



# Inteligência Artificial em Apostas Desportivas

**PEDRO ANTÓNIO DE MOURA ALVES**

Outubro de 2022





# **Inteligência Artificial em Apostas Desportivas**

**Pedro António de Moura Alves**

**1150372**

**Dissertação para obtenção do Grau de  
Mestre em Engenharia de Inteligência Artificial**

**Orientador: Doutor Tiago Manuel Campelos Ferreira Pinto, Professor Adjunto Convidado do Instituto Superior de Engenharia do Instituto Politécnico do Porto e Professor Auxiliar da Universidade de Trás-os-Montes e Alto Douro**

**Coorientador: Doutor Carlos Fernando da Silva Ramos, Professor Coordenador Principal do Instituto Superior de Engenharia do Instituto Politécnico do Porto**

**Júri:**

Presidente:

Doutora Isabel Cecília Correia da Silva Praça Gomes Pereira, Professora Coordenadora do Instituto Superior de Engenharia do Instituto Politécnico do Porto

Vogais:

Doutor Luiz Felipe Rocha de Faria, Professor Coordenador do Instituto Superior de Engenharia do Instituto Politécnico do Porto

Doutor Tiago Manuel Campelos Ferreira Pinto, Professor Adjunto Convidado do Instituto Superior de Engenharia do Instituto Politécnico do Porto e Professor Auxiliar da Universidade de Trás-os-Montes e Alto Douro

Porto, outubro 2022



«A maneira como se recolhe, gere e utiliza a informação,  
determina se uma pessoa vence ou não» Bill Gates



# Resumo

Atualmente, os dados simbolizam a matéria-prima mais valiosa do planeta. Através deles, é possível aceder a informações vitais para a sobrevivência e sucesso das empresas de todo mundo.

Neste contexto, a Inteligência Artificial (IA) assume um papel preponderante na obtenção de informação valiosa “escondida” nas diversas formas que os dados podem assumir.

Feitas estas considerações, e sabendo das proporções que a importância da IA está a tomar nos últimos anos no processo de tomada de decisão, foi desenvolvido este projeto, no âmbito da unidade curricular Projeto/Dissertação/Estágio (PROJIA). Este visa a análise dos dados existentes no mundo das apostas desportivas, para que, através destes, seja possível retirar informação valiosa para o aumento do sucesso dos prognósticos em jogos de futebol.

Para este efeito, recorreu-se a dados extraídos através de uma Application Programming Interface (API), os quais foram, posteriormente, tratados para servirem de base de informação do modelo. determinando este, por sua vez, o vencedor num determinado jogo de futebol, através de algoritmos de Deep Learning (DL).

Posto isto, com este modelo, foi possível constatar que a probabilidade de acertar o vencedor de um jogo de futebol aumenta em ligas com um maior número de jogos que ocorram de forma mais regular. Já em campeonatos em que participem clubes de diferentes países, e em que os jogos entre as equipas encerrem um grande intervalo temporal, o modelo não é capaz de acertar tão eficazmente na equipa vencedora.

**Palavras-chave:** Inteligência Artificial, Apostas Desportivas, *Machine Learning*, *Deep Learning*, *RandomForestClassifier*, LSTM, CNN



# Abstract

Nowadays, data is one of the most valuable raw materials on the planet. Through them it is possible to obtain information vital for the viability and success of companies around the world.

In this context, Artificial Intelligence (AI) assumes a leading role in obtaining valuable information "hidden" in the various forms that data can have.

Having made these considerations and knowing the proportions that AI is portraying in the decision-making process, the importance of this project, that was developed in the scope of the curricular unit Project/Dissertation/Internship (PROJIA), can be appraised. Aiming at the analysis of existing data in the world of sports betting, so that, through these, it is possible to extract valuable information to increase the success of predictions in football games.

For this purpose, data extracted through an Application Programming Interface (API) was used and then processed to feed the model's database and consequently is used to determine the winner in football matches through Deep Learning (DL) algorithms.

With this model's output, it was possible to observe that the probability of determining the winner of a football match is superior in leagues with a larger number of matches that occur more regularly. On the other hand, in leagues where clubs from different countries participate, and where the matches span over a larger time interval, the model is not able to determine the winning team as effectively.

**Keywords:** Artificial Intelligence, Sports Betting, Machine Learning, Deep Learning, RandomForestClassifier, LSTM, CNN



# Agradecimentos

Agradeço à minha família,

por todos os ensinamentos que me foram transmitidos ao longo da vida e pelo apoio incondicional no decorrer de todo o meu percurso académico.

Agradeço à minha namorada, Inês Simões,

pela ajuda e disponibilidade sempre demonstrada para comigo, mesmo nos momentos mais difíceis da minha vida, assim como pelo seu contributo prestimoso na revisão deste documento.

Agradeço à Tuna Académica do ISEP,

que me fez crescer enquanto ser humano, desempenhando um papel vital na minha adaptação a uma nova e magnífica cidade.

Por último, agradeço a todos os meus colegas que acompanharam o meu percurso académico, contribuindo, direta ou indiretamente, no desenvolvimento deste projeto.



# Índice

<b>1</b>	<b>Introdução .....</b>	<b>19</b>
1.1	Enquadramento.....	19
1.2	Descrição do Problema .....	20
1.2.1	Objetivos .....	20
1.2.2	Abordagem .....	20
1.2.3	Planeamento do Trabalho.....	21
1.3	Estrutura do Documento .....	21
<b>2</b>	<b>Estado da Arte .....</b>	<b>23</b>
2.1	Dados .....	23
2.1.1	Estrutura de Dados .....	23
2.1.2	Tratamento de Dados.....	24
2.2	Inteligência Artificial .....	24
2.2.1	Conceito de IA .....	25
2.2.2	Machine Learning.....	26
2.3	Trabalhos Relacionados .....	30
2.3.1	Modelos de apostas usando IA .....	30
2.3.2	Previsão de partidas de futebol .....	31
<b>3</b>	<b>Análise e desenho da solução.....</b>	<b>33</b>
3.1	O mundo das apostas desportivas.....	33
3.2	Obtenção dos resultados e dados estatísticos das partidas de futebol.....	34
3.3	Desenho da solução .....	37
3.3.1	Arquitetura dos dados inicialmente extraídos .....	37
3.3.2	Arquitetura da implementação.....	38
3.4	Requisitos .....	40
3.4.1	Utilizadores e as suas necessidades .....	40
3.4.2	Requisitos Funcionais .....	41
3.4.3	Requisitos Não Funcionais.....	41
<b>4</b>	<b>Implementação da solução .....</b>	<b>43</b>
4.1	Tecnologias Adotadas.....	43
4.2	Linguagens Adotadas.....	44
4.3	Bibliotecas utilizadas .....	44
4.4	Descrição da implementação .....	44
4.5	Avaliação da Solução .....	48
4.5.1	Métricas Usadas.....	48
4.5.2	Liga dos Campeões .....	49
4.5.3	Liga Francesa .....	50

4.5.4	Liga Portuguesa.....	51
4.5.5	Liga Inglesa .....	52
4.5.6	Análise Comparativa.....	52
<b>5</b>	<b>Conclusão .....</b>	<b>55</b>
5.1	Objetivos concretizados.....	55
5.2	Trabalho Futuro e Limitações .....	55
5.3	Apreciação final .....	56
	<b>Bibliografia .....</b>	<b>57</b>
	<b>Anexos .....</b>	<b>61</b>
Anexo 1	Glossário dos Requisitos.....	61
Anexo 2	Extração dos Dados da API .....	62
Anexo 3	Tratamento e Análise dos Resultados .....	63

