem inicialmente (foram adicionados no máximo 16 novas entradas a cada rodada para um total de 800 entradas iniciais), o peso dessas entradas é maior, já que elas são referentes a jogos mais recentes.

A análise feita a respeito desses dois grupos de dados foi a mesma. Na fase de grupos da Copa do Mundo foram comparados os resultados reais, os resultados dados pela previsão do código e os resultados dados pela previsão do mercado de apostas. Para fins de análise foi atribuída uma pontuação para cada situação, conforme indicado na Tabela 3.4. Além disso, foi indicado com uma marcação ao lado dos jogos as partidas que tiveram um jogador expulso, e em qual momento do jogo ocorreu a expulsão.

Essa marcação é importante para que pudesse ser feita uma análise sobre o erro ou acerto da RNAs e do mercado de apostas em relação ao resultado real. Quando as previsões foram realizadas considerava-se que estariam jogando dois times com 11 jogadores, não foi levado em consideração que um time jogou a maior parte do partida com um jogador a menos, que além da desvantagem numérica, ainda causa um maior desgaste dos outros atletas. Foi avaliado então se o erro da previsão pode ou não ser atribuído a esse fato a partir dos resultados dos outros jogos de cada time com 11 jogadores.

Situação	Pontuação
Acerto.	+1
Erro.	-1
Empate/Falta de dados.	0

Tabela 3.4: Pontuação atribuída à comparação entre realidade vs previsão.

A comparação realizada entre a previsão dada pela RNAs e a feita pelo mercado de apostas foi necessária para que se possa contrastar a capacidade da RNAs em fazer previsões a partir de um banco de dados extenso, mas limitado e uma pessoa⁷ que tem acesso às mais diversas informações além dos resultados dos jogos como, por exemplo,

- Análise de jogadores e comissão técnica;
- Conhecimento sobre as tendências dos árbitros da partida;
- Condições climáticas dos jogos;
- Rivalidade entre seleções, o que pode causar uma abordagem diferente por parte dos times;
- Resultados de competições diferentes anteriores à Copa do Mundo.

⁷Em sites de aposta, quem define as probabilidades de vitória é geralmente um *bookmaker*, ou grupo de *bookmakers*, que são especialistas no esporte.

Antes do código ser usado para prever os jogos da fase de mata-mata, foi montada a tabela de classificação da primeira fase, de acordo com os resultados reais, a classificação feita de acordo com os resultados encontrados pela previsão da RNAs e a classificação feita a partir dos resultados previstos pelo mercado de apostas, conforme apresentado na Tabela 3.5.

A	Real	Recalibrado	Cego	MA	B	Real	Recalibrado	Cego	MA
1	Uruguai	Rússia	Rússia	Uruguai	1	Espanha	Portugal	Portugal	Espanha
2	Rússia	Uruguai	Uruguai	Rússia	2	Portugal	Espanha	Espanha	Portugal
3	Arábia Saudita	Arábia Saudita	Árabia Saudita	Egito	3	Irã	Irã	Irã	Irã
4	Egito	Egito	Egito	Arábia Saudita	4	Marrocos	Marrocos	Marrocos	Marrocos
C	Real	Recalibrado	Cego	MA	D	Real	Recalibrado	Cego	MA
1	França	França	Dinamarca	França	1	Croácia	Argentina	Argentina	Argentina
2	Dinamarca	Dinamarca	França	Dinamarca	2	Argentina	Croácia	Croácia	Croácia
3	Peru	Peru	Peru	Peru	3	Nigéria	Nigéria	Nigéria	Islândia
4	Austrália	Austrália	Austrália	Austrália	4	Islândia	Islândia	Islândia	Nigéria
E	Real	Recalibrado	Cego	MA	F	Real	Recalibrado	Cego	MA
1	Brasil	Brasil	Brasil	Brasil	1	Suécia	Alemanha	Alemanha	Alemanha
2	Suíça	Empate	Empate	Sérvia	2	México	Suécia	Suécia	Suécia
3	Sérvia	Empate	Empate	Suíça	3	Coréia do Sul	México	México	México
4	Costa Rica	Empate	Empate	Costa Rica	4	Alemanha	Coréia do Sul	Coréia do Sul	Coréia do Sul
G	Real	Recalibrado	Cego	MA	H	Real	Recalibrado	Cego	MA
1	Bélgica	Inglaterra	Inglaterra	Inglaterra	1	Colômbia	Polônia	Polônia	Colômbia
2	Inglaterra	Bélgica	Bélgica	Bélgica	2	Japão	Colômbia	Senegal	Polônia
3	Tunísia	Tunísia	Tunísia	Tunísia	3	Senegal	Senegal	Colômbia	Senegal
4	Panamá	Panamá	Panamá	Panamá	4	Polônia	Japão	Japão	Japão

Tabela 3.5: Classificação de acordo com os diferentes tipos de previsão.

Essa distinção foi importante para a montagem dos chaveamentos nas fases seguintes, já que a análise foi realizada em três frentes distintas: o chaveamento real, o chaveamento de acordo com o código e o chaveamento de acordo com o mercado de apostas. É importante ressaltar, porém, que a pontuação só foi dada de acordo com os jogos que fazem parte da chave real.

Com a classificação feita, os times foram então divididos em seus devidos chaveamentos, com os primeiros colocados dos grupos A, C, E, G e os segundos colocados dos grupos B, D, F e H no lado 1; e os primeiros colocados dos grupos B, D, F e H com os segundos colocados dos grupos A, C, E e G no lado 2. A partir desse chaveamento, foram feitas as previsões nos três cenários, real, previsão da RNAs e do mercado de apostas, e a cada um foram atribuídas as devidas pontuações.

Além da distribuição da pontuação também são calculadas as porcentagens de acerto, tanto do código quanto do especialista. Esse cálculo também é feito em diversas etapas, tanto por rodada de jogos, quanto por fase e, por fim, calculados os índices de acerto geral. O cálculo da porcentagem foi feito considerando dois totais diferentes, primeiro o cálculo desconsiderando os empates, conforme apresentado na Equação 3.4, e posteriormente o cálculo dos acertos brutos, considerando o número total de jogos, de acordo com a Equação 3.5.

$$P_L = \frac{T_A}{T_J - E} \quad (3.4)$$

$$P_B = \frac{T_A}{T_J} \quad (3.5)$$

Onde P_L e P_B são, respectivamente, a porcentagem líquida e a porcentagem bruta; T_A é o total de acertos, T_J é o número total de jogos e E é a quantidade de empates.

CAPÍTULO 4

RESULTADOS E DISCUSSÃO

“Não existe gol feio, feio é não fazer gol.”

—Dadá Maravilha.

O processo de análise dos resultados foi dividido de acordo com as divisões feitas pela *FIFA* durante a Copa do Mundo, com cada fase – as três primeiras rodadas da primeira fase, as oitavas, quartas e semi finais e os dois jogos da grande final – sendo analisada separadamente. Por fim, é feita uma análise geral dos resultados obtidos, em que se obteve as porcentagens de acerto brutas e líquidas, de acordo com as Equações 3.4 a 3.5.

Em cada fase foram comparados os resultados dados pela RNAs com dados “cegos”, sem atualização a cada rodada, os resultados obtidos pela RNAs utilizando dados “recalibrados”, foram atualizados com os resultados anteriores a cada rodada, e os resultados previstos pelo mercado de apostas.

4.1 FASE DE GRUPOS

Por não terem jogos antes desta rodada, os resultados de ambos tipos de utilização da RNAs, com dados “cegos” e “recalibrados”, são iguais, já que o conjunto de dados iniciais é o mesmo. Pode-se observar, também, que não há previsão para as partidas que envolvem o Panamá e a Islândia a partir da RNAs, já que a edição de 2018 foi a primeira Copa do Mundo de ambas seleções, ou seja, a RNAs não tem dados sobre esses times. Nesse caso, para fins de análise, os resultados dos jogos foram considerados como empates, recebendo pontuação 0, conforme Tabela 3.4.

A primeira rodada é a única em que há uma certa paridade entre os três métodos de previsão, já que não há dados recentes de Copa do Mundo. Destaca-se ainda que no Mercado de Apostas, os *bookmakers*, como são chamados os responsáveis por determinar as chances iniciais de cada jogo, determinam os prováveis vencedores a partir de imensuráveis quantidades de dados, desde competições correlatas como, por exemplo, Eurocopa e Copa América, até quais as prováveis escalas e como está a performance de cada jogador.

Grupo	Jogo	Casa	Visitante	Vencedor - Real	Vencedor - Previsão	Vencedor - Cego	Vencedor - Aposta
A	1	Rússia	Arábia Saudita	Rússia	Rússia	Rússia	Rússia
A	2	Egito	Uruguai	Uruguai	Uruguai	Uruguai	Uruguai
B	3	Marrocos	Irã	Irã	Irã	Irã	Irã
B	4	Portugal	Espanha	Empate	Portugal	Portugal	Espanha
C	5	França	Austrália	França	França	França	França
D	6	Argentina	Islândia	Empate	Islândia não tem dado	Islândia não tem dado	Argentina
C	7	Peru	Dinamarca	Dinamarca	Dinamarca	Dinamarca	Dinamarca
D	8	Croácia	Nigéria	Croácia	Croácia	Croácia	Croácia
E	9	Costa Rica	Sérvia	Sérvia	Sérvia	Sérvia	Sérvia
F	10	Alemanha	México	México	Alemanha	Alemanha	Alemanha
E	11	Brasil	Suíça	Empate	Brasil	Brasil	Brasil
F	12	Suécia	Coréia do Sul	Suécia	Suécia	Suécia	Suécia
G	13	Bélgica	Panamá	Bélgica	Panamá não tem dado	Panamá não tem dado	Bélgica
G	14	Tunísia	Inglaterra	Inglaterra	Inglaterra	Inglaterra	Inglaterra
H	15*	Colômbia	Japão	Japão	Colômbia	Colômbia	Colômbia
H	16	Polônia	Senegal	Senegal	Polônia	Polônia	Polônia

Tabela 4.1: Resultados da primeira rodada.

Nota: *Jogos onde uma das seleções teve um jogador expulso no início da partida.

Conforme indicado na Tabela 4.1, os jogos em que um dos times teve um jogador expulso no início da partida são destacados. Esse destaque foi feito para que se possa observar que nesses jogos todas as previsões estão erradas, o que pode estar relacionado com o fato de que a seleção perdedora teve um jogador a menos durante a maior parte do jogo, e também com o esforço extra necessário do resto do time para compensar a falta de um jogador.

Tipo de Previsão	Número de Jogos	Acertos	Empates	Erros	Nota	% Líquida	% Bruta
Recalibrado	16	9	4	3	6	75,00%	56,25%
Cego	16	9	4	3	6	75,00%	56,25%
Aposta	16	9	4	3	6	75,00%	56,25%

Tabela 4.2: Estatísticas da primeira rodada.

As estatísticas obtidas a partir dos resultados da primeira rodada foram idênticas para os três tipos de previsão. A única previsão em que houve diferença entre os métodos foi no jogo entre Portugal e Espanha, em que o resultado foi um empate, e portanto não foi computado. Pode-se destacar que os três métodos de previsão erraram todos os jogos do grupo “H”, isso pode ter ocorrido já que esse grupo era considerado, antes do início da competição, o mais equilibrado da Copa por ser o único sem uma seleção que já foi campeã.

Também pode-se destacar a derrota da Alemanha, última campeã, para o México, seleção considerada inexpressiva, o que foi uma surpresa, o que é demonstrado pelo fato dos três métodos terem errado a previsão do jogo. Esse resultado pode ser explicado por duas razões distintas, a primeira é a queda de rendimento relacionada a times campeões nas Copas anteriores¹, a segunda é a reconstrução pela qual a seleção alemã passou entre as duas competições, com os principais jogadores se aposentando.

Na segunda rodada, esperava-se uma maior precisão dos resultados dados pelos métodos de RNAs “recalibrado” e pelo mercado de apostas, já que em ambos os casos as previsões são feitas sabendo-se o resultado da rodada anterior, o que melhora os resultados apresentados por esses dois métodos, conforme pode ser observado nas Tabelas 4.3 e 4.4.

Diferentemente de quando a expulsão ocorreu no início da partida, e que pode ser observado um impacto no resultado final, quando a expulsão ocorre ao final do jogo, parece não ter essa influência no resultado final. Além disso, da mesma forma que se espera um resultado bom da Alemanha por ter sido a campeã da última edição, a Argentina também é beneficiada pelos métodos de previsão por ter ficado em segundo lugar.

¹Especialmente os times europeus experimentaram uma queda de rendimento na Copa do Mundo seguinte à vitória.

Grupo	Jogo	Casa	Visitante	Vencedor - Real	Vencedor - Previsão	Vencedor - Cego	Vencedor - Aposta
A	17	Rússia	Egito	Rússia	Rússia	Rússia	Rússia
A	18	Uruguai	Arábia Saudita	Uruguai	Uruguai	Uruguai	Uruguai
B	19	Portugal	Marrocos	Portugal	Portugal	Portugal	Portugal
B	20	Irã	Espanha	Espanha	Espanha	Espanha	Espanha
C	21	Dinamarca	Austrália	Empate	Dinamarca	Dinamarca	Dinamarca
C	22	França	Peru	França	França	França	França
D	23	Argentina	Croácia	Croácia	Argentina	Argentina	Argentina
D	24	Nigéria	Islândia	Nigéria	Nigéria	Islândia não tem dado	Islândia
E	25	Brasil	Costa Rica	Brasil	Brasil	Brasil	Brasil
E	26	Sérvia	Suíça	Suíça	Suíça	Suíça	Sérvia
F	27**	Alemanha	Suécia	Alemanha	Alemanha	Alemanha	Alemanha
F	28	Coréia do Sul	México	México	México	México	México
G	29	Bélgica	Tunísia	Bélgica	Bélgica	Bélgica	Bélgica
G	30	Inglaterra	Panamá	Inglaterra	Inglaterra	Panamá não tem dado	Inglaterra
H	31	Japão	Senegal	Empate	Senegal	Senegal	Senegal
H	32	Polônia	Colômbia	Colômbia	Polônia	Polônia	Colômbia

Tabela 4.3: Resultados da segunda rodada.

Nota: **Jogos onde uma das seleções teve um jogador expulso no final da partida.

