

# Predição de resultados de futebol com Machine Learning

Victor Domiciano Moraes<sup>1</sup>

## Resumo

*Este trabalho busca conceber um modelo matemático capaz de prever resultados de partidas de futebol a partir de um treinamento prévio com partidas já realizadas e contidas em uma base de dados. Além, este trabalho sugere a implementação de algoritmos de aprendizado de máquina para afinar o modelo matemático e alcançar melhor taxa de acerto nas previsões.*

**Palavras-chave:** Aprendizado de Máquina. Artigo científico. Futebol. Predição de resultados.

## 1 – INTRODUÇÃO

O futebol é indubitavelmente uma paixão nacional. Acompanhada dessa paixão vem a alegria de ver o time do coração vencer jogos e, consequentemente campeonatos. Porém, essas partidas nem sempre são vencidas facilmente. Em tais partidas, há uma dosagem extra de emoção e ansiedade, até mesmo desespero, que acompanha o torcedor até o último minuto do segundo tempo.

Este tipo de emoção nem sempre é bem-vinda. O homem sempre quis prever o futuro das mais diferentes maneiras e nas mais diferentes áreas. Este trabalho tenta unir estas características e tem como objetivo fina prever o resultado de uma partida de futebol cujas equipes e data são previamente informados.

Esta é uma aplicação simples e que pode ser usada para entender, tanto na teoria quanto na prática os usos de algoritmos de aprendizado supervisionado de máquina.

## 2 – DESENVOLVIMENTO

Antes de se desenvolver o código da solução para o problema proposto, é preciso entender os desafios que acompanham este problema.

São eles: **(1)** Encontrar um volume de dados suficiente para o aprendizado **(2)** Como transformar e modelar esses dados de forma a se transformar numa entrada plausível para o modelo matemático? **(3)** Qual o melhor algoritmo de aprendizado para a predição?

### 2.1 – Primeiro desafio

Para o primeiro desafio, é preciso frisar que todos os tipos de problemas que são resolvidos com aprendizado de máquina supervisionado devem ter um conjunto de dados de treinamento de tamanho satisfatório com o objetivo de cobrir todas as possibilidades e saídas, assim permitindo que o algoritmo ‘conheça’ todo o escopo do problema e possa gerar saídas com grande nível de precisão.

No tocante à predição de resultados de partidas de futebol isso não é diferente. É necessário uma grande base de dados que contenha o histórico de milhares de partidas de futebol de várias temporadas. Por isso, foram estudados diversos conjuntos de dados e foi selecionado o conjunto de dados chamado ‘European Soccer Database’, que contém uma amostra de 25 mil partidas da principal liga de onze países da Europa entre os anos de 2008 e 2016.

Mais detalhadamente falando, essa base de dados possui informações sobre jogos de futebol, jogadores, equipes, países, campeonatos

---

<sup>5</sup> Autor correspondente: vdmoraes94@gmail.com

<sup>1</sup>Centro Federal de Educação Tecnológica de Minas Gerais

e habilidades técnicas de jogadores. Trata-se de um conjunto de dados bastante completo e apto a ser usado para a modelagem do problema proposto.

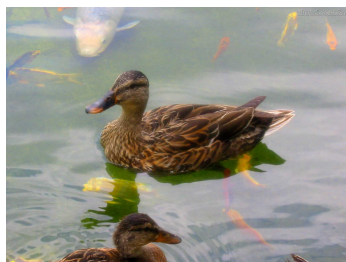
## 2.2 – Segundo desafio

Com a base de dados em mãos, foi necessário um estudo sobre como preparar os dados de entrada das amostras para servirem de dados de entrada para o algoritmo de aprendizado. A primeira decisão tomada foi a de não usar 100% dessas amostras, uma vez que em cada um dos onze países contidos na base de dados as influências no resultado de um jogo e até mesmo a maneira de praticar o esporte é diferente. Ou seja, há variáveis que influenciam mais do que outras no resultado de uma partida em um determinado campeonato, e essas variáveis podem influenciar menos em detrimento de outras em outra liga de outro país.

## 2.3 – Figura ocupando uma coluna

A **Figura 1** foi inserida, em uma única coluna, utilizando os comandos abaixo.

**Figure 1** – *Pato na lagoa. Fator de escala: 8% da original*



Artigos como mais de duas colunas não suportam o ambiente "figure" utilizado neste modelo  $\text{\LaTeX}$ . Uma alternativa para este problema é a inclusão do pacote:

```
\usepackage{caption}
```

e das seguintes linhas de comando:

```
\begin{center}
\captionof{figure}{<caption da figura>}
\includegraphics[<comandos alternativos>]
{<caminho ou nome da figura>}
\label{<nome da referencia da figura>}
\end{center}
```

## 2.4 – Figura ocupando duas colunas

O ambiente **figure** pode ser usado em um artigo quando a figura for centralizada entre as margens do artigo (ocupa um espaço maior que uma coluna). Porém é necessário a introdução do \* após o comando **figure**.