# Lista de Figuras

Figura 1 - Inteligência Artificial, <i>Machine Learning</i> e <i>Deep Learning</i> .....	25
Figura 2 - Tipos de ML [23].....	26
Figura 3 - Partes constituintes de um neurónio [28] .....	27
Figura 4 – Possível desenho de uma ANN [26]. .....	28
Figura 5 – Arquitetura de uma LSTM [33].....	29
Figura 6 - Fluxograma de uma <i>RandomForestClassifier</i> [36] .....	30
Figura 7 - Exemplo de pedido em <i>Python</i> à API de resultados .....	35
Figura 8 - Exemplo de respostas no formato JSON.....	36
Figura 9 - Arquitetura da API de resultados.....	37
Figura 10 - Desenho da arquitetura implementada.....	39
Figura 11 - Desenho da arquitetura implementada com novos dados obtidos .....	39
Figura 12 – Organização dos ficheiros .....	45
Figura 13 - Exemplo de json que a API retorna.....	45
Figura 14 – Normalização de uma variável json num dataframe .....	45
Figura 15 – Filtragem das colunas a serem analisadas .....	47
Figura 16 – Implementação da Solução .....	47
Figura 17 – <i>Accuracy and Precision</i> [42].....	48
Figura 18 - Parte do código <i>Python</i> de Extração dos Dados da API .....	62
Figura 19 - Parte do código Python de Tratamento dos Dados da API.....	63



# Lista de Tabelas

Tabela 1 - Planeamento do Trabalho .....	21
Tabela 2 - <i>Google Colab vs Jupyter Notebook</i> .....	43
Tabela 3 – Resultados obtidos na Liga dos Campeões com 3 possíveis resultados de <i>output</i> usando RandomForestClassifier.....	49
Tabela 4 - Resultados obtidos na Liga dos Campeões com 2 possíveis resultados de <i>output</i> usando RandomForestClassifier.....	50
Tabela 5 - Resultados obtidos na Liga Francesa com 3 possíveis resultados de <i>output</i> usando RandomForestClassifier .....	50
Tabela 6 - Resultados obtidos na Liga Francesa com 2 possíveis resultados de <i>output</i> usando RandomForestClassifier .....	51
Tabela 7 - Resultados obtidos na Liga Portuguesa com 3 possíveis resultados de <i>output</i> usando RandomForestClassifier .....	51
Tabela 8 - Resultados obtidos na Liga Portuguesa com 2 possíveis resultados de <i>output</i> usando RandomForestClassifier .....	51
Tabela 9 - Resultados obtidos na Liga Inglesa com 3 possíveis resultados de <i>output</i> usando RandomForestClassifier .....	52
Tabela 10 - Resultados obtidos na Liga Inglesa com 2 possíveis resultados de <i>output</i> usando RandomForestClassifier .....	52
Tabela 11 – Comparação entre a Liga dos Campeões e a Liga Francesa .....	52
Tabela 12 – Comparação entre a Liga dos Campeões e a Liga Portuguesa .....	53
Tabela 13 – Comparação entre a Liga dos Campeões e a Liga Inglesa .....	53
Tabela 14 - Glossário de Requisitos .....	61



## Lista de Acrónimos

<b>AA</b>	Aprendizagem Automática
<b>SRIJ</b>	Serviço de Regulação e Inspeção de Jogos
<b>CAGR</b>	<i>Compound Annual Growth Rate</i>
<b>IA</b>	Inteligência Artificial
<b>LSTM</b>	<i>Long-Short Term Memory</i>
<b>SVM</b>	<i>Support Vector Machine</i>
<b>DL</b>	<i>Deep Learning</i>
<b>ML</b>	<i>Machine Learning</i>
<b>API</b>	<i>Application Programming Interface</i>



# 1 Introdução

Neste capítulo, é realizada uma breve apresentação do trabalho desenvolvido, no âmbito do projeto de dissertação, para a cadeira de Projeto/Dissertação/Estágio (PROJIA).

Inicialmente, é feito um enquadramento do projeto e uma breve descrição dos seus problemas e objetivos.

De seguida, é realizado um planeamento do trabalho a ser desenvolvido e respetivas datas.

Por último, encontra-se a explicação de cada um dos capítulos que compõem a estrutura do presente documento.

## 1.1 Enquadramento

Ano após ano, o mercado global das apostas desportivas tem vindo a crescer, tendo sido avaliado no valor de 76,75 biliões de dólares, em 2021, com uma previsão da taxa de crescimento anual composta (CAGR) de 10% ao longo dos próximos 9 anos [1].

Em Portugal, estima-se que, diariamente, são movimentados 3,7 milhões de euros, de acordo com o último relatório, de agosto de 2022, do Serviço de Regulação e Inspeção de Jogos (SRIJ), sendo que desse valor 65,8% é gasto em apostas desportivas relacionados com o futebol [2].

Aos números supracitados, devemos somar o número de pessoas e grupos que vendem os seus prognósticos desportivos, prestando uma espécie de serviço de consultoria a milhares, senão milhões de pessoas que anseiam obter dinheiro sem qualquer tipo de esforço [3].

É importante salientar que, em Portugal, existem, em média, cerca de 1800 novos registos diários em entidades exploradoras destes mercados [2], dos quais 58,6% dos apostadores apresentam menos de 35 anos e 81,9% menos de 45 anos [2].

Considerando o crescimento acentuado do número de jogadores portugueses, e atendendo ao facto da população mais jovem representar a maior fatia de apostadores, surge a necessidade de prestar um serviço de qualidade, intuitivo e automatizado. Desta forma, aqueles passam a ter a possibilidade de usufruir de um programa de consultoria em apostas desportivas que os possa auxiliar a maximizar os seus lucros, otimizando a gestão dos riscos.

Por outro lado, as casas de apostas desportivas pretendem maximizar os seus lucros através dos ajustes que vão realizando às probabilidades de determinados eventos acontecerem à medida que o tempo passa e novas apostas são feitas.

O cálculo destas probabilidades, sobretudo quando feito de forma dinâmica de acordo com os acontecimentos em tempo real, é uma tarefa complexa e que exige a análise de grandes

quantidades de dados correlacionados. A IA apresenta diversas abordagens interessantes e promissoras para lidar com problemas deste género.

O presente projeto tem como principal intuito a construção de um modelo de Inteligência Artificial (IA) capaz de prever a equipa vencedora num determinado jogo de futebol.

Para a implementação do projeto, foi concebido um processo de extração e tratamento dos dados e, posteriormente, procedeu-se ao desenvolvimento de um modelo de IA capaz de prever os resultados, recorrendo à aprendizagem supervisionada, utilizando nomeadamente algoritmos de *Deep Learning*.

## **1.2 Descrição do Problema**

Nos dias que correm, os dados apresentam-se como uma das matérias mais valiosas para as empresas e para um grande número de pessoas. Deste modo, para que sejam utilizados de forma eficiente, têm de ser tratados e analisados de forma correta.

Esta análise requer uma automatização no processo de tratamento e análise dos dados para que seja possível extrair o máximo de informação possível, no caso deste trabalho, relativamente aos jogos de futebol.

Para isso, é necessário desenvolver um modelo de IA capaz de entender o que poderá vir a acontecer num jogo de futebol.

### **1.2.1 Objetivos**

Este projeto teve como principal objetivo a previsão do resultado de jogos de futebol.

Os restantes objetivos inerentes ao projeto são os seguintes:

- Levantar o estado arte relacionado com extração e tratamento de dados, modelos de previsão baseados em IA, e soluções para previsão no âmbito de apostas desportivas;
- Implementar um processo de extração e tratamento de dados;
- Implementar um modelo de IA para previsão da equipa vencedora de um jogo de futebol;
- Analisar os resultados e retirar valor de negócio com a informação gerada pelo modelo de IA;
- Discutir e reportar as principais conclusões tiradas através deste estudo.