Com o maior índice de erros da rodada, o mercado de apostas acreditou que, após um empate com uma das seleções favoritas, Argentina, a Islândia conseguiria uma vitória contra a contestada Nigéria, o que não aconteceu. O mercado de apostas também acreditou que a Sérvia, após uma vitória na primeira rodada, continuaria tendo um bom desempenho. Esse confronto foi marcado por polêmicas extra campo, como o fato de vários jogadores da Suíça são naturais de Kosovo, país não reconhecido pela Sérvia, e o número de jogadores albaneses naturalizados suíços, o que foi criticado por jogadores da Sérvia antes do jogo.

Tipo de Previsão	Número de Jogos	Acertos	Empates	Erros	Nota	% Líquida	% Bruta
Recalibrado	16	12	2	2	10	85,71%	75,00%
Cego	16	11	4	1	10	91,67%	68,75%
Aposta	16	11	2	3	8	78,57%	68,75%

Tabela 4.4: Estatísticas da segunda rodada.

É interessante observar que todos os métodos de previsão apresentaram um melhor desempenho na segunda rodada. Essa melhora já era esperada para a RNAs “recalibrada” e para o mercado de apostas, já que nesses casos são considerados os resultados anteriores, melhorando a precisão das previsões.

O que surpreende é que há uma melhora também nos resultados dados pela RNAs “cega”, o que pode indicar que os resultados dessa rodada foram, de maneira geral, mais condizentes com o que era esperado. Isso significa que o aumento da precisão dos outros dois métodos deve ser interpretado não apenas como resultante de um maior número de dados, mas deve-se considerar também que o número de resultados não esperados foi menor.

A terceira rodada da fase de grupos da Copa do Mundo é sempre uma incógnita. Nesse estágio da competição há três tipos de times: os que já estão classificados e apenas vão decidir a posição de classificação, os que já estão desclassificados, e aqueles que ainda não se classificaram, mas ainda tem chances de passar para a próxima fase. Os dois primeiros grupos são os principais responsáveis pela imprevisibilidade dessa rodada, conforme pode ser observado na Tabela 4.5.

É relevante observar que nos casos de “Empate”, há uma divergência entre as previsões feitas pelos três métodos, o que é um indício da imprevisibilidade desses jogos. Um jogo que merece destaque nessa rodada é a partida entre Inglaterra e Bélgica, em que ambos os times já estavam classificados, mas que o jogo definiu a posição na tabela de classificação de cada um. Das duas chaves, uma era considerada mais “fácil”, e neste caso, o time que perdesse ficaria em uma situação mais confortável durante o mata-mata.

Grupo	Jogo	Casa	Visitante	Vencedor - Real	Vencedor - Previsão	Vencedor - Cego	Vencedor - Aposta
A	33	Arábia Saudita	Egito	Arábia Saudita	Arábia Saudita	Arábia Saudita	Egito
A	34*	Uruguai	Rússia	Uruguai	Rússia	Rússia	Uruguai
B	35	Irã	Portugal	Empate	Portugal	Portugal	Portugal
B	36	Espanha	Marrocos	Empate	Espanha	Espanha	Espanha
C	37	Austrália	Peru	Peru	Peru	Peru	Peru
C	38	Dinamarca	França	Empate	França	Dinamarca	França
D	39	Nigéria	Argentina	Argentina	Argentina	Argentina	Argentina
D	40	Islândia	Croácia	Croácia	Croácia	Islândia não tem dado	Croácia
E	41	Suíça	Costa Rica	Empate	Costa Rica	Costa Rica	Suíça
E	42	Sérvia	Brasil	Brasil	Brasil	Brasil	Brasil
F	43	México	Suécia	Suécia	Suécia	Suécia	Suécia
F	44	Coréia do Sul	Alemanha	Coréia do Sul	Alemanha	Alemanha	Alemanha
G	45	Inglaterra	Bélgica	Bélgica	Inglaterra	Inglaterra	Inglaterra
G	46	Panamá	Tunísia	Tunísia	Panamá	Panamá não tem dado	Tunísia
H	47	Senegal	Colômbia	Colômbia	Colômbia	Senegal	Colômbia
H	48	Japão	Polônia	Polônia	Polônia	Polônia	Polônia

Tabela 4.5: Resultados da terceira rodada.

Nota: *Jogos onde uma das seleções teve um jogador expulso no início da partida.

Tipo de Previsão	Número de Jogos	Acertos	Empates	Erros	Nota	% Líquida	% Bruta
Recalibrado	16	8	4	4	4	67,00%	50,00%
Cego	16	6	6	4	2	60,00%	37,50%
Aposta	16	9	4	3	6	75,00%	56,25%

Tabela 4.6: Estatísticas da terceira rodada.

Na Tabela 4.6, vê-se a brusca queda de rendimento dos três métodos nessa rodada, principalmente na RNAs “cega”. A menor queda foi a do mercado de apostas, fenômeno que pode ser explicado pela compreensão que os *bookmakers* têm da situação do campeonato e das necessidades de cada time. Outro fator que afetou o desempenho dos métodos de previsão foi o alto número de empates nessa rodada.

4.2 MATA-MATA

Nesta fase da competição, a pressão nos times é muito maior, já que a derrota significa eliminação da competição. A partir desse momento, não importam apenas as habilidades de cada time, mas o fator psicológico dos jogadores ao enfrentar uma decisão também é importante nesse estágio. Pode-se observar na Tabela 4.7 que as previsões foram heterogêneas, com alguns resultados inesperados.

Jogo	Casa	Visitante	Vencedor - Real	Vencedor - Previsão	Vencedor - Cego	Vencedor - Aposta
49	Uruguai	Portugal	Uruguai	Portugal	Portugal	Portugal
50	França	Argentina	França	Argentina	Argentina	França
51	Brasil	México	Brasil	Brasil	Brasil	Brasil
52	Bélgica	Japão	Bélgica	Bélgica	Bélgica	Bélgica
53*	Espanha	Rússia	Rússia	Espanha	Espanha	Espanha
54*	Croácia	Dinamarca	Croácia	Croácia	Dinamarca	Croácia
55*	Suécia	Suíça	Suécia	Suécia	Suíça	Suíça
56*	Colômbia	Inglaterra	Inglaterra	Inglaterra	Inglaterra	Inglaterra

Tabela 4.7: Resultados das oitavas de final.

Nota: *Jogos em que uma das seleções teve um jogador expulso no início da partida.

*Jogos em que houveram cobranças de pênalti.

Na fase de mata-mata, outro fator deve ser considerado: a possibilidade de haver disputa por pênaltis, que foi o caso dos jogos 53, 54 e 56. A previsão desses resultados é mais difícil, já que o resultado do jogo é um empate e o vencedor definido por quem consegue marcar mais

pênaltis no adversário, o que é mais relacionado às habilidades pessoais de cada jogador do que ao habitual desempenho do time.

Tipo de Previsão	Número de Jogos	Acertos	Erros	Nota	% Líquida	% Bruta
Recalibrado	8	5	3	2	62,50%	62,50%
Cego	8	4	4	0	50,00%	50,00%
Aposta	8	6	2	4	75,00%	75,00%

Tabela 4.8: Estatísticas das oitavas de final.

Na Tabela 4.8 o impacto dessa incerteza relacionada às cobranças de pênalti, pode ser observado, os três métodos têm índices de acerto menores do que as duas primeiras rodadas, mas ainda assim melhores do que o da terceira rodada. Nas oitavas de final também pode ser percebido o impacto que uma expulsão tem no resultado de um jogo, na partida entre Suíça e Suécia, a equipe com o jogador expulso foi também a eliminada da competição.

Jogo	Casa	Visitante	Vencedor - Real	Vencedor - Previsão	Vencedor - Cego	Vencedor - Aposta
57	Uruguai	França	França	França	França	França
58	Brasil	Bélgica	Bélgica	Brasil	Brasil	Brasil
59	Suécia	Inglaterra	Inglaterra	Inglaterra	Inglaterra	Inglaterra
60*	Rússia	Croácia	Croácia	Rússia	Croácia	Croácia

Tabela 4.9: Resultados das quartas de final.

Nota: *Jogos em que houveram cobranças de pênalti.

Nas quartas de final a necessidade de cobrança de pênaltis também deve ser levada em conta para análise das previsões, já que o jogo entre Croácia e Rússia foi decidido dessa maneira. Outro ponto de destaque nessa rodada foi a eliminação do Brasil, que era, assim como a Alemanha e a Espanha, favorita para vencer essa edição. Diversos fatores podem ser considerados como responsáveis pela derrota, desde a questionável escalação dos jogadores até a baixa eficiência do ataque brasileiro.

Tipo de Previsão	Número de Jogos	Acertos	Erros	Nota	% Líquida	% Bruta
Recalibrado	4	2	2	0	50,00%	50,00%
Cego	4	3	1	2	75,00%	75,00%
Aposta	4	3	1	2	75,00%	75,00%

Tabela 4.10: Estatísticas das quartas de final.

Como o número de jogos em cada fase diminui progressivamente, o impacto de um erro de previsão é maior nas estatísticas de cada método a medida que a competição avança. No caso das quartas de final, esse fenômeno pôde ser verificado principalmente quando observa-se o resultado obtido pelo método “recalibrado” ser muito abaixo do que o método obteve nas fases anteriores. Outro fato curioso dessa fase é o melhor resultado de porcentagem bruta do método “cego” em relação ao “recalibrado”, sendo a primeira vez que isso ocorreu em todo o teste.

Nas semi finais, a vitória da França sobre a Bélgica era esperada, não apenas pela tradição da seleção francesa, mas também pela insurgência da seleção belga que começou em 2010, mas cujo último título foi em nos Jogos Olímpicos de 1920. No outro jogo, a Inglaterra foi mais favorecida por ter um melhor histórico em Copas do Mundo, mas também, no caso do mercado de apostas, o fato de que a seleção Croata vir de duas decisões por pênalti consecutivas, ou seja, jogou duas prorrogações a mais que o time inglês, resultando em 60 minutos a mais de jogo, e acredita-se que em um maior cansaço dos jogadores.

Destaca-se do desempenho nessa fase a RNAs “recalibrada” foi perfeito, acertando o resultado de ambos os jogos. Pode-se dizer que esse desempenho está relacionado à atualização de dados, já que quando a previsão foi feita “às cegas”, o resultado foi diferente. Novamente

Jogo	Casa	Visitante	Vencedor - Real	Vencedor - Previsão	Vencedor - Cego	Vencedor - Aposta
61	França	Bélgica	França	França	França	França
62 [†]	Croácia	Inglaterra	Croácia	Croácia	Inglaterra	Inglaterra

Tabela 4.11: Resultados das semi-finais.

Nota: [†]Jogo em que houve prorrogação.

Tipo de Previsão	Número de Jogos	Acertos	Erros	Nota	% Líquida	% Bruta
Recalibrado	2	2	0	2	100%	100%
Cego	2	1	1	0	50,00%	50,00%
Aposta	2	1	1	0	50,00%	50,00%

Tabela 4.12: Estatísticas das semi-finais.

deve-se levar em consideração que graças ao pequeno número de jogos nessa fase, os percentuais são distorcidos fazendo com que os erros pareçam ter um peso maior.

Jogo	Casa	Visitante	Vencedor - Real	Vencedor - Previsão	Vencedor - Cego	Vencedor - Aposta
63	Bélgica	Inglaterra	Bélgica	Bélgica	Inglaterra	Bélgica
64	França	Croácia	França	França	França	França

Tabela 4.13: Resultados das finais.

A disputa pelo terceiro lugar é um jogo difícil de se prever do ponto de vista do mercado de apostas, já que é uma disputa praticamente simbólica, em que os times não têm muita motivação em vencer. No caso das RNAs, o fato do conjunto de dados do método “recalibrado” incluir um confronto direto entre as duas seleções, com vitória belga, influencia na previsão feita. Já em relação a final, a tradição francesa em Copas do Mundo foi importante para as previsões, mas, além disso, com a prorrogação na semi-final, os croatas jogaram um total de 90 minutos a mais que a seleção francesa, o equivalente a um jogo completo, e considerando que a seleção croata teve uma das maiores médias de idade da Copa (27,8) e a francesa uma das menores (26), o que leva a suposição do condicionamento físico como fator decisivo da partida.

Tipo de Previsão	Número de Jogos	Acertos	Erros	Nota	% Líquida	% Bruta
Recalibrado	2	2	0	2	100%	100%
Cego	2	1	1	0	50,00%	50,00%
Aposta	2	2	0	2	100%	100%

Tabela 4.14: Estatísticas das finais.

Nesta fase fica clara a importância de se ter os dados anteriores recentes para a realização de previsões precisas. Nos dois casos em que a previsão foi feita a partir do conhecimento dos resultados anteriores, os métodos de previsão obtiveram 100% de acerto, enquanto no método cujos dados não foram atualizados, a taxa de acerto foi de 50%.

4.3 RESULTADOS FINAIS

Por fim, foram comparados os resultados totais dos três métodos de previsão com o objetivo de se definir qual é o mais eficaz. Para isso, os resultados de cada fase foram somados e suas devidas estatísticas globais calculadas, conforme exibido na Tabela 4.15.

Um ponto a ser destacado sobre a Tabela 4.15 é a grande diferença entre os resultados obtidos utilizando o método “cego” com aqueles obtidos pelo método “recalibrado”. Com a

Tipo de Previsão	Número de Jogos	Acertos	Empates	Erros	Nota	% Bruta	% Líquida
Recalibrado	64	40	10	14	26	63,00%	74,07%
Cego	64	35	14	15	20	55,00%	70,00%
Aposta	64	41	10	13	28	64,00%	75,93%

Tabela 4.15: Estatísticas globais.

RNAs aprendendo com os erros anteriores e, conseqüentemente, fazendo os devidos ajustes para corrigi-los, a eficiência do método subiu 8% em termos de porcentagem bruta, um número alto ao considerar-se que a diferença entre o método “recalibrado” e o mercado de apostas é de apenas 1%.

Este é, inclusive, o outro ponto que merece ser evidenciado. O mercado de apostas, de acordo com Stekler et al. (2010), é o padrão a ser comparado quando se trata de previsão de jogos de futebol. Que a RNAs proposta tenha obtido resultados apenas 1% menos precisos que o mercado de apostas é um bom indicativo para a eficácia do método. Isso porque, as previsões feitas pela RNAs têm como base de dados apenas os resultados de jogos da Copa do Mundo de edições anteriores, além da atualização com os novos dados dessa Copa.