A **Figura 2** foi inserida no artigo utilizando o comando **figure\*** para o ambiente de figura.

```
\begin{figure*}
\centering
\caption{Pato na lagoa. Fator de escala: 20\% da original}
\includegraphics[scale=0.2]{../02-figuras/pato}
\label{fig:patoB}
\end{figure*}
```

## 2.5 – Tabelas

O ambiente de tabelas é inserido no texto de modo análogo àquele feito no ambiente de figuras.

## 2.6 – Citações de referências

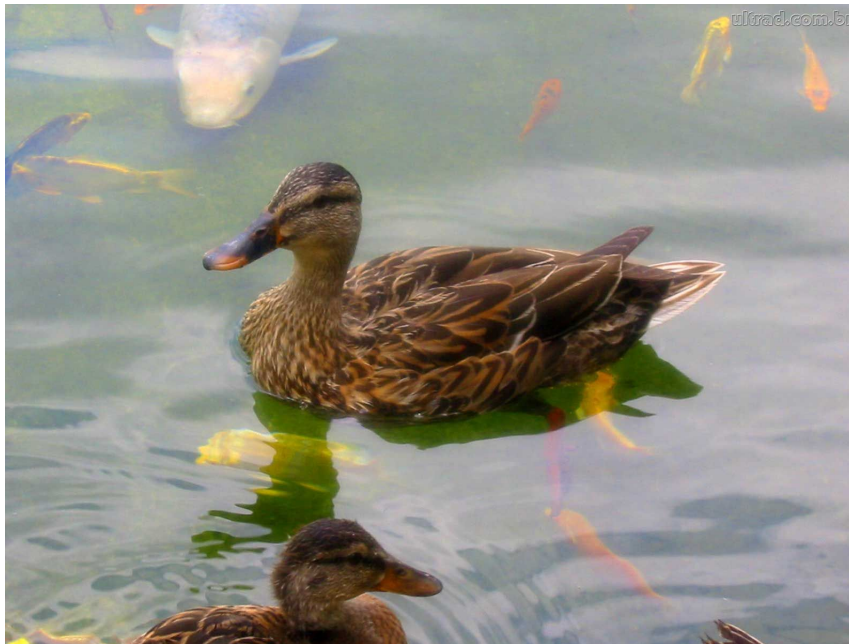
As referências são inseridas no texto como em qualquer documento em  $\text{\LaTeX}$ . Quando o nome do autor da referência faz parte do texto que você está escrevendo use o comando **\citeonline{}** e quando este não for o caso use o comando **\cite{}**. Veja a diferença entre os dois nas seguintes frases:

- (1) Conforme discute [Kim et al. \(1996\)](#) o resultado [...].
- (2) Alternativamente a literatura([KIM et al., 1996](#)) indica que o resultado [...].

Quando se tem mais de uma referência a ser citada em um mesmo certo trecho, há duas possibilidades de referenciá-las:

- (1) colocar todas as referências em um único colchete (*i.e.*, num mesmo comando **\cite**),  
([KIM et al., 1996](#); [WIKIBOOKS, 2009](#))

**Figure 2** – *Pato na lagoa. Fator de escala: 20% da original*



(2) colocar cada referência em seu próprio colchete (*i.e.*, usando vários comandos `\cite` consecutivos),

(KIM et al., 1996)(WIKIBOOKS, 2009)

É obrigatório o agradecimento às instituições de fomento à pesquisa que financiaram total ou parcialmente o trabalho, inclusive no que diz respeito à concessão de bolsas.

### 3 – CONCLUSÃO

Edite esta seção para colocar a conclusão de seu trabalho de pesquisa.

Procure fazer uma análise crítica de seu trabalho, destacando os principais resultados e as contribuições deste trabalho para a área de pesquisa.

Também deve indicar, se possível e/ou conveniente, como este trabalho pode ser estendido ou aprimorado.

### ABSTRACT

Translation of the abstract into english, possibly adapting or slightly changing the text in order to adjust it to the grammar of Standard English.

### BIBLIOGRAPHY

### AGRADECIMENTOS

Edite e coloque aqui os agradecimentos às pessoas e/ou instituições que contribuíram para a realização do trabalho.

KIM, J. I. et al. Perturbative expansion for coherence loss. *Phys. Rev. Lett.*, American Physical Society, v. 77, p. 207–210, Jul 1996. Disponível em: <http://link.aps.org/doi/10.1103/PhysRevLett.77.207>.

WIKIBOOKS. *LaTeX*. 2009. Disponível em: <http://en.wikibooks.org/wiki/LaTeX>.

Acesso em: 8 de novembro de 2009.