### **1.2.2 Abordagem**

De forma a concretizar os objetivos supracitados, foi realizado um estudo dos trabalhos relacionados com o tema em questão, bem como das tecnologias e metodologias a adotar.

Posteriormente, foi realizada uma análise ao mundo das apostas desportivas e às bases de dados existentes que poderiam facultar os dados necessários para este projeto.

Por último, foi delineado o planeamento de todo o trabalho, desde o estudo dos trabalhos relacionados, até à implementação e análise do modelo de IA.

### 1.2.3 Planeamento do Trabalho

É possível observar, na Tabela 1, o planeamento do projeto.

Tabela 1 - Planeamento do Trabalho



Nos primeiros 4 meses, procedeu-se ao estudo do modelo preliminar do projeto a ser, posteriormente, implementado, assim como à análise de projetos similares, que perseguiram o mesmo objetivo. Neste sentido, foram estudadas as tecnologias adotadas por esses projetos e, paralelamente, foram determinadas aquelas que seriam adequadas à implementação deste projeto em específico.

Posteriormente, realizaram-se os processos de extração e tratamento de dados e, logo de seguida, iniciou-se o desenvolvimento do modelo de Inteligência Artificial.

Por último, efetuou-se uma análise aos resultados obtidos, sendo esta acompanhada pelo término da redação do presente documento.

## 1.3 Estrutura do Documento

Este documento encontra-se dividido em cinco capítulos, cada um com o seu tema e respetiva importância para o entendimento e visão holística do relatório.

O **capítulo 1** tem como objetivo apresentar o trabalho desenvolvido e a estrutura do relatório.

No **capítulo 2**, é realizada uma breve introdução aos temas necessários para o desenvolvimento do projeto e dos seus componentes. De igual modo, são apresentadas as diferentes tecnologias que existem para a implementação do mesmo.

Em relação ao **capítulo 3**, verifica-se que este apresenta a análise e o desenho da solução, onde se pode encontrar a introdução ao tema “O Mundo das Apostas Desportivas”. Cumulativamente, é realizada uma apresentação dos locais de extração dos dados estatísticos para o desenvolvimento do projeto, sendo ainda possível ver o desenho da solução e a sua arquitetura.

Após a análise e desenho da solução estarem concluídos, segue-se a fase da sua implementação, que se encontra descrita no **capítulo 3.4**. Neste capítulo, também é possível encontrar a avaliação e os resultados obtidos relativamente à solução implementada, no ponto **4.5**.

Por último, é realizada a conclusão do projeto, tanto ao nível dos objetivos concretizados como ao nível do trabalho a desenvolver no futuro. Esta é inserida no **capítulo 5**, que contém, de igual forma, a apreciação final de todo o projeto.

## 2 Estado da Arte

Este capítulo, inicialmente, expõe o conceito de dados e de todo o processo de extração e tratamento necessário para a alimentação de um modelo de IA.

De seguida, é apresentado um pouco do mundo da Inteligência Artificial e dos principais algoritmos que poderão ser úteis na implementação do projeto.

Por último são apresentados e analisados alguns trabalhos relacionados com o projeto desenvolvido.

### 2.1 Dados

Os dados correspondem a um conjunto de valores, num estado bruto, a partir dos quais se podem retirar informações úteis para o negócio [4].

A obtenção de um dado pode ter uma diminuta importância para uma empresa/pessoa. Contudo, quando a empresa/pessoa obtém milhões de dados, estes assumem uma extrema relevância, uma vez que, com processos de tratamento de dados e com recurso a modelos de IA, é possível retirar mais informação. Esta, por sua vez, poderá ser utilizada para aumentar o conhecimento da empresa/pessoa em relação aos dados em questão de uma forma mais ampla e precisa [5][6].

#### 2.1.1 Estrutura de Dados

No que concerne à sua organização, os dados podem-se enquadrar em três tipos [7]:

- Estruturados
  - Estrutura estática e previamente planeada;
  - Todas as estruturas de dados têm as suas relações definidas;
  - Ex: Base de Dados Relacional.
- Semiestruturados
  - Estrutura intermédia com parte devidamente organizada planeada e outra parte sem estrutura previamente definida;
  - O Esquema de representação vem em conjunto com os dados;
  - Ex: XML.
- Não Estruturados
  - Sem estrutura previamente definida;
  - Cerca de 80% dos dados das empresas;
  - Ex: Relatórios, Documentos, Imagens, Áudios, Vídeos.

Os dados previamente estruturados necessitam de menos memória e são de mais fácil gestão, enquanto os dados não estruturados requerem mais memória e apresentam uma maior dificuldade na sua gestão e proteção. Neste projeto, os dados devem estar devidamente estruturados e interligados para que o modelo de AI consiga compreender os diversos conjuntos de dados, definindo assim os atributos necessários para a aprendizagem do modelo [8].

### **2.1.2 Tratamento de Dados**

A grande maioria dos dados que se encontram disponíveis nas mais variadas bases de dados encontram-se incorretamente estruturados para a alimentação de um modelo de AI, sendo necessário realizar um processo de tratamento de dados para que este seja capaz de extrair a máxima informação dos dados [9].

O processo de tratamento de dados consiste numa série de transformações às quais os mesmos são sujeitos, para que daí advenha uma uniformização das mais diversas fontes de dados, bem como a eliminação de ruído que possa existir [10]:

- Transformações Básicas:
  - Limpeza: Trocar os NULL por 0, “Male” por “M”, formatações de data, etc;
  - Duplicação: Eliminação dos registos repetidos;
  - Revisão do formato dos dados;
  - Restruturação das chaves: Restruir as relações entre tabelas.
- Transformações Avançadas:
  - Derivação: Criação de novas métricas com base no negócio;
  - Filtrar: Selecionar apenas algumas linhas/colunas/tabelas;
  - Juntar: Juntar dados de diferentes tabelas mediante o modelo dimensional;
  - Cortar: Transformar uma coluna em inúmeras colunas;
  - Validar Dados: Através das regras de negócio, validar as diferentes linhas;
  - Agregação: Juntar dados de diferentes fontes de dados;
  - Integração: Definir cada elemento que seja único com um valor padrão.

## **2.2 Inteligência Artificial**

A Inteligência Artificial (IA) atravessou uma grande evolução nos últimos anos. Neste contexto, os pequenos aparelhos de hardware, inicialmente, utilizados deram lugar a redes neuronais e sistemas ambientes [11]. Estes requerem uma grande quantidade de dados, que se deverão apresentar de forma limpa e tratada.

Efetivamente, a qualidade dos dados recebidos e o tratamento a que são submetidos constituem fatores determinantes do sucesso do desenvolvimento destes sistemas [12].

### 2.2.1 Conceito de IA

A Inteligência Artificial emergiu com uma grande influência para um grande número de pessoas durante muito tempo [13], tendo sido utilizada pela primeira vez por *Allan Turing*, em 1950, questionando se **as máquinas conseguiriam pensar** [14].

A IA, igualmente denominada de Inteligência Demonstrada Pelas Máquinas ou Computadores [15], tem sido utilizada, ao longo dos últimos anos, em diversas áreas, como, por exemplo, na associação de processos de **deteção e previsão**, sendo estes, por vezes, comparados a um cérebro humano [16].

Existem vários ramos de investigação associados à IA, destacando-se o **Machine Learning (ML)**. Este consiste na incorporação de algoritmos capazes de ensinarem a máquina a ser inteligente, recorrendo a um processo automático de aprendizagem capaz de retirar inteligência dos dados existentes nas inúmeras fontes [13] [17].