Já o mercado de apostas não apenas leva em consideração jogos de Copa do Mundo, mas ele também tem acesso às escalafões de cada time, aos resultados de jogos de outras competições internacionais – como Copa América e Eurocopa, informações sobre os árbitros da partida – que, como foi visto nas Tabelas 4.1, 4.5 e 4.1, podem afetar o resultado de uma partida – sobre as condições financeiras de cada seleção, entre outros infinitos dados que chegam às pessoas responsáveis pela definição das chances diariamente.

CAPÍTULO 5

CONCLUSÃO

“Fiz o gol com a minha cabeça e com a
mão de Deus.”
—Diego Maradona.

Em sua grande popularidade, o futebol tornou-se tema de interesse não apenas de fãs, mas também de investidores e do mercado de apostas. Em um mercado que movimenta anualmente entre R\$ 455 bilhões e R\$ 577 bilhões em uma escala global (Oswald, 2014), todo tipo de conhecimento extra é uma vantagem competitiva a ser examinada.

Dentro do esporte, a competição mais importante, e que movimenta o maior volume de dinheiro, é a Copa do Mundo *FIFA* realizada a cada quatro anos. Nela, 32 times disputam o título de campeão ao longo de um mês¹, além de uma premiação em dinheiro. Só a competição movimenta, em média, R\$20 bilhões para o país sede (*FIFA*, 2014) entre aumento no turismo e a captação de investimento por parte de empresas internacionais.

Apesar da Copa do Mundo, em geral, e a *FIFA*, em particular, serem responsáveis por uma significativa parte da movimentação de dinheiro relacionada ao futebol, a maior parcela dos valores movimentados relacionados ao esporte é ligada aos clubes. Desde transferência de atletas entre times até os contratos de patrocínio pago pelas empresas e o dinheiro arrecadado de torcedores, seja pela venda de produtos ou de ingressos, o chamado “mercado da bola” merece atenção.

Uma das maneiras possíveis de se obter uma vantagem competitiva no futebol é ter uma forma de prever os resultados dos jogos. Para empresas patrocinadoras, essa previsão pode auxiliar na escolha de em que time investir, para a *FIFA* qual país precisa de investimento e para o indivíduo, para ajudá-lo a apostar contra o mercado, por exemplo.

Existem várias ferramentas para realizar a previsão de resultados de jogos de futebol, desde a consulta a especialistas até métodos mais técnicos, como a distribuição de Poisson e a regressão linear. Apesar de ser um método cada vez mais utilizado para previsões em geral, a IA ainda não é amplamente utilizada na previsão esportiva, sendo uma área de aplicação dessas ferramentas que está sendo expandida agora (Groll et al., 2018).

¹A partir de 2026 o formato da Copa do Mundo mudará, serão 64 seleções na competição que se estenderá a dois meses.

O objetivo deste trabalho foi a proposição de uma RNAs que seja capaz de realizar previsões de resultados de jogos de futebol da Copa do Mundo com um índice de acertos maior que aquele oferecido pelo mercado de apostas. O método apresentado neste trabalho pode ser utilizado em outros campeonatos e esportes, bastando a adaptação dos dados a serem alimentados na RNAs de acordo com a necessidade e vontade do usuário.

O método utilizado para atingir esse objetivo pode ser dividido em três partes. Na primeira parte o código em Python foi apresentado e cada função sua é explicada. Em um segundo momento, discutiu-se os dados utilizados para as previsões, qual sua origem, como serão utilizados e quais manipulações foram necessárias. Por fim, foi apresentado como foi realizada a análise dos resultados obtidos.

A análise de resultados comparou os resultados reais com as previsões dadas por três métodos diferentes: no primeiro a RNAs foi alimentada com dados que foram chamados de “cegos”, isto é, os dados iam até a final da Copa do Mundo de 2014 e não eram atualizados a cada rodada; no segundo método, a RNAs foi alimentada com dados chamados de “recalibrados”, já que a cada nova rodada eram introduzidos os novos dados gerados; no terceiro método foram extraídos dados de um site de apostas para se comparar com os resultados obtidos pela RNAs.

A análise de resultados mostrou ainda a superioridade do método “recalibrado” quando comparado ao método “cego”, resultado esperado, já que no primeiro a RNAs aprende com seus erros em rodadas anteriores e é reajustada para oferecer melhores previsões, enquanto no segundo as previsões são feitas sem relação com as rodadas anteriores. A diferença entre os dois métodos foi 63% de acerto bruto para o método “recalibrado” e 55% para o método “cego”.

Já em relação ao mercado de apostas, que foi de fato o objetivo desse trabalho, a RNAs “recalibrada” teve um desempenho um pouco abaixo do esperado, já que o mercado de apostas teve um percentual bruto de acerto de 64%. Essa diferença, na verdade, representa um acerto a mais que o mercado de apostas teve em relação a RNAs, ou seja, apenas um jogo.

Considerando as diferentes quantidades de informação envolvidas nesses dois métodos, essa diferença mostra que a RNAs, comparativamente, teve um desempenho melhor do que o mercado de apostas. Para a RNAs, toda previsão feita era baseada única e exclusivamente nos resultados de jogos passados em Copas do Mundo, não foram fornecidos resultados de outras competições como, por exemplo, Eliminatórias da Copa² e Copa América.

Já o mercado de apostas teve acesso a uma diversidade e quantidade muito superior de informações. Todos os resultados de todas as competições internacionais estavam disponíveis, sabia-se o desempenho individual de cada jogador das seleções, qual seriam as condições climáticas em cada jogo³, era conhecido quem eram os juizes em cada partida⁴, o mercado de apostas também sabia quando um jogador de um time se lesionava e se ele jogaria a próxima fase, entre outros inúmeros dados aos quais o mercado de aposta tinha acesso e não foram fornecidos à RNAs.

Portanto, levando em consideração a disparidade nas informações iniciais que cada um dos métodos tinha e a diferença entre o desempenho entre eles, pode-se dizer que o desempenho alcançado pela RNAs é satisfatório não apenas quando comparado com o mercado de apostas nessa situação, mas principalmente ao considerar que o desempenho do mercado de apostas é, em média, de 62,5% (Stekler et al., 2010, p. 615).

No entanto, exatamente pelo fato dos resultados encontrados terem sido positivos, não é absurdo concluir que caso a RNAs tivesse sido alimentada com um maior número e variedade de dados, o desempenho da mesma seria superior. O problema na utilização de dados exclusi-

²Eliminatórias são competições continentais que ocorrem antes da Copa do Mundo em que todas as seleções que desejem se classificar devem competir.

³Em dias muito quentes, seleções de países nórdicos como, por exemplo, Islândia e Suécia, podiam ter suas performances afetadas.

⁴Alguns juizes tem uma maior tendência a expulsar jogadores, e algumas seleções são conhecidas por terem jogadores mais propensos a serem expulsos, o que pode afetar o resultado dos jogos.

vamente da Copa do Mundo é que isso significa que os dados estão espaçados em períodos de 4 em 4 anos, e, mesmo assim, no caso de seleções que não são consistentemente classificadas para a Copa do Mundo, esses dados são ainda mais antigos⁵ ou no caso de seleções que fazem suas estreias na competição⁶, não há dados para previsão.

Uma forma de corrigir essa lacuna seria o acréscimo dos resultados das Eliminatórias, que têm participação de todas as seleções da Copa do Mundo e geralmente as escalações das seleções são similares às utilizadas no torneio, de jogos Amistosos que traria resultados de jogos das seleções do período entre-Copas e de jogos entre times de continentes diferentes, resultados da Copa das Confederações⁷, que apesar de ser um torneio com poucos participantes, pode ser útil para a previsão por serem recentes às Copas do Mundo.

Uma crítica frequentemente feita a utilização de métodos matemáticos e computacionais para se prever resultados de jogos de futebol é que essas ferramentas desconsideram o fator humano do esporte. Não é possível quantificar a vontade ou paixão dos jogadores, nem tentar moldar o futebol como algo puramente racional, Vicente Verdun, escritor mexicano, defende que

Se o futebol estivesse baseado na razão, não haveria razão para existir o futebol.

Mas, a partir do que é possível ser mensurado, existem diversas técnicas de previsão que podem ser empregadas para se obter resultados de partidas futuras. Entre elas, pode-se destacar o *deep learning*, técnica de AM baseada na representação de dados de aprendizado, ao invés de algoritmos específicos a tarefa, isto é a partir do conjunto de dados, o algoritmo automaticamente descobre as representações necessárias para detectar características ou classificações de dados puros (LeCun, Bengio, & Hinton, 2015; Schmidhuber, 2015). Essa técnica é considerada uma evolução das RNAs e apresenta, em geral, resultados mais precisos (LeCun et al., 2015).

Para trabalhos futuros, além da incorporação de resultados de outras competições, seria interessante estudar como a performance da RNAs evolui caso seja alimentada com diferentes tipos de dados, não apenas resultados dos jogos. Um dado que poderia ser incorporado na previsão é a nota de cada jogador convocado aribuída pela *FIFA*⁸ e as convocações de cada seleção, isso ajudaria a ter uma noção mais realista de cada seleção de acordo com a qualidade dos jogadores atuais, ou invés de resultados passados.

Outro estudo que pode ser feito é sobre como a mudança no formato da competição que passará a valer a partir de 2026 impactará na acurácia desses métodos de previsão, principalmente considerando que a competição terá o dobro de seleções participando e que muitas dessas estarão competindo pela primeira vez, ou seja, não teriam dados suficientes para realizar muitas previsões da primeira rodada, ao mesmo tempo que aumentaria o número de jogos nas rodadas subsequentes, como isso afetaria o desempenho da RNAs?

A RNAs usada para previsão da Copa do Mundo poderia também ser aplicada a outras competições, não se restringindo apenas a disputa entre seleções. No caso do Brasil, a RNAs podia ser usada para prever os resultados do Brasileirão ou da Copa do Brasil e até mesmo em escalas menores, para prever os vencedores das partidas dos campeonatos estaduais. E não apenas no Brasil, mas em qualquer campeonato nacional e regional ao redor do mundo.

Uma modalidade do futebol em que as RNAs poderiam ser utilizadas é o futebol feminino. A categoria feminina do esporte tem ganhado cada vez mais destaque, e é uma boa oportunidade para a aplicação dessa ferramenta já que a quantidade reduzida de jogos e, conse-

⁵O Peru é um exemplo, já que a última vez que jogou a Copa do Mundo antes de 2018 foi em 1982.

⁶Caso da Islândia e do Panamá em 2018.

⁷Competição organizada pela *FIFA* que ocorre no ano anterior à Copa do Mundo entre 8 times das 6 federações – um time de cada federação, a seleção do país sede e o último campeão.

⁸Para seu jogo de futebol de realidade virtual, a *FIFA* atribui uma nota a cada jogador baseado em suas habilidades.

quentemente, de dados, tornaria a aprendizagem, da RNAs mais fácil.

Além do futebol, a RNAs também tem aplicações em outros esportes, no tênis e no vôlei, por exemplo, o número de *sets* vencidos poderia ser usado como placar sem nenhuma alteração no formato dos dados. Esportes americanos⁹, que possuem uma grande audiência mundialmente, e aos quais a RNAs poderia ser aplicada diretamente.

⁹Futebol Americano, Basquete, Beisebol e Hóquei no gelo.

ANEXO

Neste anexo estão representados os arquivos completos de dados utilizados pela RNAs, além dos códigos integrais em Python que foram utilizados para a realização desse trabalho. A Tabela 1 representa todos os resultados de jogos anteriores que são considerados para a previsão¹⁰, as últimas linhas em cor **verde** são os dados que foram adicionados, posteriormente, para que se pudesse realizar as previsões “recalibradas”.

Número do Jogo	Ano	Time A	Gols Time A	Gols Time B	Time B
53	1950	Brasil	4	0	México
54	1950	Inglaterra	2	0	Chile
55	1950	Espanha	3	1	EUA
56	1950	Suécia	3	2	Itália
57	1950	Iugoslávia	3	0	Suíça
58	1950	Brasil	2	2	Suíça
59	1950	Iugoslávia	4	1	México
60	1950	Espanha	2	0	Chile
61	1950	Suécia	2	2	Paraguai
62	1950	EUA	1	0	Inglaterra
63	1950	Brasil	2	0	Iugoslávia
64	1950	Espanha	1	0	Inglaterra
65	1950	Itália	2	0	Paraguai
66	1950	Uruguai	8	0	Bolívia
67	1950	Suíça	2	1	México
68	1950	Chile	5	2	EUA
69	1950	Uruguai	2	2	Espanha
70	1950	Brasil	7	1	Suécia
71	1950	Brasil	6	1	Espanha
72	1950	Uruguai	3	2	Suécia
73	1950	Suécia	3	1	Espanha
74	1950	Uruguai	2	1	Brasil
75	1954	Uruguai	2	0	Tchecoslováquia
76	1954	Áustria	1	0	Escócia
77	1954	Brasil	5	0	México
78	1954	Iugoslávia	1	0	França
79	1954	Alemanha FR	4	1	Turquia
80	1954	Hungria	9	0	Coreia do Sul
81	1954	Inglaterra	4	4	Bélgica
82	1954	Suíça	2	1	Itália
83	1954	Uruguai	7	0	Escócia
84	1954	Áustria	5	0	Tchecoslováquia
85	1954	França	3	2	México
86	1954	Brasil	1	1	Iugoslávia
87	1954	Hungria	8	3	Alemanha FR
88	1954	Turquia	7	0	Coreia do Sul
89	1954	Inglaterra	2	0	Suíça
90	1954	Itália	4	1	Bélgica
91	1954	Alemanha FR	7	2	Turquia
92	1954	Suíça	4	1	Itália
93	1954	Uruguai	4	2	Inglaterra
94	1954	Áustria	7	5	Suíça

Continua.

¹⁰As primeiras 52 partidas jogadas em Copa do Mundo foram descartadas, por conta da diferença entre o formato antigo e o atual da Copa do Mundo.

