

# Football Events

---

## Temática 3: Aprendizagem Supervisionada – Classificação

Gustavo Magalhães - [up201705072@fe.up.pt](mailto:up201705072@fe.up.pt)

Nuno Cardoso - [up201706162@fe.up.pt](mailto:up201706162@fe.up.pt)

Tiago Silva - [up201705985@fe.up.pt](mailto:up201705985@fe.up.pt)

Turma 1

10/05/2020

# Especificação

Através de algoritmos de aprendizagem supervisionada por classificação, pretendemos que o computador agrupe os dados dos vários eventos que ocorrem durante jogos de futebol, com o objetivo final de o capacitar a classificar jogos quanto ao seu país de origem e a determinar a equipa vencedora.

Estes eventos podem ser faltas, remates, golos entre outros, identificando-se também a zona no campo da sua ocorrência no decorrer de uma partida de futebol.

Dado o vasto e completo conjunto de dados, decidimos fazer uma análise a dois tipos de problemas, de modo a identificar qual fará mais sentido explorar. São estes:

- **Classificar o país onde ocorre o jogo** consoante o número de ocorrências de golos, faltas, cartões, penáltis, cantos, contra-ataques e número de vezes em que se rematou a bola numa dada zona do campo;
- **Classificar o vencedor do jogo ou empate** não contabilizando o número de golos, usando para isso, número de faltas, cartões, penáltis, cantos, contra-ataques e número de remates de uma dada zona do campo, isto tanto para a equipa da casa como a visitante.

# Referências

1. Duarte, Luís Miguel da Silva. Projeto de Dissertação, 1x2 - Previsão de Resultados de Jogos de Futebol. FEUP. Porto. 29 de Junho de 2015. <https://repositorio-aberto.up.pt/bitstream/10216/79327/2/35444.pdf>.
2. Fernandes, Felipe Augusto Pereira. Previsão de Resultados no Futebol Por Meio de Técnicas de Aprendizado de Máquina. CEFET-MG. Belo Horizonte. Fevereiro de 2019.  
<https://sig.cefetmg.br/sigaa/verArquivo?idArquivo=2256419&key=6a8ee9c8bd115b18683f4b40fb5ec585>.
3. Olson, Randal S. Data-Analysis-and-Machine-Learning-Projects. University of Pennsylvania Institute for Bioinformatics. 3 de Julho de 2019.  
<https://github.com/rhiever/Data-Analysis-and-Machine-Learning-Projects/blob/master/example-data-science-notebook/Example%20Machine%20Learning%20Notebook.ipynb>.

# Identificação de Algoritmos

Iremos implementar e testar o nosso projeto com os seguintes algoritmos:

- Árvore de decisão;
- K-ésimo vizinho mais próximo;
- Rede neuronal artificial;

# Descrição das Ferramentas a Usar

Uma vez que a linguagem a usar será **Python na versão 3.7**, para facilitar a nossa implementação iremos usar as seguintes bibliotecas:

- **Numpy** - para uso facilitado de matrizes e arrays multidimensionais;
- **Pandas** - para guardar dados a analisar de modo fácil e eficiente;
- **Seaborn** e **Matplotlib** - para visualização de gráficos e análise estatística;
- **Scikit-Learn** - para auxílio na implementação dos algoritmos de machine learning;

Iremos também utilizar **Jupyter** para agrupar toda a análise estatística de dados feita, assim como a sua exploração.

# Trabalho realizado até ao momento

Com programação em Python e com recurso ao conjunto de dados fornecidos, construímos dois *data sets*, de modo a que os dados de um jogo fossem agrupados numa linha de uma tabela. Deste modo, foram criados dois ficheiros .csv distintos, correspondentes às duas opções de classificação apresentadas:

1. gameEvents.csv - com dados de classificação relativos ao país onde o jogo decorreu;
2. gameEvents[HomeAway].csv - com dados de classificação relativos ao vencedor do jogo.