É bastante usual pensar, erroneamente, que a IA é igual ao *Machine Learning*. Este aspeto constitui um argumento totalmente inválido, já que a IA pode ser um sistema codificado em diversos conjuntos de regras que são capazes de tomar decisões, à medida que os novos dados vão chegando, sem intervenção humana. Esse conjunto de regras pode ser definido manualmente ou através de outras técnicas, como, por exemplo, o ML, bastando, para isso, fornecer um conjunto de dados históricos para que a máquina seja capaz de aprender sozinha [13]. Na Figura 1, é possível observar a relação do ML com a IA, assim como um subdomínio do ML, o *Deep Learning*.

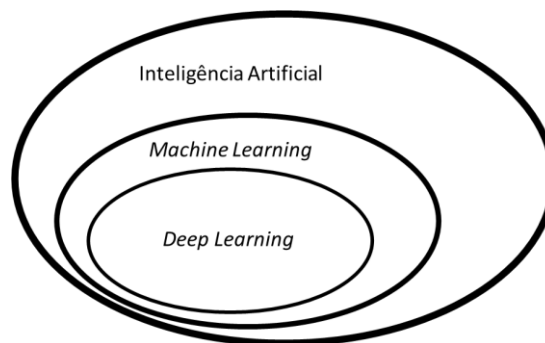


Figura 1 - Inteligência Artificial, *Machine Learning* e *Deep Learning*

O DL desencadeou uma grande revolução no ML, devido à similaridade com os mecanismos do cérebro humano. Este subdomínio é capaz de resolver processos complexos através de várias camadas de unidades matemáticas mais simples que, em conjunto, solucionam estes desafios [18]. Uma das vantagens do DL é a sua capacidade de aprendizagem a partir de dados brutos, heterogêneos e que apresentam bastante ruído [19].

## 2.2.2 Machine Learning

Machine Learning é a ciência de programar uma máquina para que esta seja capaz de retirar conhecimento através dos dados fornecidos [20].

De acordo com Arthur Samuel, o ML é a área de investigação que fornece às máquinas a capacidade de aprendizagem sem que esta seja, anteriormente e de forma explícita, programada [21].

Numa definição mais orientada para a engenharia moderna, Tom Mitchell define este conceito como sendo a capacidade de um programa de computador aprimorar a execução de uma determinada tarefa, após ter tido alguma experiência que lhe permita reter informação, e, desta forma, aprender a otimizar a tarefa previamente realizada [22].

O ML abrange diversos algoritmos/processos que são usados para resolver as diversas tarefas, aprendendo através dos dados fornecidos. A Figura 2 ilustra uma árvore dos tipos de implementação e abordagens que se poderão seguir dentro do ML.

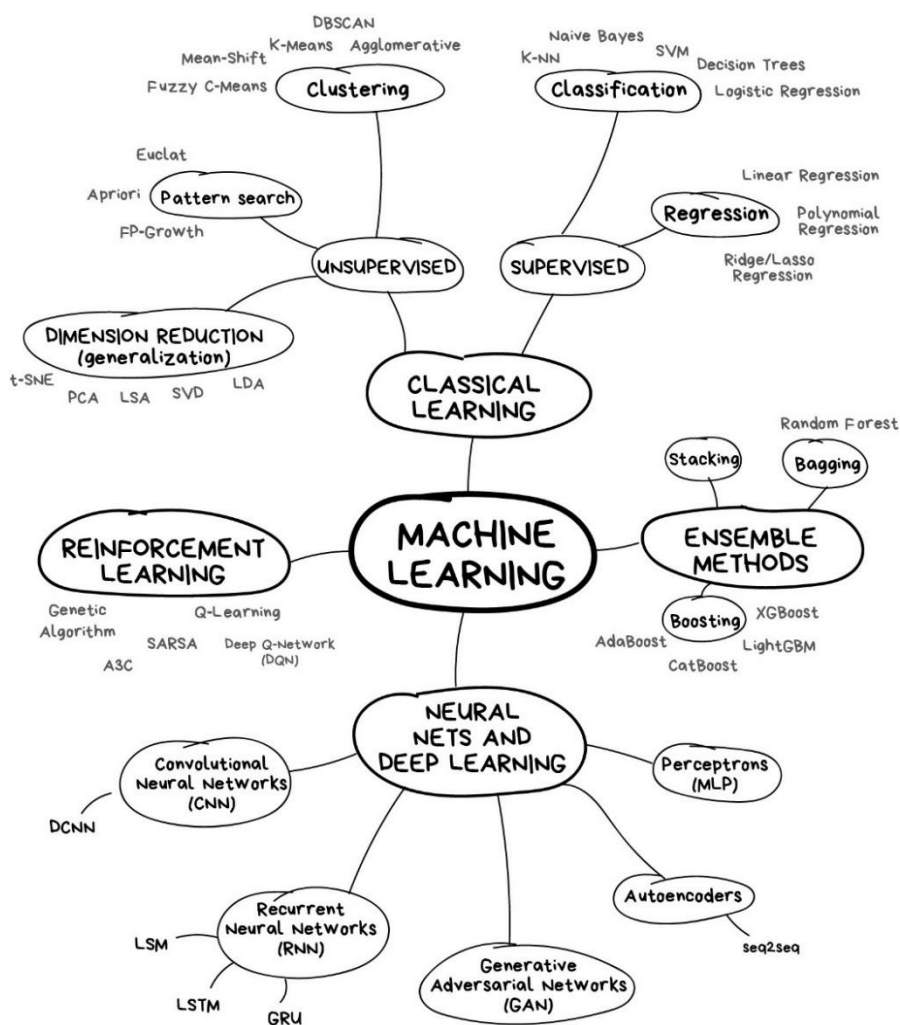


Figura 2 - Tipos de ML [23]

No caso das **aprendizagens automáticas supervisionadas**, quando os algoritmos processam os dados, estas conseguem reter toda a informação necessária, bem como o resultado esperado em cada situação, e, com base nisto, são capazes de aplicar algoritmos de classificação/regressão para perceber de forma automática a relação entre os mais diversos dados [20]. Ou seja, deste modo, é possível estabelecer correlações entre as diversas métricas e o resultado final.

No que diz respeito às **aprendizagens não supervisionadas**, o computador pretende adquirir conhecimentos através de algoritmos que identifiquem homogeneidade nos diversos elementos fornecidos sem saberem o resultado final no decorrer do processo de aprendizagem [24].

Quando a quantidade de dados começa a tomar proporções acrescidas, e, simultaneamente, o processo de aprendizagem e de estudo de dados complexifica, devem ser adotados algoritmos de **Deep Learning (DL)** que recorrem a uma espécie de rede neuronal artificial (**Artificial Neural Network – ANN**). Esta é inspirada no cérebro humano e nas ligações que lá ocorrem no decorrer do processo de aprendizagem [25].

#### 2.2.2.1 ANN

Uma rede neuronal artificial (Artificial Neural Network – ANN) é um modelo que é desenvolvido tendo como base o cérebro humano [26].

O cérebro é composto por diversos neurónios. Estes são compostos pelo **corpo celular**, local onde está presente o **núcleo celular**, bem como por múltiplos **dendritos**, que representam um prolongamento do neurónio responsável por garantir a receção dos estímulos nervosos e transmiti-los ao corpo celular. Outro elemento importante numa célula nervosa é o prolongamento designado de **Axónio**, que assegura a condução do impulso nervoso até outras células ou órgãos. A Figura 3 apresenta a estrutura de um neurónio [27].