Número do Jogo	Ano	Time A	Gols Time A	Gols Time B	Time B
95	1954	Alemanha FR	2	0	Iugoslávia
96	1954	Hungria	4	2	Brasil
97	1954	Alemanha FR	6	1	Áustria
98	1954	Hungria	4	2	Uruguai
99	1954	Áustria	3	1	Uruguai
100	1954	Alemanha FR	3	2	Hungria
101	1958	Suécia	3	0	México
102	1958	União Soviética	2	2	Inglaterra
103	1958	Argentina	1	3	Alemanha FR
104	1958	Hungria	1	1	Wales
105	1958	França	7	3	Paraguai
106	1958	Iugoslávia	1	1	Escócia
107	1958	Brasil	3	0	Áustria
108	1958	Irlanda do Norte	1	0	Tchecoslováquia
109	1958	Brasil	0	0	Inglaterra
110	1958	México	1	1	Wales
111	1958	Paraguai	3	2	Escócia
112	1958	Iugoslávia	3	2	França
113	1958	União Soviética	2	0	Áustria
114	1958	Alemanha FR	2	2	Tchecoslováquia
115	1958	Argentina	3	1	Irlanda do Norte
116	1958	Suécia	2	1	Hungria
117	1958	Suécia	0	0	Wales
118	1958	Brasil	2	0	União Soviética
119	1958	Alemanha FR	2	2	Irlanda do Norte
120	1958	Hungria	4	0	México
121	1958	Paraguai	3	3	Iugoslávia
122	1958	França	2	1	Escócia
123	1958	Inglaterra	2	2	Áustria
124	1958	Tchecoslováquia	6	1	Argentina
125	1958	União Soviética	1	0	Inglaterra
126	1958	Irlanda do Norte	2	1	Tchecoslováquia
127	1958	Wales	2	1	Hungria
128	1958	Brasil	1	0	Wales
129	1958	Alemanha FR	1	0	Iugoslávia
130	1958	Suécia	2	0	União Soviética
131	1958	França	4	0	Irlanda do Norte
132	1958	Suécia	3	1	Alemanha FR
133	1958	Brasil	5	2	França
134	1958	França	6	3	Alemanha FR
135	1958	Brasil	5	2	Suécia
136	1962	Uruguai	2	1	Colômbia
137	1962	Brasil	2	0	México
138	1962	Argentina	1	0	Bulgária
139	1962	Chile	3	1	Suíça
140	1962	União Soviética	2	0	Iugoslávia
141	1962	Tchecoslováquia	1	0	Espanha
142	1962	Hungria	2	1	Inglaterra
143	1962	Alemanha FR	0	0	Itália
144	1962	Iugoslávia	3	1	Uruguai
145	1962	Brasil	0	0	Tchecoslováquia
146	1962	Inglaterra	3	1	Argentina
147	1962	Chile	2	0	Itália
148	1962	União Soviética	4	4	Colômbia
149	1962	Espanha	1	0	México
150	1962	Hungria	6	1	Bulgária
151	1962	Alemanha FR	2	1	Suíça
152	1962	União Soviética	2	1	Uruguai
153	1962	Brasil	2	1	Espanha
154	1962	Hungria	0	0	Argentina
155	1962	Alemanha FR	2	0	Chile
156	1962	Iugoslávia	5	0	Colômbia
157	1962	México	3	1	Tchecoslováquia
158	1962	Inglaterra	0	0	Bulgária
159	1962	Itália	3	0	Suíça
160	1962	Chile	2	1	União Soviética
161	1962	Brasil	3	1	Inglaterra
162	1962	Tchecoslováquia	1	0	Hungria
163	1962	Iugoslávia	1	0	Alemanha FR
164	1962	Tchecoslováquia	3	1	Iugoslávia
165	1962	Brasil	4	2	Chile
166	1962	Chile	1	0	Iugoslávia
167	1962	Brasil	3	1	Tchecoslováquia
168	1966	Inglaterra	0	0	Uruguai
169	1966	Alemanha FR	5	0	Suíça
170	1966	Brasil	2	0	Bulgária
171	1966	União Soviética	3	0	Coreia do Norte
172	1966	França	1	1	México
173	1966	Portugal	3	1	Hungria
174	1966	Argentina	2	1	Espanha

Continua.

Número do Jogo	Ano	Time A	Gols Time A	Gols Time B	Time B
175	1966	Itália	2	0	Chile
176	1966	Uruguai	2	1	França
177	1966	Espanha	2	1	Suíça
178	1966	Hungria	3	1	Brasil
179	1966	Coreia do Norte	1	1	Chile
180	1966	Portugal	3	0	Bulgária
181	1966	Alemanha FR	0	0	Argentina
182	1966	União Soviética	1	0	Itália
183	1966	Inglaterra	2	0	México
184	1966	Uruguai	0	0	México
185	1966	Argentina	2	0	Suíça
186	1966	Portugal	3	1	Brasil
187	1966	Coreia do Norte	1	0	Itália
188	1966	Inglaterra	2	0	França
189	1966	Hungria	3	1	Bulgária
190	1966	Alemanha FR	2	1	Espanha
191	1966	União Soviética	2	1	Chile
192	1966	Inglaterra	1	0	Argentina
193	1966	Alemanha FR	4	0	Uruguai
194	1966	União Soviética	2	1	Hungria
195	1966	Portugal	5	3	Coreia do Norte
196	1966	Alemanha FR	2	1	União Soviética
197	1966	Inglaterra	2	1	Portugal
198	1966	Portugal	2	1	União Soviética
199	1966	Inglaterra	4	2	Alemanha FR
200	1970	México	0	0	União Soviética
201	1970	Uruguai	2	0	Israel
202	1970	Peru	3	2	Bulgária
203	1970	Inglaterra	1	0	Romênia
204	1970	Itália	1	0	Suécia
205	1970	Alemanha FR	2	1	Marrocos
206	1970	Brasil	4	1	Tchecoslováquia
207	1970	Bélgica	3	0	El Salvador
208	1970	Uruguai	0	0	Itália
209	1970	Peru	3	0	Marrocos
210	1970	Romênia	2	1	Tchecoslováquia
211	1970	União Soviética	4	1	Bélgica
212	1970	Suécia	1	1	Israel
213	1970	Alemanha FR	5	2	Bulgária
214	1970	Brasil	1	0	Inglaterra
215	1970	México	4	0	El Salvador
216	1970	Suécia	1	0	Uruguai
217	1970	Alemanha FR	3	1	Peru
218	1970	Brasil	3	2	Romênia
219	1970	União Soviética	2	0	El Salvador
220	1970	Itália	0	0	Israel
221	1970	Bulgária	1	1	Marrocos
222	1970	Inglaterra	1	0	Tchecoslováquia
223	1970	México	1	0	Bélgica
224	1970	Itália	4	1	México
225	1970	Alemanha FR	3	2	Inglaterra
226	1970	Brasil	4	2	Peru
227	1970	Uruguai	1	0	União Soviética
228	1970	Brasil	3	1	Uruguai
229	1970	Itália	4	3	Alemanha FR
230	1970	Alemanha FR	1	0	Uruguai
231	1970	Brasil	4	1	Itália
232	1974	Brasil	0	0	Iugoslávia
233	1974	Alemanha FR	1	0	Chile
234	1974	German DR	2	0	Austrália
235	1974	Zaire	0	2	Escócia
236	1974	Uruguai	0	2	Holanda
237	1974	Suécia	0	0	Bulgária
238	1974	Itália	3	1	Haiti
239	1974	Polônia	3	2	Argentina
240	1974	Austrália	0	3	Alemanha FR
241	1974	Escócia	0	0	Brasil
242	1974	Chile	1	1	German DR
243	1974	Iugoslávia	9	0	Zaire
244	1974	Holanda	0	0	Suécia
245	1974	Bulgária	1	1	Uruguai
246	1974	Haiti	0	7	Polônia
247	1974	Argentina	1	1	Itália
248	1974	Escócia	1	1	Iugoslávia
249	1974	Austrália	0	0	Chile
250	1974	Zaire	0	3	Brasil
251	1974	German DR	1	0	Alemanha FR
252	1974	Bulgária	1	4	Holanda
253	1974	Argentina	4	1	Haiti
254	1974	Suécia	3	0	Uruguai

Continua.

Número do Jogo	Ano	Time A	Gols Time A	Gols Time B	Time B
255	1974	Polônia	2	1	Itália
256	1974	Iugoslávia	0	2	Alemanha FR
257	1974	Brasil	1	0	German DR
258	1974	Holanda	4	0	Argentina
259	1974	Suécia	0	1	Polônia
260	1974	Argentina	1	2	Brasil
261	1974	Polônia	2	1	Iugoslávia
262	1974	German DR	0	2	Holanda
263	1974	Alemanha FR	4	2	Suécia
264	1974	Polônia	0	1	Alemanha FR
265	1974	Holanda	2	0	Brasil
266	1974	Argentina	1	1	German DR
267	1974	Suécia	2	1	Iugoslávia
268	1974	Brasil	0	1	Polônia
269	1974	Holanda	1	2	Alemanha FR
270	1978	Alemanha FR	0	0	Polônia
271	1978	Itália	2	1	França
272	1978	Tunísia	3	1	México
273	1978	Argentina	2	1	Hungria
274	1978	Suécia	1	1	Brasil
275	1978	Áustria	2	1	Espanha
276	1978	Peru	3	1	Escócia
277	1978	Holanda	3	0	IR Irã
278	1978	Itália	3	1	Hungria
279	1978	Polônia	1	0	Tunísia
280	1978	Alemanha FR	6	0	México
281	1978	Argentina	2	1	França
282	1978	Brasil	0	0	Espanha
283	1978	Áustria	1	0	Suécia
284	1978	Escócia	1	1	IR Irã
285	1978	Holanda	0	0	Peru
286	1978	França	3	1	Hungria
287	1978	Polônia	3	1	México
288	1978	Alemanha FR	0	0	Tunísia
289	1978	Itália	1	0	Argentina
290	1978	Brasil	1	0	Áustria
291	1978	Espanha	1	0	Suécia
292	1978	Peru	4	1	Irã
293	1978	Escócia	3	2	Holanda
294	1978	Alemanha FR	0	0	Itália
295	1978	Holanda	5	1	Áustria
296	1978	Brasil	3	0	Peru
297	1978	Argentina	2	0	Polônia
298	1978	Polônia	1	0	Peru
299	1978	Itália	1	0	Áustria
300	1978	Alemanha FR	2	2	Holanda
301	1978	Argentina	0	0	Brasil
302	1978	Holanda	2	1	Itália
303	1978	Áustria	3	2	Alemanha FR
304	1978	Brasil	3	1	Polônia
305	1978	Argentina	6	0	Peru
306	1978	Brasil	2	1	Itália
307	1978	Argentina	3	1	Holanda
308	1982	Argentina	0	1	Bélgica
309	1982	Itália	0	0	Polônia
310	1982	Brasil	2	1	União Soviética
311	1982	Peru	0	0	Camarões
312	1982	Hungria	10	1	El Salvador
313	1982	Escócia	5	2	Nova Zelândia
314	1982	Alemanha FR	1	2	Argélia
315	1982	Inglaterra	3	1	França
316	1982	Espanha	1	1	Honduras
317	1982	Chile	0	1	Áustria
318	1982	Tchecoslováquia	1	1	Kuwait
319	1982	Iugoslávia	0	0	Irlanda do Norte
320	1982	Itália	1	1	Peru
321	1982	Argentina	4	1	Hungria
322	1982	Brasil	4	1	Escócia
323	1982	Polônia	0	0	Camarões
324	1982	Bélgica	1	0	El Salvador
325	1982	União Soviética	3	0	Nova Zelândia
326	1982	Alemanha FR	4	1	Chile
327	1982	Inglaterra	2	0	Tchecoslováquia
328	1982	Espanha	2	1	Iugoslávia
329	1982	Argélia	0	2	Áustria
330	1982	França	4	1	Kuwait
331	1982	Honduras	1	1	Irlanda do Norte
332	1982	Polônia	5	1	Peru
333	1982	Bélgica	1	1	Hungria
334	1982	União Soviética	2	2	Escócia

Continua.

Número do Jogo	Ano	Time A	Gols Time A	Gols Time B	Time B
335	1982	Itália	1	1	Camarões
336	1982	Argentina	2	0	El Salvador
337	1982	Brasil	4	0	Nova Zelândia
338	1982	Algéria	3	2	Chile
339	1982	França	1	1	Tchecoslováquia
340	1982	Honduras	0	1	Iugoslávia
341	1982	Alemanha FR	1	0	Áustria
342	1982	Inglaterra	1	0	Kuwait
343	1982	Irlanda do Norte	1	0	Espanha
344	1982	Áustria	0	1	França
345	1982	Polônia	3	0	Bélgica
346	1982	Itália	2	1	Argentina
347	1982	Alemanha FR	0	0	Inglaterra
348	1982	Áustria	2	2	Irlanda do Norte
349	1982	Bélgica	0	1	União Soviética
350	1982	Argentina	1	3	Brasil
351	1982	Alemanha FR	2	1	Espanha
352	1982	França	4	1	Irlanda do Norte
353	1982	Polônia	0	0	União Soviética
354	1982	Itália	3	2	Brasil
355	1982	Espanha	0	0	Inglaterra
356	1982	Polônia	0	2	Itália
357	1982	Alemanha FR	3	3	França
358	1982	Polônia	3	2	França
359	1982	Itália	3	1	Alemanha FR
360	1986	Bulgária	1	1	Itália
361	1986	Canadá	0	1	França
362	1986	Espanha	0	1	Brasil
363	1986	União Soviética	6	0	Hungria
364	1986	Argentina	3	1	Coreia do Sul
365	1986	Marrocos	0	0	Polônia
366	1986	Algéria	1	1	Irlanda do Norte
367	1986	Bélgica	1	2	México
368	1986	Portugal	1	0	Inglaterra
369	1986	Paraguai	1	0	Iraque
370	1986	Escócia	0	1	Dinamarca
371	1986	Uruguai	1	1	Alemanha FR
372	1986	Itália	1	1	Argentina
373	1986	França	1	1	União Soviética
374	1986	Coreia do Sul	1	1	Bulgária
375	1986	Hungria	2	0	Canadá
376	1986	Brasil	1	0	Algéria
377	1986	Inglaterra	0	0	Marrocos
378	1986	Irlanda do Norte	1	2	Espanha
379	1986	México	1	1	Paraguai
380	1986	Polônia	1	0	Portugal
381	1986	Iraque	1	2	Bélgica
382	1986	Dinamarca	6	1	Uruguai
383	1986	Alemanha FR	2	1	Escócia
384	1986	Hungria	0	3	França
385	1986	União Soviética	2	0	Canadá
386	1986	Coreia do Sul	2	3	Itália
387	1986	Argentina	2	0	Bulgária
388	1986	Paraguai	2	2	Bélgica
389	1986	Iraque	0	1	México
390	1986	Portugal	1	3	Marrocos
391	1986	Inglaterra	3	0	Polônia
392	1986	Irlanda do Norte	0	3	Brasil
393	1986	Algéria	0	3	Espanha
394	1986	Escócia	0	0	Uruguai
395	1986	Dinamarca	2	0	Alemanha FR
396	1986	União Soviética	3	4	Bélgica
397	1986	México	2	0	Bulgária
398	1986	Argentina	1	0	Uruguai
399	1986	Brasil	4	0	Polônia
400	1986	Itália	0	2	França
401	1986	Marrocos	0	1	Alemanha FR
402	1986	Inglaterra	3	0	Paraguai
403	1986	Dinamarca	1	5	Espanha
404	1986	Brasil	1	1	França
405	1986	Alemanha FR	0	0	México
406	1986	Espanha	1	1	Bélgica
407	1986	Argentina	2	1	Inglaterra
408	1986	França	0	2	Alemanha FR
409	1986	Argentina	2	0	Bélgica
410	1986	França	4	2	Bélgica
411	1986	Argentina	3	2	Alemanha FR
412	1990	Argentina	0	1	Camarões
413	1990	União Soviética	0	2	Romênia
414	1990	Emirados Árabes Unidos	0	2	Colômbia

Continua.