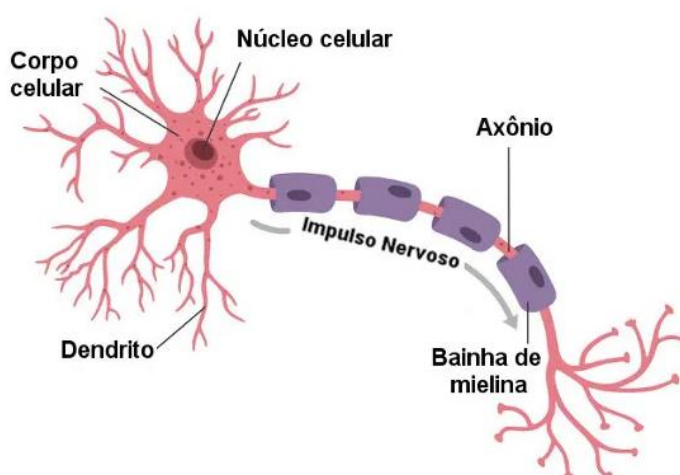


Figura 3 - Partes constituintes de um neurónio [28]

A ANN é composta por diversos processadores interligados, os quais são denominados de **neurónios artificiais**, fazendo um paralelismo aos neurónios do cérebro humano [26].

Cada neurónio recebe um determinado número de sinais de entrada, através das suas ligações, produzindo um único sinal de saída. Posteriormente, este é transmitido à camada de neurónios seguinte, cujas células nervosas se encontram ligadas às anteriores. Deste modo, os sinais de entrada irão produzir novos sinais de saída nos restantes neurónios até ser alcançada a camada de output onde uma decisão/ação é tomada, mediante os sinais que forem produzidos e comunicados ao longo do sistema [26].

Na Figura 4, é possível constatar as relações que poderão suceder num sistema neuronal artificial, sendo que os círculos representam cada neurónio artificial que o sistema possui.

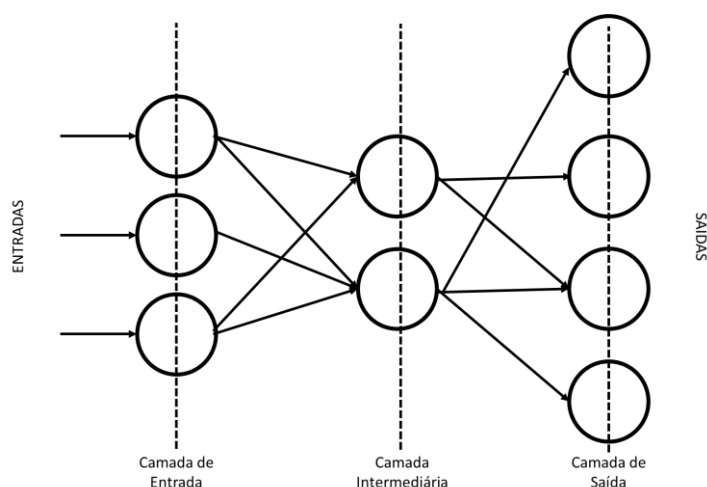


Figura 4 – Possível desenho de uma ANN [26].

#### 2.2.2.2 CNN

As redes neuronais convulsionais (**Convolutional neural network – CNN**) correspondem a um tipo de ANN utilizado principalmente para processar e classificar os dados no formato de imagens 2D. No entanto, pode, de igual forma, ser utilizado para áudio, vídeo e texto. As redes neurais convulsionais na primeira camada têm o objetivo de detetar características específicas nos dados. Na segunda camada, todas as características detetadas na camada anterior são unidas de forma a gerar características maiores, a fim de detetar a totalidade da amostra de dados [29].

#### 2.2.2.3 LSTM

Um tipo de rede neuronal artificial é o **Long Short Term (LSTM)**, que foi definida como sendo capaz de resolver o problema de expansão e fuga dos ANN [30] [31].

Uma LSTM possui quatro redes neuronais, ligadas entre si, com diferentes blocos de memória chamados de células [32]. Atente-se na Figura 5 uma possível estrutura de uma LSTM.

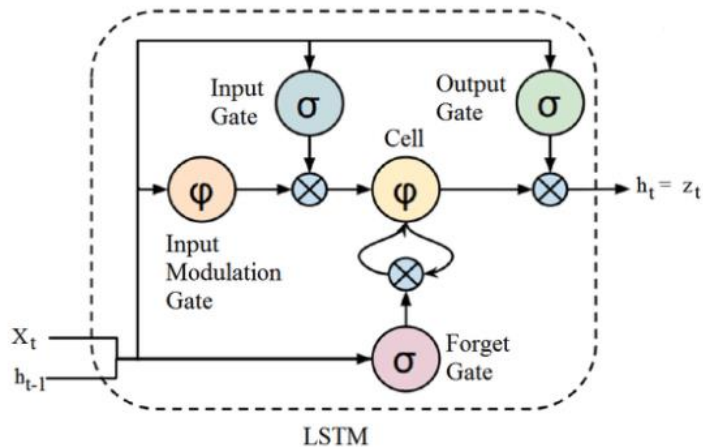


Figura 5 – Arquitetura de uma LSTM [33]

As portas de entrada (**Input Gate** e **Input Modulation Gate**) têm como intuito a adição de informações úteis ao estado da célula que poderá produzir uma ação. Quando estas informações perdem a sua utilidade, são esquecidas e levadas para a **Forget Gate**, de forma a serem limpas e esquecidas. Desta forma, o vasto número de dados e a sua complexidade é reduzido [33].

Na célula, são, também, ativados estados mediante as informações que lhes sejam fornecidas. Eles, por seu turno, irão produzir ações na porta de saída, local que possui a missão de extrair a informação útil do estado da célula [33].

#### 2.2.2.4 RandomForestClassifier

Uma **RandomForestClassifier** é um tipo de árvore de decisão. As árvores de decisão utilizam uma variedade de algoritmos ou diferentes configurações de um mesmo algoritmo. para decidir em quantos nós estas se subdividem. Os nós têm como intuito direcionar a decisão mediante os dados de entrada [34].

Através desta decisão, é realizada uma classificação dos dados fornecidos inicialmente, sendo que, no caso do **RandomForestClassifier**, são criadas múltiplas árvores de decisão para que a distribuição seja idêntica e, em cada árvore, seja apenas possível obter um valor para o input inicial [35]. Na Figura 6, consegue-se observar um exemplo de um fluxograma de uma **RandomForestClassifier**.

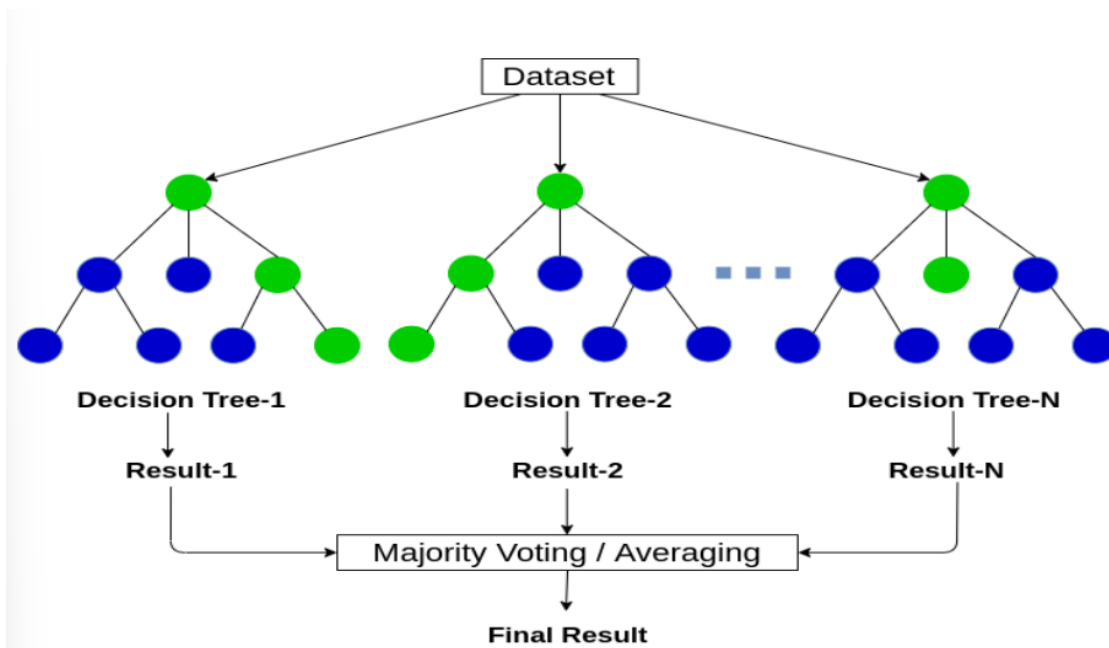


Figura 6 - Fluxograma de uma *RandomForestClassifier* [36]

## 2.3 Trabalhos Relacionados

Na presente secção, é possível encontrar trabalhos relacionados com o atual projeto onde irá ser detalhado a análise de cada um dos projetos, bem como os objetivos iniciais e os concretizados em cada um dos projetos.

### 2.3.1 Modelos de apostas usando IA

Os modelos de apostas têm vindo a ganhar cada vez mais interesse nos dias que correm, isto devido em grande parte ao número crescente de apostadores e de entusiastas do desporto.

Neste projeto, o autor teve como principal objetivo a utilização da Inteligência Artificial para auxiliar os apostadores a preverem os resultados de eventos desportivos [26].

Para isso, ele explicou, em traços gerais, de que forma é que as redes neuronais artificiais (ANN) poderiam ser usadas para prever os resultados dos jogos de futebol. Adicionalmente, foi apresentada uma técnica designada por **SVM**, **Support Vector Machine**, com o propósito de antever o resultado de um jogo de futebol [26].

Por último, foi explanada a cadeia de **Markov**, sendo apresentada como uma opção viável para a aplicação na previsão dos resultados desportivos. Isto porque, conforme explicado, a cadeia presume o próximo estado de uma amostra, neste caso, o próximo resultado do jogo, mediante os estados anteriores [26].

Em suma, este estudo viabilizou a compreensão da parte teórica inerente à implementação de um projeto de previsão de resultados. Não obstante, não foi possível observar a concretização empírica dos fundamentos teóricos.

### **2.3.2 Previsão de partidas de futebol**

Nesta tese, Jorn Boksem procurou aumentar a precisão na previsão de um modelo de IA em partidas de futebol.

Para isso, foi realizado um estudo intensivo de outros trabalhos similares, analisando-se os resultados dos mesmos com valores percentuais para a precisão/sucesso de cada tipo de algoritmo/método utilizado [37].

Com efeito, ao longo do projeto, ele foi testando vários tipos de *features* que pudessem influenciar a previsão do modelo. Desta forma, ele conseguiu estabelecer comparações que lhe permitiram descobrir os inputs que dariam para otimizar a aprendizagem do modelo.

Para finalizar, os dados deste trabalho foram bastante reduzidos, quer a nível de volume, quer a nível da sua diversidade, tendo-se debruçado apenas sobre os dados da primeira liga inglesa, o que, no caso do mercado de apostas português, representa uma pequena fatia do mercado de apostas desportivas.



## 3 Análise e desenho da solução

Neste capítulo, é descrito o Domínio do Problema, sendo abordado o mundo das apostas desportivas. De igual modo, é apresentada a API que foi utilizada para obter os resultados dos jogos, bem como as estatísticas de todas as equipas e campeonatos existentes em todo o mundo.

No que concerne ao Domínio do Problema, são referidos os recursos existentes para a implementação do projeto.

Após a descrição do problema, são detalhados os requisitos necessários para o desenvolvimento do projeto.

Para finalizar, é exposto o desenho da solução, desde a elaboração da arquitetura dos dados inicialmente extraídos até à arquitetura da implementação.

### 3.1 O mundo das apostas desportivas

Desde os tempos do Império Romano, em que as classes altas apostavam em jogos de tabuleiros, dados e corridas de cavalos, as apostas se têm vindo a tornar num hábito humano, possibilitando uma fácil e rápida obtenção de dinheiro [38].

Atualmente, as casas de apostas movimentam, em média, 340 milhões de euros por jornada, valor referente, apenas, às apostas dos seus clientes no mercado português [39].

Ao número supracitado, deve-se somar o número de grupos de prognósticos que vendem os seus serviços a milhares, senão milhões, de pessoas que anseiam obter dinheiro sem qualquer tipo de esforço [3]. Todavia, não faz parte do conhecimento das mesmas que a grande maioria destes grupos são financiados pelas casas de apostas com o propósito de incentivar hipotéticos clientes a arriscar o seu dinheiro [3].

Por outro lado, as casas de apostas diminuem o seu lucro de cada vez que as suas projeções dos resultados de um determinado jogo não correspondem à realidade. Por conseguinte, estas precisam de realizar ajustes ao longo do tempo de modo a conseguirem aumentar o potencial lucro num determinado evento [40].

Posto isto, o presente projeto visa expor as equipas vencedoras em cada jogo de futebol.

## 3.2 Obtenção dos resultados e dados estatísticos das partidas de futebol

Neste contexto, a solução encontrada consistiu em recorrer a uma **Application Programming Interface**<sup>1</sup> - API. Esta fornece, diretamente, a lista de equipas e de campeonatos existentes, permitindo, a cada minuto, extrair os resultados de cada jogo, bem como os dados históricos relativos às estatísticas das diversas partidas, jogadores e treinadores.

Após a observação de várias APIs, optou-se pela seleção da API fornecida pelo site “www.api-football.com”. Este apresentava maior credibilidade e possuía um período de teste experimental, o que permitiu realizar alguns testes aos dados que esta API fornecia.

Ademais, a API em questão apresenta a vantagem de incluir todo o histórico de resultados desde o ano de 2010 até à atualidade. Este fator assume uma extrema importância para a execução da análise estatística dos resultados passados nos mais diversos campeonatos, com especial foco nas 7 maiores ligas europeias.

Na Figura 7, pode-se visualizar os diferentes tipos de pedidos que podem ser executados através da API. Os referidos pedidos são realizados através de API *Endpoints*<sup>2</sup>, que nos restituirá as informações desejadas em formato de JSON<sup>3</sup>.

---

<sup>1</sup> Traduzindo do Inglês “Interface de Programação de Aplicações”, que é construída com a missão de integrar sistemas, possibilitando a segurança dos dados e a facilidade na troca de informações com diferentes linguagens de programação [44]

<sup>2</sup> São usados para uma API comunicar com outros sistemas, normalmente, através de URL, partilhando apenas os dados pretendidos pela API e, desta forma, aumentando a proteção dos restantes dados [45]

<sup>3</sup> JSON é um formato leve e simples na realização de troca de informação/dados entre sistemas [46]

The image shows a two-part interface. The top part is a menu with a dropdown arrow on the right. It contains the following items:

- Players
- GET Seasons
- GET Players (highlighted)
- GET Squads
- GET Top Scorers
- GET Top Assists
- GET Top Yellow Cards
- GET Top Red Cards

The bottom part is a dark-themed panel with a 'Request samples' section. It features a dropdown menu with 'Python' selected, and buttons for 'Php', 'Node', 'JavaScript', 'Curl', 'Ruby', and 'Use Cases'. Below the menu is a code editor with the following Python code:

```
import http.client
conn = http.client.HTTPSConnection("v3.football.api-sports.io")
headers = {
    'x-rapidapi-host': "v3.football.api-sports.io",
    'x-rapidapi-key': "XXXXXXXXXXXXXXXXXXXXXXXX"
}
conn.request("GET", "/players?id=276&season=2019", headers=headers)
res = conn.getresponse()
data = res.read()
print(data.decode("utf-8"))
```

A 'Copy' button is located at the bottom right of the code editor.