Número do Jogo	Ano	Time A	Gols Time A	Gols Time B	Time B
415	1990	Itália	1	0	Áustria
416	1990	EUA	1	5	Tchecoslováquia
417	1990	Brasil	2	1	Suécia
418	1990	Alemanha FR	4	1	Iugoslávia
419	1990	Costa Rica	1	0	Escócia
420	1990	Inglaterra	1	1	República da Irlanda
421	1990	Bélgica	2	0	Coreia do Sul
422	1990	Holanda	1	1	Egito
423	1990	Uruguai	0	0	Espanha
424	1990	Argentina	2	0	União Soviética
425	1990	Camarões	2	1	Romênia
426	1990	Iugoslávia	1	0	Colômbia
427	1990	Itália	1	0	EUA
428	1990	Áustria	0	1	Tchecoslováquia
429	1990	Alemanha FR	5	1	Emirados Árabes Unidos
430	1990	Brasil	1	0	Costa Rica
431	1990	Suécia	1	2	Escócia
432	1990	Inglaterra	0	0	Holanda
433	1990	República da Irlanda	0	0	Egito
434	1990	Coreia do Sul	1	3	Espanha
435	1990	Bélgica	3	1	Uruguai
436	1990	Argentina	1	1	Romênia
437	1990	Camarões	0	4	União Soviética
438	1990	Alemanha FR	1	1	Colômbia
439	1990	Iugoslávia	4	1	Emirados Árabes Unidos
440	1990	Itália	2	0	Tchecoslováquia
441	1990	Áustria	2	1	EUA
442	1990	Brasil	1	0	Escócia
443	1990	Suécia	1	2	Costa Rica
444	1990	Coreia do Sul	0	1	Uruguai
445	1990	Bélgica	1	2	Espanha
446	1990	República da Irlanda	1	1	Holanda
447	1990	Inglaterra	1	0	Egito
448	1990	Camarões	2	1	Colômbia
449	1990	Tchecoslováquia	4	1	Costa Rica
450	1990	Brasil	0	1	Argentina
451	1990	Alemanha FR	2	1	Holanda
452	1990	República da Irlanda	0	0	Romênia
453	1990	Itália	2	0	Uruguai
454	1990	Espanha	1	2	Iugoslávia
455	1990	Inglaterra	1	0	Bélgica
456	1990	Iugoslávia	0	0	Argentina
457	1990	Itália	1	0	República da Irlanda
458	1990	Alemanha FR	1	0	Tchecoslováquia
459	1990	Inglaterra	3	2	Camarões
460	1990	Itália	1	1	Argentina
461	1990	Alemanha FR	1	1	Inglaterra
462	1990	Itália	2	1	Inglaterra
463	1990	Alemanha FR	1	0	Argentina
464	1994	Espanha	2	2	Coreia do Sul
465	1994	Alemanha	1	0	Bolívia
466	1994	EUA	1	1	Suíça
467	1994	Itália	0	1	República da Irlanda
468	1994	Colômbia	1	3	Romênia
469	1994	Bélgica	1	0	Marrocos
470	1994	Noruega	1	0	México
471	1994	Camarões	2	2	Suécia
472	1994	Holanda	2	1	Arábia Saudita
473	1994	Brasil	2	0	Rússia
474	1994	Argentina	4	0	Grécia
475	1994	Nigéria	3	0	Bulgária
476	1994	Alemanha	1	1	Espanha
477	1994	Romênia	1	4	Suíça
478	1994	EUA	2	1	Colômbia
479	1994	Itália	1	0	Noruega
480	1994	Coreia do Sul	0	0	Bolívia
481	1994	México	2	1	República da Irlanda
482	1994	Suécia	3	1	Rússia
483	1994	Brasil	3	0	Camarões
484	1994	Bélgica	1	0	Holanda
485	1994	Arábia Saudita	2	1	Marrocos
486	1994	Argentina	2	1	Nigéria
487	1994	Bulgária	4	0	Grécia
488	1994	EUA	0	1	Romênia
489	1994	Suíça	0	2	Colômbia
490	1994	Alemanha	3	2	Coreia do Sul
491	1994	Bolívia	1	3	Espanha
492	1994	República da Irlanda	0	0	Noruega
493	1994	Brasil	1	1	Suécia
494	1994	Itália	1	1	México

Continua.

Número do Jogo	Ano	Time A	Gols Time A	Gols Time B	Time B
495	1994	Rússia	6	1	Camarões
496	1994	Marrocos	1	2	Holanda
497	1994	Bélgica	0	1	Arábia Saudita
498	1994	Grécia	0	2	Nigéria
499	1994	Argentina	0	2	Bulgária
500	1994	Alemanha	3	2	Bélgica
501	1994	Espanha	3	0	Suíça
502	1994	Arábia Saudita	1	3	Suécia
503	1994	Romênia	3	2	Argentina
504	1994	Holanda	2	0	República da Irlanda
505	1994	Brasil	1	0	EUA
506	1994	Nigéria	1	2	Itália
507	1994	México	1	1	Bulgária
508	1994	Itália	2	1	Espanha
509	1994	Holanda	2	3	Brasil
510	1994	Bulgária	2	1	Alemanha
511	1994	Romênia	2	2	Suécia
512	1994	Bulgária	1	2	Itália
513	1994	Suécia	0	1	Brasil
514	1994	Suécia	4	0	Bulgária
515	1994	Brasil	0	0	Itália
516	1998	Brasil	2	1	Escócia
517	1998	Marrocos	2	2	Noruega
518	1998	Itália	2	2	Chile
519	1998	Camarões	1	1	Áustria
520	1998	Paraguai	0	0	Bulgária
521	1998	Arábia Saudita	0	1	Dinamarca
522	1998	França	3	0	África do Sul
523	1998	Espanha	2	3	Nigéria
524	1998	Coreia do Sul	1	3	México
525	1998	Holanda	0	0	Bélgica
526	1998	Argentina	1	0	Japão
527	1998	Iugoslávia	1	0	Irã
528	1998	Jamaica	1	3	Croácia
529	1998	Inglaterra	2	0	Tunísia
530	1998	Romênia	1	0	Colômbia
531	1998	Alemanha	2	0	EUA
532	1998	Escócia	1	1	Noruega
533	1998	Brasil	3	0	Marrocos
534	1998	Chile	1	1	Áustria
535	1998	Itália	3	0	Camarões
536	1998	África do Sul	1	1	Dinamarca
537	1998	França	4	0	Arábia Saudita
538	1998	Nigéria	1	0	Bulgária
539	1998	Espanha	0	0	Paraguai
540	1998	Japão	0	1	Croácia
541	1998	Bélgica	2	2	México
542	1998	Holanda	5	0	Coreia do Sul
543	1998	Alemanha	2	2	Iugoslávia
544	1998	Argentina	5	0	Jamaica
545	1998	EUA	1	2	Irã
546	1998	Colômbia	1	0	Tunísia
547	1998	Romênia	2	1	Inglaterra
548	1998	Itália	2	1	Áustria
549	1998	Chile	1	1	Camarões
550	1998	Escócia	0	3	Marrocos
551	1998	Brasil	1	2	Noruega
552	1998	França	2	1	Dinamarca
553	1998	África do Sul	2	2	Arábia Saudita
554	1998	Espanha	6	1	Bulgária
555	1998	Nigéria	1	3	Paraguai
556	1998	Bélgica	1	1	Coreia do Sul
557	1998	Holanda	2	2	México
558	1998	Alemanha	2	0	Irã
559	1998	EUA	0	1	Iugoslávia
560	1998	Japão	1	2	Jamaica
561	1998	Argentina	1	0	Croácia
562	1998	Romênia	1	1	Tunísia
563	1998	Colômbia	0	2	Inglaterra
564	1998	Itália	1	0	Noruega
565	1998	Brasil	4	1	Chile
566	1998	França	1	0	Paraguai
567	1998	Nigéria	1	4	Dinamarca
568	1998	Alemanha	2	1	México
569	1998	Holanda	2	1	Iugoslávia
570	1998	Romênia	0	1	Croácia
571	1998	Argentina	2	2	Inglaterra
572	1998	Itália	0	0	França
573	1998	Brasil	3	2	Dinamarca
574	1998	Holanda	2	1	Argentina

Continua.

Número do Jogo	Ano	Time A	Gols Time A	Gols Time B	Time B
575	1998	Alemanha	0	3	Croácia
576	1998	Brasil	1	1	Holanda
577	1998	França	2	1	Croácia
578	1998	Holanda	1	2	Croácia
579	1998	Brasil	0	3	França
580	2002	França	0	1	Senegal
581	2002	Uruguai	1	2	Dinamarca
582	2002	República da Irlanda	1	1	Camarões
583	2002	Alemanha	8	0	Arábia Saudita
584	2002	Argentina	1	0	Nigéria
585	2002	Paraguai	2	2	África do Sul
586	2002	Inglaterra	1	1	Suécia
587	2002	Espanha	3	1	Eslovênia
588	2002	Brasil	2	1	Turquia
589	2002	Itália	2	0	Equador
590	2002	Croácia	0	1	México
591	2002	China PR	0	2	Costa Rica
592	2002	Japão	2	2	Bélgica
593	2002	Coreia do Sul	2	0	Polônia
594	2002	Rússia	2	0	Tunísia
595	2002	EUA	3	2	Portugal
596	2002	Alemanha	1	1	República da Irlanda
597	2002	Dinamarca	1	1	Senegal
598	2002	Camarões	1	0	Arábia Saudita
599	2002	França	0	0	Uruguai
600	2002	Suécia	2	1	Nigéria
601	2002	Argentina	0	1	Inglaterra
602	2002	Espanha	3	1	Paraguai
603	2002	África do Sul	1	0	Eslovênia
604	2002	Brasil	4	0	China PR
605	2002	Itália	1	2	Croácia
606	2002	Costa Rica	1	1	Turquia
607	2002	Japão	1	0	Rússia
608	2002	México	2	1	Equador
609	2002	Coreia do Sul	1	1	EUA
610	2002	Portugal	4	0	Polônia
611	2002	Tunísia	1	1	Bélgica
612	2002	Dinamarca	2	0	França
613	2002	Senegal	3	3	Uruguai
614	2002	Arábia Saudita	0	3	República da Irlanda
615	2002	Camarões	0	2	Alemanha
616	2002	Suécia	1	1	Argentina
617	2002	Nigéria	0	0	Inglaterra
618	2002	Eslovênia	1	3	Paraguai
619	2002	África do Sul	2	3	Espanha
620	2002	Costa Rica	2	5	Brasil
621	2002	Turquia	3	0	China PR
622	2002	Equador	1	0	Croácia
623	2002	México	1	1	Itália
624	2002	Tunísia	0	2	Japão
625	2002	Portugal	0	1	Coreia do Sul
626	2002	Polônia	3	1	EUA
627	2002	Bélgica	3	2	Rússia
628	2002	Dinamarca	0	3	Inglaterra
629	2002	Alemanha	1	0	Paraguai
630	2002	Espanha	1	1	República da Irlanda
631	2002	Suécia	1	2	Senegal
632	2002	Brasil	2	0	Bélgica
633	2002	México	0	2	EUA
634	2002	Japão	0	1	Turquia
635	2002	Coreia do Sul	2	1	Itália
636	2002	Inglaterra	1	2	Brasil
637	2002	Alemanha	1	0	EUA
638	2002	Senegal	0	1	Turquia
639	2002	Espanha	0	0	Coreia do Sul
640	2002	Alemanha	1	0	Coreia do Sul
641	2002	Brasil	1	0	Turquia
642	2002	Coreia do Sul	2	3	Turquia
643	2002	Alemanha	0	2	Brasil
644	2006	Alemanha	4	2	Costa Rica
645	2006	Polônia	0	2	Equador
646	2006	Inglaterra	1	0	Paraguai
647	2006	Trinidade e Tobago	0	0	Suécia
648	2006	Argentina	2	1	Costa do Marfim
649	2006	Sérvia e Montenegro	0	1	Holanda
650	2006	México	3	1	Irã
651	2006	Angola	0	1	Portugal
652	2006	Austrália	3	1	Japão
653	2006	EUA	0	3	Czech Republic
654	2006	Itália	2	0	Gana

Continua.

Número do Jogo	Ano	Time A	Gols Time A	Gols Time B	Time B
655	2006	Coreia do Sul	2	1	Togo
656	2006	França	0	0	Suíça
657	2006	Brasil	1	0	Croácia
658	2006	Espanha	4	0	Ucrânia
659	2006	Tunísia	2	2	Arábia Saudita
660	2006	Alemanha	1	0	Polónia
661	2006	Equador	3	0	Costa Rica
662	2006	Inglaterra	2	0	Trinidade e Tobago
663	2006	Suécia	1	0	Paraguai
664	2006	Argentina	6	0	Sérvia e Montenegro
665	2006	Holanda	2	1	Costa do Marfim
666	2006	México	0	0	Angola
667	2006	Portugal	2	0	IR Irã
668	2006	Czech Republic	0	2	Gana
669	2006	Itália	1	1	EUA
670	2006	Japão	0	0	Croácia
671	2006	Brasil	2	0	Austrália
672	2006	França	1	1	Coreia do Sul
673	2006	Togo	0	2	Suíça
674	2006	Arábia Saudita	0	4	Ucrânia
675	2006	Espanha	3	1	Tunísia
676	2006	Equador	0	3	Alemanha
677	2006	Costa Rica	1	2	Polónia
678	2006	Suécia	2	2	Inglaterra
679	2006	Paraguai	2	0	Trinidade e Tobago
680	2006	Irã	1	1	Angola
681	2006	Portugal	2	1	México
682	2006	Holanda	0	0	Argentina
683	2006	Costa do Marfim	3	2	Sérvia e Montenegro
684	2006	Czech Republic	0	2	Itália
685	2006	Gana	2	1	EUA
686	2006	Japão	1	4	Brasil
687	2006	Croácia	2	2	Austrália
688	2006	Ucrânia	1	0	Tunísia
689	2006	Arábia Saudita	0	1	Espanha
690	2006	Togo	0	2	França
691	2006	Suíça	2	0	Coreia do Sul
692	2006	Alemanha	2	0	Suécia
693	2006	Argentina	2	1	México
694	2006	Inglaterra	1	0	Equador
695	2006	Portugal	1	0	Holanda
696	2006	Itália	1	0	Austrália
697	2006	Suíça	0	0	Ucrânia
698	2006	Brasil	3	0	Gana
699	2006	Espanha	1	3	França
700	2006	Alemanha	1	1	Argentina
701	2006	Itália	3	0	Ucrânia
702	2006	Inglaterra	0	0	Portugal
703	2006	Brasil	0	1	França
704	2006	Alemanha	0	2	Itália
705	2006	Portugal	0	1	França
706	2006	Alemanha	3	1	Portugal
707	2006	Itália	1	1	França
708	2010	África do Sul	1	1	México
709	2010	Uruguai	0	0	França
710	2010	Coreia do Sul	2	0	Grécia
711	2010	Argentina	1	0	Nigéria
712	2010	Inglaterra	1	1	EUA
713	2010	Algéria	0	1	Eslovênia
714	2010	Sérvia	0	1	Gana
715	2010	Alemanha	4	0	Austrália
716	2010	Holanda	2	0	Dinamarca
717	2010	Japão	1	0	Camarões
718	2010	Itália	1	1	Paraguai
719	2010	Nova Zelândia	1	1	Slováquia
720	2010	Costa do Marfim	0	0	Portugal
721	2010	Brasil	2	1	Coreia do Norte
722	2010	Honduras	0	1	Chile
723	2010	Espanha	0	1	Suíça
724	2010	África do Sul	0	3	Uruguai
725	2010	Argentina	4	1	Coreia do Sul
726	2010	Grécia	2	1	Nigéria
727	2010	França	0	2	México
728	2010	Alemanha	0	1	Sérvia
729	2010	Eslovênia	2	2	EUA
730	2010	Inglaterra	0	0	Algéria
731	2010	Holanda	1	0	Japão
732	2010	Gana	1	1	Austrália
733	2010	Camarões	1	2	Dinamarca
734	2010	Slováquia	0	2	Paraguai

Continua.

Número do Jogo	Ano	Time A	Gols Time A	Gols Time B	Time B
735	2010	Itália	1	1	Nova Zelândia
736	2010	Brasil	3	1	Costa do Marfim
737	2010	Portugal	7	0	Coreia do Norte
738	2010	Chile	1	0	Suíça
739	2010	Espanha	2	0	Honduras
740	2010	México	0	1	Uruguai
741	2010	França	1	2	África do Sul
742	2010	Nigéria	2	2	Coreia do Sul
743	2010	Grécia	0	2	Argentina
744	2010	Eslovênia	0	1	Inglaterra
745	2010	EUA	1	0	Algéria
746	2010	Gana	0	1	Alemanha
747	2010	Austrália	2	1	Sérvia
748	2010	Slovakia	3	2	Itália
749	2010	Paraguai	0	0	Nova Zelândia
750	2010	Dinamarca	1	3	Japão
751	2010	Camarões	1	2	Holanda
752	2010	Portugal	0	0	Brasil
753	2010	Coreia do Norte	0	3	Costa do Marfim
754	2010	Chile	1	2	Espanha
755	2010	Suíça	0	0	Honduras
756	2010	Uruguai	2	1	Coreia do Sul
757	2010	EUA	1	2	Gana
758	2010	Alemanha	4	1	Inglaterra
759	2010	Argentina	3	1	México
760	2010	Holanda	2	1	Slovakia
761	2010	Brasil	3	0	Chile
762	2010	Paraguai	0	0	Japão
763	2010	Espanha	1	0	Portugal
764	2010	Holanda	2	1	Brasil
765	2010	Uruguai	1	1	Gana
766	2010	Argentina	0	4	Alemanha
767	2010	Paraguai	0	1	Espanha
768	2010	Uruguai	2	3	Holanda
769	2010	Alemanha	0	1	Espanha
770	2010	Uruguai	2	3	Alemanha
771	2010	Holanda	0	1	Espanha
772	2014	Brasil	3	1	Croácia
773	2014	México	1	0	Camarões
774	2014	Espanha	1	5	Holanda
775	2014	Chile	3	1	Austrália
776	2014	Colômbia	3	0	Grécia
777	2014	Uruguai	1	3	Costa Rica
778	2014	Inglaterra	1	2	Itália
779	2014	Costa do Marfim	2	1	Japão
780	2014	Suíça	2	1	Equador
781	2014	França	3	0	Honduras
782	2014	Argentina	2	1	Bósnia e Herzegovina
783	2014	Alemanha	4	0	Portugal
784	2014	IR Irã	0	0	Nigéria
785	2014	Gana	1	2	EUA
786	2014	Bélgica	2	1	Algéria
787	2014	Brasil	0	0	México
788	2014	Rússia	1	1	Coreia do Sul
789	2014	Austrália	2	3	Holanda
790	2014	Espanha	0	2	Chile
791	2014	Camarões	0	4	Croácia
792	2014	Colômbia	2	1	Costa do Marfim
793	2014	Uruguai	2	1	Inglaterra
794	2014	Japão	0	0	Grécia
795	2014	Itália	0	1	Costa Rica
796	2014	Suíça	2	5	França
797	2014	Honduras	1	2	Equador
798	2014	Argentina	1	0	IR Irã
799	2014	Alemanha	2	2	Gana
800	2014	Nigéria	1	0	Bósnia e Herzegovina
801	2014	Bélgica	1	0	Rússia
802	2014	Coreia do Sul	2	4	Algéria
803	2014	EUA	2	2	Portugal
804	2014	Austrália	0	3	Espanha
805	2014	Holanda	2	0	Chile
806	2014	Camarões	1	4	Brasil
807	2014	Croácia	1	3	México
808	2014	Itália	0	1	Uruguai
809	2014	Costa Rica	0	0	Inglaterra
810	2014	Japão	1	4	Colômbia
811	2014	Grécia	2	1	Costa do Marfim
812	2014	Nigéria	2	3	Argentina
813	2014	Bósnia e Herzegovina	3	1	IR Irã
814	2014	Honduras	0	3	Suíça

Continua.

Número do Jogo	Ano	Time A	Gols Time A	Gols Time B	Time B
815	2014	Equador	0	0	França
816	2014	EUA	0	1	Alemanha
817	2014	Portugal	2	1	Gana
818	2014	Coreia do Sul	0	1	Bélgica
819	2014	Algéria	1	1	Rússia
820	2014	Brasil	1	1	Chile
821	2014	Colômbia	2	0	Uruguai
822	2014	França	2	0	Nigéria
823	2014	Alemanha	2	1	Algéria
824	2014	Brasil	2	1	Colômbia
825	2014	França	0	1	Alemanha
826	2014	Brasil	1	7	Alemanha
827	2014	Brasil	0	3	Holanda
828	2014	Alemanha	1	0	Argentina
829	2014	Holanda	0	0	Argentina
830	2014	Holanda	0	0	Costa Rica
831	2014	Argentina	1	0	Bélgica
832	2014	Holanda	2	1	México
833	2014	Costa Rica	1	1	Grécia
834	2014	Argentina	1	0	Suíça
835	2014	Bélgica	2	1	EUA
836	2014	Brasil	1	1	Chile
837	2014	Colômbia	2	0	Uruguai
838	2014	Holanda	2	1	México
839	2014	Costa Rica	1	1	Grécia
840	2014	França	2	0	Nigéria
841	2014	Alemanha	2	1	Algéria
842	2014	Argentina	1	0	Suíça
843	2014	Bélgica	2	1	EUA
844	2014	França	0	1	Alemanha
845	2014	Brasil	2	1	Colômbia
846	2014	Argentina	1	0	Bélgica
847	2014	Holanda	0	0	Costa Rica
848	2014	Brasil	1	7	Alemanha
849	2014	Holanda	0	0	Argentina
850	2014	Brasil	0	3	Holanda
851	2014	Alemanha	1	0	Argentina
852	2018	Rússia	5	0	Arábia Saudita
853	2018	Egito	0	1	Uruguai
854	2018	Marrocos	0	1	Irã
855	2018	Portugal	3	3	Espanha
856	2018	França	2	1	Austrália
857	2018	Argentina	1	1	Islândia
858	2018	Peru	0	1	Dinamarca
859	2018	Croácia	2	0	Nigéria
860	2018	Costa Rica	0	1	Sérvia
861	2018	Alemanha	0	1	México
862	2018	Brasil	1	1	Suíça
863	2018	Suécia	1	0	Coreia do Sul
864	2018	Bélgica	3	0	Panamá
865	2018	Tunísia	1	2	Inglaterra
866	2018	Colômbia	1	2	Japão
867	2018	Polônia	1	2	Senegal
868	2018	Rússia	3	1	Egito
869	2018	Portugal	1	0	Marrocos
870	2018	Uruguai	1	0	Arábia Saudita
871	2018	Irã	0	1	Espanha
872	2018	Dinamarca	1	1	Austrália
873	2018	França	1	0	Peru
874	2018	Argentina	0	3	Croácia
875	2018	Brasil	2	0	Costa Rica
876	2018	Nigéria	2	0	Islândia
877	2018	Sérvia	1	2	Suíça
878	2018	Bélgica	5	2	Tunísia
879	2018	Coreia do Sul	1	2	México
880	2018	Alemanha	2	1	Suécia
881	2018	Inglaterra	6	1	Panamá
882	2018	Japão	2	2	Senegal
883	2018	Polônia	0	3	Colômbia
884	2018	Arábia Saudita	2	1	Egito
885	2018	Uruguai	3	0	Rússia
886	2018	Irã	1	1	Portugal
887	2018	Espanha	2	2	Marrocos
888	2018	Austrália	0	2	Peru
889	2018	Dinamarca	0	0	França
890	2018	Nigéria	1	2	Argentina
891	2018	Islândia	1	2	Croácia
892	2018	México	0	3	Suécia
893	2018	Coreia do Sul	2	0	Alemanha
894	2018	Suíça	2	2	Costa Rica

Continua.

Número do Jogo	Ano	Time A	Gols Time A	Gols Time B	Time B
895	2018	Sérvia	0	2	Brasil
896	2018	Senegal	0	1	Colômbia
897	2018	Japão	0	1	Polônia
898	2018	Inglaterra	0	1	Bélgica
899	2018	Panamá	1	2	Tunísia
900	2018	França	4	3	Argentina
901	2018	Uruguai	2	1	Portugal
902	2018	Espanha	1	1	Rússia
903	2018	Croácia	1	1	Dinamarca
904	2018	Brasil	2	0	México
905	2018	Bélgica	3	2	Japão
906	2018	Suécia	1	0	Suíça
907	2018	Colômbia	1	1	Inglaterra
908	2018	Uruguai	0	2	França
909	2018	Brasil	1	2	Bélgica
910	2018	Suécia	0	2	Inglaterra
911	2018	Rússia	2	2	Croácia
912	2018	França	1	0	Bélgica
913	2018	Croácia	2	1	Inglaterra

Tabela 1: Resultados das partidas.

Na Tabela 2 são apresentados os dados de posição de cada seleção nas Copas do Mundo desde 1950 até 2014. O cálculo da pontuação atribuída a cada posição é feito pela função “*to_score*” definida no arquivo auxiliar “*utils*”.

Ano	Posição	País	Ano	Posição	País	Ano	Posição	País	Ano	Posição	País
1950	1	Uruguai	1966	2	Alemanha	1982	3	Polônia	1998	4	Holanda
1950	2	Brasil	1966	3	Portugal	1982	4	França	2002	1	Brasil
1950	3	Suécia	1966	4	União Soviética	1986	1	Argentina	2002	2	Alemanha
1950	4	Espanha	1970	1	Brasil	1986	2	Alemanha	2002	3	Turquia
1954	1	Alemanha	1970	2	Itália	1986	3	França	2002	4	Coreia do Sul
1954	2	Hungria	1970	3	Alemanha	1986	4	Bélgica	2006	1	Itália
1954	3	Áustria	1970	4	Uruguai	1990	1	Alemanha	2006	2	França
1954	4	Uruguai	1974	1	Alemanha	1990	2	Argentina	2006	3	Alemanha
1958	1	Brasil	1974	2	Holanda	1990	3	Itália	2006	4	Uruguai
1958	2	Suécia	1974	3	Polônia	1990	4	Inglaterra	2010	1	Espanha
1958	3	França	1974	4	Brasil	1994	1	Brasil	2010	2	Holanda
1958	4	Alemanha	1978	1	Argentina	1994	2	Itália	2010	3	Alemanha
1962	1	Brasil	1978	2	Holanda	1994	3	Suécia	2010	4	Uruguai
1962	2	Tchecoslováquia	1978	3	Brasil	1994	4	Bulgária	2014	1	Alemanha
1962	3	Chile	1978	4	Itália	1998	1	França	2014	2	Argentina
1962	4	Iugoslávia	1982	1	Itália	1998	2	Brasil	2014	3	Holanda
1966	1	Inglaterra	1982	2	Alemanha	1998	3	Croácia	2014	4	Brasil

Tabela 2: Pódios ao longo dos anos.