Figura 7 - Exemplo de pedido em *Python* à API de resultados

Esta API retorna as respostas em formato de JSON, com as informações determinadas para cada *endpoint*, podendo as mesmas variar mediante o sucesso da chamada do *endpoint*.

A Figura 8 é ilustrativa de um exemplo de resultado retornado pelo *endpoint*, relativo aos dados estatísticos de um jogador, no formato JSON.

```
Copy Expand all Collapse all
{
  "get": "players",
+ "parameters": { ... },
  "errors": [ ],
  "results": 1,
- "paging": {
    "current": 1,
    "total": 1
  },
- "response": [
  - {
    - "player": {
      "id": 276,
      "name": "Neymar",
      "firstname": "Neymar",
      "lastname": "da Silva Santos Júnior",
      "age": 28,
+ "birth": { ... },
      "nationality": "Brazil",
      "height": "175 cm",
      "weight": "68 kg",
      "injured": false,
      "photo": "https://media.api-sports.io/football/players/276.¿
    },
+ "statistics": [ ... ]
  }
]
}
```

Figura 8 - Exemplo de respostas no formato JSON

### 3.3 Desenho da solução

Neste tópico, é apresentada a arquitetura geral da solução, assim como a análise e arquitetura de cada uma das fases de implementação.

#### 3.3.1 Arquitetura dos dados inicialmente extraídos

Na Figura 9, é possível observar a arquitetura existente na API de resultados.

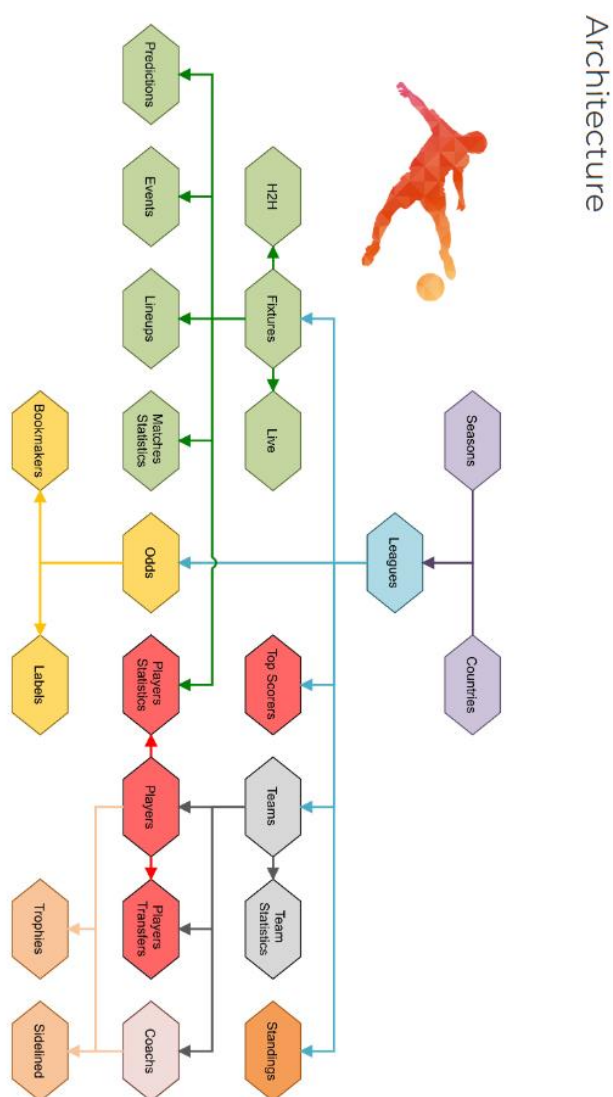


Figura 9 - Arquitetura da API de resultados<sup>4</sup>

<sup>4</sup> Imagem retirada da documentação fornecida pela API de resultados "www.api-football.com"

Assim, verifica-se que existem 4 blocos principais de dados que se vão relacionando entre eles:

- **Jogos de Futebol:** Neste grande bloco, encontram-se os resultados anteriormente obtidos entre as 2 equipas em questão (H2H), os jogadores que participaram ou irão participar, no caso de ser um jogo no futuro (*Lineups*), e os dados estatísticos do jogo (*Matches Statistics*). De igual modo, os eventos que ocorreram durante o jogo (*Events*), bem como as previsões que o algoritmo da API efetua para o mesmo (*Predictions*), são informações que aqui se conseguem obter. Simultaneamente, é possível encontrar os dados relativos ao jogo, tais como o estádio onde irá ser realizado e o ID relacionado ao jogo, o qual servirá para unir com outros blocos (*Fixtures*). Por último, também é possível obter todas as informações em tempo real, aquando da ocorrência dos jogos (*Live*). Este bloco possui relação com os demais.
- **Equipas:** Aqui, tem-se a possibilidade de tomar conhecimento de toda a informação relacionada com as equipas (*Team*), como, por exemplo, os dados estatísticos da mesma (*Team Statistics*), os jogadores (*Players*), as suas estatísticas (*Players Statistics*) e as transferências (*Players Transfers*). De igual forma, é possível observar os treinadores (*Coachs*) das equipas e todo o seu histórico (*Sidelined*) e, ainda, os trofeus dos jogadores e treinadores da equipa (*Trophies*). Este bloco relaciona-se com o dos Jogos de Futebol e das Ligas.
- **Odds:** Permitindo a obtenção das *odds* (*Odds*) existentes para um determinado jogo nos diferentes mercados (*Labels*) e nas diferentes casas de apostas desportivas (*Bookmakers*), este bloco apresenta uma relação direta com o bloco dos Jogos de Futebol e das Ligas.
- **Ligas:** Lugar onde se situam as informações das ligas (*League*), das épocas (*Season*) e dos países (*Countries*), este bloco relaciona-se com os mencionados anteriormente.

### 3.3.2 Arquitetura da implementação

Numa fase inicial, foram extraídos os dados da fonte de dados, neste caso, da API.

Posteriormente, procedeu-se a uma análise e exploração dos dados obtidos para aprofundar o conhecimento acerca dos mesmos.

Logo de seguida, foi implementado um processo de transformação dos dados, através do procedimento de limpeza de dados inúteis e recorrendo a uma uniformização dos mesmos, com o intuito de facilitar a aprendizagem por parte dos modelos.

Cumulativamente, neste passo, foram criadas novas *features* que se revelaram úteis para o modelo em questão aprender.

Igualmente, foi realizada a filtragem dos dados que seguiriam para teste e para treino, pois, no caso dos jogos de futebol, a sequência dos resultados tem extrema importância no cálculo dos resultados finais.

Por último, foi construído o modelo de IA e respetiva validação.

Na Figura 10, é possível acompanhar o processo em cima detalhado de modo mais elucidativo.