A seguir está representado o código em Python anteriormente retratado pelo Algoritmo 1 do pseudocódigo, em sua totalidade. Além do que foi apresentado pelo pseudocódigo, nesta versão tem-se a utilização de uma função para a contagem de tempo entre o início e o fim do processamento.

```

1 #Código para previsão de resultados de jogos da Copa do Mundo
2 #Autor: João Pedro Fisanotti
3 #Adaptação: Laura Peconick
4
5 #Carregando os pacotes necessários:
6 from random import random
7
8 from IPython.display import SVG
9 import pygal
10
11 from pybrain.structure import SigmoidLayer
12 from pybrain.tools.shortcuts import buildNetwork
13 from pybrain.supervised.trainers import BackpropTrainer
14 from pybrain.datasets import ClassificationDataSet
15 from pybrain.utilities import percentError
16
17 import time
18 import timeit

```

```

19
20 from utils import get_matches, get_team_stats, extract_samples, normalize,
21 split_samples
22
23 #Inicio da contagem do tempo:
24 inicio = timeit.default_timer()
25
26 #Dados de input
27 input_features = ['year',
28                  'matches_won_percent',
29                  'podium_score_yearly',
30                  'matches_won_percent_2',
31                  'podium_score_yearly_2',]
32
33 #Definindo o que é desejado como output:
34 output_feature = 'winner'
35
36 #Excluindo os empates (melhora a acurácia do código) (Não funciona caso seja
37 #substituído por 'False', já que a função SigmoidLayer só permite duas opções
38 #de saída):
39 exclude_ties = True
40
41 #Duplicando os resultados das partidas, para que não tenha time 'de casa',
42 #favorecido pela torcida:
43 duplicate_with_reversed = True
44
45 #Calculando algumas estatísticas gerais dos times:
46 team_stats = get_team_stats()
47 team_stats
48
49 #Para visualizar, especificadamente, a estatística de porcentagem de vitórias
50 #de cada time:
51 team_stats.sort_values('matches_won_percent')
52
53 #Pegar os dados das partidas, duplicar os resultados e excluindo os empates:
54 matches = get_matches(with_team_stats=True,
55                       duplicate_with_reversed=duplicate_with_reversed,
56                       exclude_ties=exclude_ties)
57
58 matches
59
60 #Iniciando o processo de aprendizagem, definindo os inputs:
61 inputs, outputs = extract_samples(matches,
62                                  input_features,
63                                  output_feature)
64
65 #Normalizando os dados de input:
66 normalizer, inputs = normalize(inputs)
67
68 #Definindo a porcentagem dos dados de input que serão utilizados para treinar
69 #a RNA (Nesse caso, 10%):
70 train_inputs, train_outputs, test_inputs, test_outputs = split_samples(inputs,
71                                                                           outputs)
72
73 n = buildNetwork(len(input_features),
74                 10 * len(input_features),
75                 10 * len(input_features),
76                 1,
77                 outclass=SigmoidLayer,
78                 bias=True)
79
80 #Chama a rede neural e traduz o resultado no output final do vencedor da

```

```

81 #partida:
82 def neural_result(input):
83     n_output = n.activate(input)
84     if n_output >= 0.5:
85         return 2
86     else:
87         return 1
88
89 #Calcula, treina e testa os possíveis erros:
90 def test_network():
91     print (100 - percentError(map(neural_result, train_inputs), train_outputs),
92           100 - percentError(map(neural_result, test_inputs), test_outputs))
93
94 #Cria um grupo de teste para treinar a rede neural e mostra a acurácia do
95 #teste:
96 train_set = ClassificationDataSet(len(input_features))
97
98 for i, input_line in enumerate(train_inputs):
99     train_set.addSample(train_inputs[i], [train_outputs[i] - 1])
100
101 trainer = BackpropTrainer(n, dataset=train_set, momentum=0.5, weightdecay=0.0)
102
103 train_set.assignClasses()
104
105 test_network()
106
107 #Treina a rede dado um dado número de interações (Nesse caso, 20 - pode ser
108 #alterado):
109 for i in range(20):
110     trainer.train()
111     test_network()
112
113 #Define a função 'predict' que será usada para realizar as previsões, no
114 #formato predict(ANO, 'TimeA', 'TimeB') e que dará como output o vencedor da
115 #partida:
116 def predict(year, team1, team2, reverse=True):
117     inputs = []
118
119     if reverse:
120         assert predict(year, team2, team1, False)
121
122     for feature in input_features:
123         from_team_2 = '_2' in feature
124         feature = feature.replace('_2', '')
125
126         if feature in team_stats.columns.values:
127             team = team2 if from_team_2 else team1
128             value = team_stats.loc[team, feature]
129         elif feature == 'year':
130             value = year
131         else:
132             raise ValueError("Don't know where to get feature: " + feature)
133
134         inputs.append(value)
135     print (inputs, normalizer)
136     inputs = normalizer.transform([inputs])[0]
137     result = neural_result(inputs)
138
139     if result == 0:
140         return 'tie'
141     elif result == 1:
142         return team1

```

```

143     elif result == 2:
144         return team2
145     else:
146         return 'Unknown result: ' + str(result)
147
148 #Finalizando a contagem do tempo e imprimindo a duração total:
149 fim = timeit.default_timer()
150 print ('Duracao: %f' % (fim - inicio))

```

Código 1: Código para previsões.

O arquivo auxiliar “utils”, no qual os dados são inseridos e as estatísticas utilizadas pela RNAs são geradas e os dados são normalizados.

```

1 #Código 'utils' para manipulação dos dados.
2 #Autor: João Pedro Fisanotti
3 #Adaptação: Laura Peconick
4
5 #Carregando os pacotes necessários:
6 from random import random
7 import pandas as pd
8 import pygal
9 import sklearn
10 from sklearn.preprocessing import StandardScaler
11
12 #Carregando os arquivos de resultados, pódio e atualização de nomes,
13 #respectivamente.
14 RAW_MATCHES_FILE = 'raw_matches_new_semis.csv'
15 RAW_WINNERS_FILE = 'raw_winners.csv'
16 TEAM_RENAMES_FILE = 'team_renames.csv'
17
18 #Definindo a função que faz a atualização dos nomes:
19 def apply_renames(column):
20     with open(TEAM_RENAMES_FILE) as renames_file:
21         renames = dict(l.strip().split(',')
22                        for l in renames_file.readlines()
23                        if l.strip())
24
25     def renamer(team):
26         return renames.get(team, team)
27
28     return column.map(renamer)
29
30 #Definindo a função que cria as tabelas com resultados dos jogos:
31 def get_matches(with_team_stats=False, duplicate_with_reversed=False,
32                exclude_ties=False):
33     matches = pd.DataFrame.from_csv(RAW_MATCHES_FILE)
34     for column in ('team1', 'team2'):
35         matches[column] = apply_renames(matches[column])
36
37     if duplicate_with_reversed:
38         id_offset = len(matches)
39
40         matches2 = matches.copy()
41         matches2.rename(columns={'team1': 'team2',
42                                'team2': 'team1',
43                                'score1': 'score2',
44                                'score2': 'score1'},
45                        inplace=True)
46         matches2.index = matches2.index.map(lambda x: x + id_offset)
47
48     matches = pd.concat((matches, matches2))

```

```

49
50 def winner_from_score_diff(x):
51     if x > 0:
52         return 1
53     elif x < 0:
54         return 2
55     else:
56         return 0
57
58 matches['score_diff'] = matches['score1'] - matches['score2']
59 matches['winner'] = matches['score_diff']
60 matches['winner'] = matches['winner'].map(winner_from_score_diff)
61
62 if exclude_ties:
63     matches = matches[matches['winner'] != 0]
64
65 if with_team_stats:
66     stats = get_team_stats()
67
68     matches = matches.join(stats, on='team1')\
69         .join(stats, on='team2', rsuffix='_2')
70
71     return matches
72
73 #Definindo as tabelas com os dados de pódio:
74 def get_winners():
75     winners = pd.DataFrame.from_csv(RAW_WINNERS_FILE)
76     winners.team = apply_renames(winners.team)
77
78     return winners
79
80 #Calculando estatísticas diversas para cada time, como média de jogos:
81 def get_team_stats():
82     winners = get_winners()
83     matches = get_matches()
84
85     teams = set(matches.team1.unique()).union(matches.team2.unique())
86
87     stats = pd.DataFrame(list(teams), columns=['team'])
88
89     stats = stats.set_index('team')
90
91     for team in teams:
92         team_matches = matches[(matches.team1 == team) |
93                                 (matches.team2 == team)]
94         stats.loc[team, 'matches_played'] = len(team_matches)
95
96         # wins where the team was on the left side (team1)
97         wins1 = team_matches[(team_matches.team1 == team) &
98                               (team_matches.score1 > team_matches.score2)]
99         # wins where the team was on the right side (team2)
100        wins2 = team_matches[(team_matches.team2 == team) &
101                               (team_matches.score2 > team_matches.score1)]
102
103        stats.loc[team, 'matches_won'] = len(wins1) + len(wins2)
104
105        stats.loc[team, 'years_played'] = len(team_matches.year.unique())
106
107        team_podiums = winners[winners.team == team]
108        to_score = lambda position: 2 ** (5 - position) # better position -> more score,
109        exponential
110        stats.loc[team, 'podium_score'] = team_podiums.position.map(to_score).sum()

```



```

110
111     stats.loc[team, 'cups_won'] = len(team_podiums[team_podiums.position == 1])
112
113     stats['matches_won_percent'] = stats['matches_won'] / stats['matches_played'] * 100.0
114     stats['podium_score_yearly'] = stats['podium_score'] / stats['years_played']
115     stats['cups_won_yearly'] = stats['cups_won'] / stats['years_played']
116
117     return stats
118
119 #Definindo o conjunto de dados para teste:
120 def extract_samples(matches, origin_features, result_feature):
121     inputs = [tuple(matches.loc[i, feature]
122                   for feature in origin_features)
123              for i in matches.index]
124
125     outputs = tuple(matches[result_feature].values)
126
127     assert len(inputs) == len(outputs)
128
129     return inputs, outputs
130
131 #Definindo a função de normalização:
132 def normalize(array):
133     scaler = StandardScaler()
134     array = scaler.fit_transform(array)
135
136     return scaler, array
137
138 #Definindo as amostras de dados para treinamento:
139 def split_samples(inputs, outputs, percent=0.75):
140     assert len(inputs) == len(outputs)
141
142     inputs1 = []
143     inputs2 = []
144     outputs1 = []
145     outputs2 = []
146
147     for i, inputs_row in enumerate(inputs):
148         if random() < percent:
149             input_to = inputs1
150             output_to = outputs1
151         else:
152             input_to = inputs2
153             output_to = outputs2
154
155         input_to.append(inputs_row)
156         output_to.append(outputs[i])
157
158     return inputs1, outputs1, inputs2, outputs2

```

Código 2: Arquivo auxiliar para manipulação de dados.

REFERÊNCIAS BIBLIOGRÁFICAS

- Alpaydin, E. (2004). *Introduction to machine learning*. Cambridge, MA: MIT Press. (Citado 1 vez na página 16.)
- Andersson, P., Edman, J., & Ekman, M. (2005). [Predicting the World Cup 2002 in soccer: Performance and confidence of experts and non-experts](#). *International Journal of Forecasting*, 21(3), 565–576. (Citado 2 vezes nas páginas 2 e 12.)
- Ariely, D., & Zakay, D. (2001). [A timely account of the role of duration in decision making](#). *Acta Psychologica*, 108(2), 187–207. (Citado 1 vez na página 22.)
- Baker, R. D., & McHale, I. G. (2013). [Forecasting exact scores in National Football League games](#). *International Journal of Forecasting*, 29(1), 122–130. (Citado 2 vezes nas páginas 2 e 3.)
- Basheer, I., & Hajmeer, M. (2000). [Artificial neural networks: Fundamentals, computing, design, and application](#). *Journal of Microbiological Methods*, 43(1), 3–31. (Citado 1 vez na página 17.)
- Boulier, B. L., & Stekler, H. (1999). [Are sports seedings good predictors? An evaluation](#). *International Journal of Forecasting*, 15(1), 83–91. (Citado 1 vez na página 2.)
- Boyle, N. (2016). [Teaching history and political economy through soccer](#). *Soccer & Society*, 18(2-3), 407–417. (Citado 1 vez na página 7.)
- Breiman, L. (2001). [Random Forests](#). *Machine Learning*, 45(1), 5–32. (Citado 1 vez na página 13.)
- Brunette, E. S., Flemmer, R. C., & Flemmer, C. L. (2009). [A review of artificial intelligence](#). In *2009 4th international conference on autonomous robots and agents*. IEEE. (Citado 1 vez na página 15.)
- Buchanan, B. G. (2006). A (very) brief history of artificial intelligence. *AI Magazine*, 26(4), 53–60. (Citado 2 vezes nas páginas 14 e 15.)
- Bunk, B. D. (2015). [Sardinero and Not a Can of Sardines](#). *Journal of Urban History*, 41(3), 444–459. (Citado 2 vezes nas páginas 8 e 10.)
- Carling, C., Bloomfield, J., Nelsen, L., & Reilly, T. (2008). [The Role of Motion Analysis in Elite Soccer](#). *Sports Medicine*, 38(10), 839–862. (Citado 1 vez na página 6.)
- Codd, E. F. (1970). [A relational model of data for large shared data banks](#). *Communications of the ACM*, 13(6), 377–387. (Citado 1 vez na página 24.)
- Condon, E. M., Golden, B. L., & Wasil, E. A. (1999). [Predicting the success of nations at the Summer Olympics using neural networks](#). *Computers & Operations Research*, 26(13), 1243–1265. (Citado 1 vez na página 2.)
- CONMEBOL. (2016). *Estatutos de la CONMEBOL* (Tech. Rep.). Lima, Peru: Confederación Sudamericana de Fútbol. (Citado 1 vez na página 10.)
- Constantinou, A., & Fenton, N. (2017). [Towards smart-data: Improving predictive accuracy in long-term football team performance](#). *Knowledge-Based Systems*, 124(1), 93–104. (Citado 2 vezes nas páginas 2 e 12.)

- Constantinou, A., Fenton, N. E., & Neil, M. (2012). [pi-football: A Bayesian network model for forecasting Association Football match outcomes](#). *Knowledge-Based Systems*, 36(1), 322–339.
(Citado 1 vez na página 12.)
- Constantinou, A., Fenton, N. E., & Neil, M. (2013). [Profiting from an inefficient association football gambling market: Prediction, risk and uncertainty using Bayesian networks](#). *Knowledge-Based Systems*, 50(1), 60–86.
(Citado 3 vezes nas páginas 2, 3, e 12.)
- Cortes, C., & Vapnik, V. (1995). [Support-vector networks](#). *Machine Learning*, 20(3), 273–297.
(Citado 1 vez na página 13.)
- Dawson, A. (2018). [The 20 richest football clubs in the world](#). *Business Insider*.
(Citado 1 vez na página 2.)
- Dietze, G. (2012). [Intersektionalität im nationalen Strafraum: Race, Gender und Sexualität und die deutsche Nationalmannschaft](#). *Feministische Studien*, 30(1), 53–65.
(Citado 1 vez na página 8.)
- Dixon, M. J., & Coles, S. G. (1997). [Modelling Association Football Scores and Inefficiencies in the Football Betting Market](#). *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(2), 265–280.
(Citado 2 vezes nas páginas 2 e 10.)
- Dyte, D., & Clarke, S. R. (2000). [A ratings based Poisson model for World Cup soccer simulation](#). *Journal of the Operational Research Society*, 51(8), 993–998.
(Citado 2 vezes nas páginas 2 e 11.)
- Elo, A. E. (1978). *The Rating Of Chess Players, Past & Present*. Ann Arbor, MI: Arco Pub.
(Citado 1 vez na página 11.)
- Elsy, B., & Pugliese, S. G. (2017). *Football and the Boundaries of History*. Nova Iorque: Palgrave Macmillan US.
(Citado 2 vezes nas páginas 7 e 10.)
- Fernandez, J., Medina, D., Gomez, A., Arias, M., & Gavalda, R. (2016). [From Training to Match Performance: A Predictive and Explanatory Study on Novel Tracking Data](#). In *2016 IEEE 16th international conference on data mining workshops (ICDMW)*. IEEE.
(Citado 2 vezes nas páginas 2 e 12.)
- FIFA. (2014). [FIFA Financial Report 2014](#). *Reports*.
(Citado 3 vezes nas páginas 2, 10, e 39.)
- FIFA. (2017). [FIFA Financial Report 2017](#). *Reports*.
(Citado 2 vezes nas páginas 2 e 10.)
- FIFA. (2018). *Associations – CBF*. (Acessado em 30/05/2018.)
(Citado 1 vez na página 1.)
- FIFA. (2018). *World Cup Archives*.
(Citado 1 vez na página 2.)
- Freedman, D. (2005). *Statistical models: Theory and practice*. Cambridge: Cambridge University Press.
(Citado 1 vez na página 10.)
- Giske, R., Benestad, B., Haraldstad, K., & Høigaard, R. (2013). [Decision-Making Styles among Norwegian Soccer Coaches: An Analysis of Decision-Making Style in Relation to Elite and Non-Elite Coaching and Level of Playing History](#). *International Journal of Sports Science & Coaching*, 8(4), 689–701.
(Citado 1 vez na página 6.)
- Giulianotti, R. (2012). *Football*. Hoboken, NJ: The Wiley-Blackwell Encyclopedia of Globalization.
(Citado 1 vez na página 1.)
- Goddard, J., & Asimakopoulos, I. (2004). [Forecasting football results and the efficiency of fixed-odds betting](#). *Journal of Forecasting*, 23(1), 51–66.
(Citado 2 vezes nas páginas 3 e 11.)
- Goldstein, I., & Papert, S. (1977). [Artificial Intelligence, Language, and the Study of Knowledge](#). *Cognitive Science*, 1(1), 84–123.
(Citado 1 vez na página 14.)
- Governo Federal. (2014). *Caderno de turismo*. (Acessado em 10/06/2018.)
(Citado 1 vez na página 2.)
- Groll, A., Ley, C., Schauburger, G., & Eetvelde, H. (2018). [Prediction of the fifa world cup 2018 - a random forest approach with an emphasis on estimated team ability parameters](#). (In Press.)
(Citado 4 vezes nas páginas 2, 3, 13, e 39.)
- Haight, F. A. (1967). *Handbook of the poisson distribution (orsas)*. Hoboken, NJ: Wiley.
(Citado 1 vez na página 10.)
- Hamilton, M., Hoang, P., Layne, L., Murray, J., Padgett, D., CoreyStafford, & Tran, H. (2014). [Applying Machine Learning Techniques to Baseball Pitch Prediction](#). In *Proceedings of the 3rd international conference on pattern recognition applications and methods*. Angers, França: SCITEPRESS - Science and and Technology

Publications.

- (Citado 1 vez na página 2.)
- Hart, P. E. (1984). Artificial intelligence in transition. *AI Magazine*, 5(3), 17–20.
- (Citado 2 vezes nas páginas 14 e 15.)
- Haykin, S. (1998). *Neural networks: A comprehensive foundation* (2nd ed.). Upper Saddle River, NJ, USA: Prentice Hall PTR.
- (Citado 5 vezes nas páginas 3, 14, 16, 17, e 21.)
- Hennig, E. M., & Sterzing, T. (2010). [The influence of soccer shoe design on playing performance: A series of biomechanical studies](#). *Footwear Science*, 2(1), 3–11.
- (Citado 1 vez na página 6.)
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression (Wiley Series in Probability and Statistics)*. Hoboken, NJ: Wiley-Interscience Publication.
- (Citado 1 vez na página 11.)
- Howson, C., & Urbach, P. (2005). *Scientific reasoning: The bayesian approach*. Chicago, IL: Open Court Publishing Company.
- (Citado 1 vez na página 11.)
- Hughes, M., & Franks, I. (2005). [Analysis of passing sequences, shots and goals in soccer](#). *Journal of Sports Sciences*, 23(5), 509–514.
- (Citado 1 vez na página 6.)
- Hvattum, L. M., & Arntzen, H. (2010). [Using ELO ratings for match result prediction in association football](#). *International Journal of Forecasting*, 26(3), 460–470.
- (Citado 1 vez na página 11.)
- Kaelberer, M. (2016). [From Bern to Rio: Soccer and National Identity Discourses in Germany](#). *International Journal of Politics, Culture, and Society*, 30(3), 275–294.
- (Citado 2 vezes nas páginas 8 e 10.)
- Keogh, F., & Rose, G. (2013). [Football Betting - The Global Gambling Industry Worth Billions](#). *BBC - Sports*. (Acessado em 30/05/2018.)
- (Citado 1 vez na página 2.)
- Kittleston, R. (2017). Fausto dos Santos: The Wonders and Challenges of Blackness in Brazil's "Mulatto Football". In *Football and the Boundaries of History*. Nova Iorque: Palgrave Macmillan US.
- (Citado 1 vez na página 7.)
- Klapproth, F. (2008). [Time and decision making in humans](#). *Cognitive, Affective, & Behavioral Neuroscience*, 8(4), 509–524.
- (Citado 1 vez na página 22.)
- Koopman, S. J., & Lit, R. (2013). [A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League](#). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(1), 167–186.
- (Citado 2 vezes nas páginas 2 e 11.)
- Kose, U., & Vasant, P. (2017). [Fading intelligence theory: A theory on keeping artificial intelligence safety for the future](#). In *2017 international artificial intelligence and data processing symposium (IDAP)*. Malatya, Turquia: IEEE.
- (Citado 1 vez na página 15.)
- Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, 31(1), 249–268.
- (Citado 1 vez na página 16.)
- Kuhlman, D. (2012). *A python book: Beginning python, advanced python, and python exercises*. Cambridge, MA: MIT Press.
- (Citado 1 vez na página 4.)
- Langley, P. (2011). [The changing science of machine learning](#). *Machine Learning*, 82(3), 275–279.
- (Citado 1 vez na página 16.)
- Lechner, F. J. (2007). [Imagined communities in the global game: Soccer and the development of Dutch national identity](#). *Global Networks*, 7(2), 215–229.
- (Citado 2 vezes nas páginas 8 e 10.)
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). [Deep learning](#). *Nature*, 521(7553), 436–444.
- (Citado 1 vez na página 41.)
- Leung, C. K., & Joseph, K. W. (2014). [Sports Data Mining: Predicting Results for the College Football Games](#). *Procedia Computer Science*, 35(1), 710–719.
- (Citado 1 vez na página 2.)
- Maher, M. J. (1982). [Modelling association football scores](#). *Statistica Neerlandica*, 36(3), 109–118.
- (Citado 1 vez na página 10.)

- Martins, R. G., Martins, A. S., Neves, L. A., Lima, L. V., Flores, E. L., & do Nascimento, M. Z. (2017). [Exploring polynomial classifier to predict match results in football championships](#). *Expert Systems with Applications*, 83(1), 79–93.
(Citado 1 vez na página 13.)
- Maszczyk, A., Gołaś, A., Pietraszewski, P., Roczniok, R., Zając, A., & Stanula, A. (2014). [Application of Neural and Regression Models in Sports Results Prediction](#). *Procedia - Social and Behavioral Sciences*, 117(1), 482–487.
(Citado 1 vez na página 2.)
- Mazzoni, T. (1950). *História do futebol no brasil*. Rio de Janeiro: Leya.
(Citado 1 vez na página 1.)
- Milanovic, B. (2005). [Globalization and goals: Does soccer show the way?](#) *Review of International Political Economy*, 12(5), 829–850.
(Citado 1 vez na página 9.)
- Min, B., Kim, J., Choe, C., Eom, H., & McKay, R. B. (2008). [A compound framework for sports results prediction: A football case study](#). *Knowledge-Based Systems*, 21(7), 551–562.
(Citado 1 vez na página 2.)
- Minsky, M. (1961). [Steps toward Artificial Intelligence](#). *Proceedings of the IRE*, 49(1), 8–30.
(Citado 1 vez na página 13.)
- Ning, S., & Yan, M. (2010). [Discussion on research and development of artificial intelligence](#). In *2010 IEEE international conference on advanced management science(ICAMS 2010)*. Chengdu, China: IEEE.
(Citado 1 vez na página 15.)
- Oswald, V. (2014). [Brasil movimenta apenas 2% do mercado da bola](#). *O Globo*, 1(1).
(Citado 1 vez na página 39.)
- Pelak, C. F. (2009). [Women and gender in South African soccer: A brief history](#). *Soccer & Society*, 11(1-2), 63–78.
(Citado 1 vez na página 8.)
- Reep, C., & Benjamin, B. (1968). [Skill and Chance in Association Football](#). *Journal of the Royal Statistical Society. Series A (General)*, 131(4), 581.
(Citado 1 vez na página 6.)
- Rutkowski, L. (2004). [Adaptive Probabilistic Neural Networks for Pattern Classification in Time-Varying Environment](#). *IEEE Transactions on Neural Networks*, 15(4), 811–827.
(Citado 1 vez na página 19.)
- Samuel, A. L. (1959). [Some Studies in Machine Learning Using the Game of Checkers](#). *IBM Journal of Research and Development*, 3(3), 210–229.
(Citado 1 vez na página 15.)
- Schmidhuber, J. (2015). [Deep learning in neural networks: An overview](#). *Neural Networks*, 61(1), 85–117.
(Citado 1 vez na página 41.)
- Sorek, T. (2003). [Palestinian Nationalism has left the Field: A Shortened History of Arab Soccer in Israel](#). *International Journal of Middle East Studies*, 35(3), 417–437.
(Citado 1 vez na página 9.)
- Stekler, H., Sendor, D., & Verlander, R. (2010). [Issues in sports forecasting](#). *International Journal of Forecasting*, 26(3), 606–621.
(Citado 4 vezes nas páginas 2, 3, 37, e 40.)
- Suzuki, A. K., Salasar, L. E. B., Leite, J. G., & Louzada-Neto, F. (2010). [A Bayesian approach for predicting match outcomes: The 2006 \(Association\) Football World Cup](#). *Journal of the Operational Research Society*, 61(10), 1530–1539.
(Citado 2 vezes nas páginas 2 e 11.)
- Szymanski, S., & Smith, R. (1997). [The English Football Industry: Profit, performance and industrial structure](#). *International Review of Applied Economics*, 11(1), 135–153.
(Citado 1 vez na página 5.)
- Tam, K. Y., & Kiang, M. Y. (1992). [Managerial Applications of Neural Networks: The Case of Bank Failure Predictions](#). *Management Science*, 38(7), 926–947.
(Citado 2 vezes nas páginas 3 e 18.)
- Tsai, C., & Wu, J. (2008). [Using neural network ensembles for bankruptcy prediction and credit scoring](#). *Expert Systems with Applications*, 34(4), 2639–2649.
(Citado 2 vezes nas páginas 3 e 19.)
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 49(1), 433–460.
(Citado 2 vezes nas páginas 2 e 13.)
- Warner, B., & Misra, M. (1996). [Understanding Neural Networks as Statistical Tools](#). *The American Statistician*, 50(4), 284–293.
(Citado 1 vez na página 18.)
- Williams, A. M., & Hodges, N. J. (2005). [Practice, instruction and skill acquisition in soccer: Challenging tradition](#).

Journal of Sports Sciences, 23(6), 637–650.

(Citado 1 vez na página 6.)

Yao, X., & Liu, Y. (1997). [A new evolutionary system for evolving artificial neural networks](#). *IEEE Transactions on Neural Networks*, 8(3), 694–713.

(Citado 2 vezes nas páginas 18 e 19.)

Zhang, G., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1999). [Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis](#). *European Journal of Operational Research*, 116(1), 16–32.

(Citado 2 vezes nas páginas 3 e 19.)

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). [Forecasting with artificial neural networks: The state of the art](#). *International Journal of Forecasting*, 14(1), 35–62.

(Citado 3 vezes nas páginas 17, 18, e 21.)

Zobel, J. (2014). *Writing for Computer Science*. Londres: Springer London.

(Citado 1 vez na página 22.)

GLOSSÁRIO

CONMEBOL

Confederación Sudamericana de Fútbol. xvii, 10

DFB

Deutschen Fußball Bund. xvii, 8

FIFA

Fédération Internationale de Football Association. xvii, 1, 2, 10–12, 26, 31, 39, 41

ID3

Iterative Dichotomiser 3. xvii, 18

RF

Random Forests. xvii, 13

SVM

Support Vector Machines. xvii, 13

kNN

k-Nearest Neighbors. xvii, 18

AM

Aprendizagem de Máquina. xvii, 2–4, 10, 12, 13, 15, 16, 18, 19, 41

CBF

Confederação Brasileira de Futebol. xvii, 1

IA

Inteligência Artificial. ix, xvii, 2, 3, 5, 13–15, 19, 39

RNAPs

Redes Neurais Artificiais Probabilísticas. xvii, 19

RNAs

Redes Neurais Artificiais. ix, xiii, xvii, 3–5, 13, 16–19, 21–29, 31–37, 40–43, 57