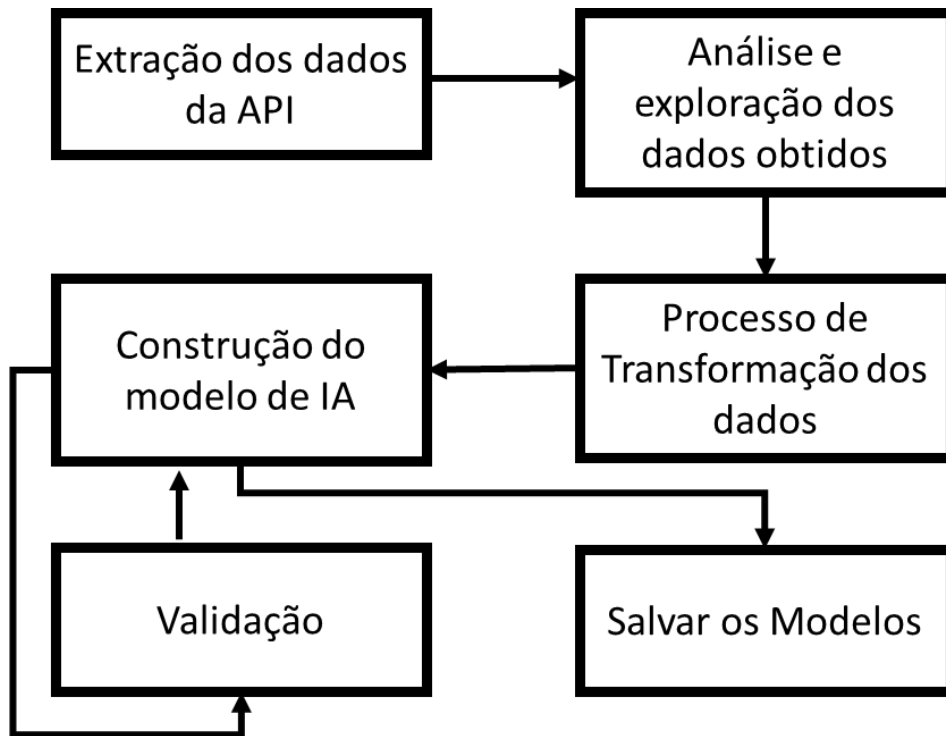


Figura 10 - Desenho da arquitetura implementada

Após esta primeira abordagem, foi possível otimizar, gradualmente, o modelo de IA, à medida que novos dados surgiam, tendo-se, para esse efeito, utilizado a arquitetura presente na Figura 11.

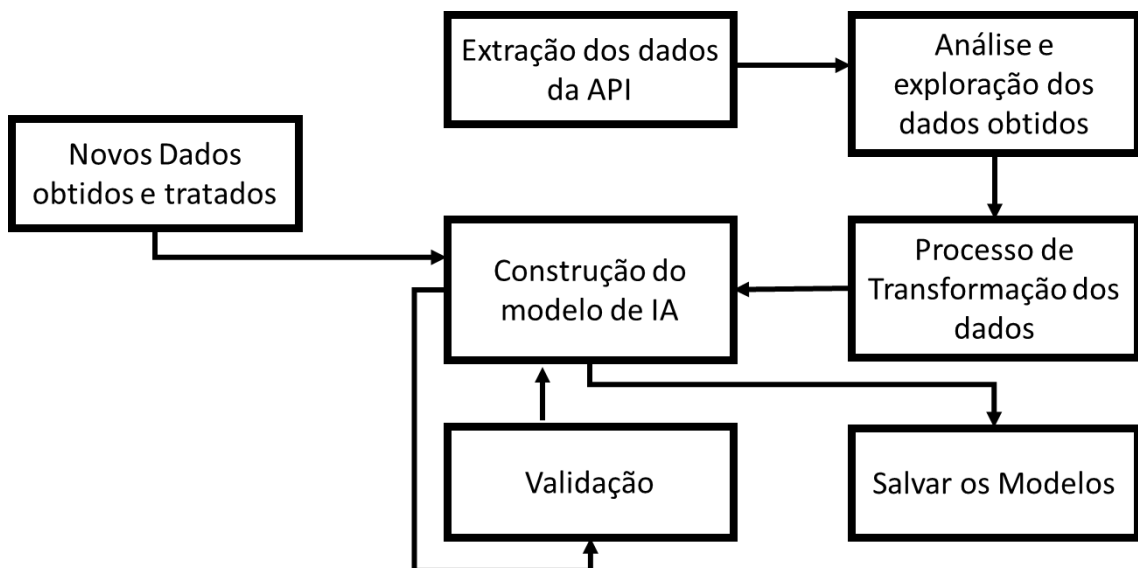


Figura 11 - Desenho da arquitetura implementada com novos dados obtidos

## 3.4 Requisitos

Nesta parte, são apresentados os utilizadores e respetivas necessidades, seguindo-se a descrição dos requisitos funcionais e não funcionais deste projeto.

O glossário<sup>5</sup> referente à engenharia de requisitos encontra-se presente no Anexo 1 Glossário dos Requisitos.

### 3.4.1 Utilizadores e as suas necessidades

A principal funcionalidade deste projeto corresponde à determinação da previsão correta do resultado de um jogo de futebol. Nesse sentido, os utilizadores interessados nesta previsão são os apostadores e os gestores de *odds* das casas de apostas.

Assim sendo, surgiram os seguintes casos de uso<sup>6</sup> relacionados com os apostadores:

- Eu, apostador, quero consultar o sucesso dos prognósticos dos resultados por tipo de liga.
- Eu, apostador, quero consultar o sucesso dos prognósticos dos resultados por tipo de país.
- Eu, apostador, quero consultar o sucesso dos prognósticos dos resultados por tipo de competição.
- Eu, apostador, quero consultar o sucesso dos prognósticos dos resultados por tipo de aposta, isto é, se o prognóstico é “Live” ou “Pré-Live”.
- Eu, apostador, quero consultar o sucesso dos prognósticos dos resultados por ano, mês, semana e dia.
- Eu, apostador, quero consultar os prognósticos futuros que apresentam maior probabilidade de sucesso.
- Eu, apostador, quero visualizar um relatório construído dinamicamente com os campos e filtros escolhidos por mim.
- Eu, gestor, quero exportar prognósticos, que possa escolher, para um ficheiro.

Relativamente aos gestores de *odds* das casas de apostas, é possível enumerar os seguintes casos de uso:

- Eu, gestor de *odds*, quero consultar o sucesso dos prognósticos dos resultados por tipo de liga.
- Eu, gestor de *odds*, quero consultar o sucesso dos prognósticos dos resultados por tipo de país.

---

<sup>5</sup> O glossário é visto como um dicionário, com intuito de definir a terminologia de alguns termos ou expressões [47].

<sup>6</sup> História concisa de um requisito do utilizador sob o ponto de vista desse mesmo utilizador [47][48][49].

- Eu, gestor de *odds*, quero consultar o sucesso dos prognósticos dos resultados por tipo de competição.
- Eu, gestor de *odds*, quero visualizar um relatório construído dinamicamente com os campos e filtros escolhidos por mim.
- Eu, gestor de *odds*, quero exportar prognósticos, que possa escolher, para um ficheiro.
- Eu, gestor de *odds*, quero consultar em tempo real todos os prognósticos, através de uma API ou outro processo automatizado de integração em tempo real.
- Eu, gestor de *odds*, pretendo saber todo o tipo de previsões e a probabilidade associada de estas previsões estarem corretas.

### **3.4.2 Requisitos Funcionais**

Os requisitos funcionais servem para definir os serviços que o sistema deve suportar, bem como os seus comportamentos.

Desta forma, o projeto contém os seguintes requisitos funcionais:

- O sistema deve atualizar o conteúdo do modelo diariamente;
- O sistema deve controlar os dados a que cada utilizador terá acesso;
- O sistema deve permitir a comparação de equipas, recorrendo a dados estatísticos;
- O sistema deve permitir a comparação de prognósticos por país, campeonato, data, entre outros filtros;
- O sistema deve apresentar os melhores prognósticos futuros, analisando estatisticamente dados anteriores.

### **3.4.3 Requisitos Não Funcionais**

Os requisitos não funcionais correspondem às características que o sistema deve incluir no que concerne à usabilidade, confiabilidade, portabilidade, funcionalidade, eficiência e manutenibilidade. Fazem, também, parte destes requisitos as restrições de design, enquadrando-se, desta forma, no modelo ISO 912615 [41].

De seguida, são detalhados alguns dos requisitos não funcionais:

- A API de resultados deve apresentar dados reais sem estes serem adulterados;
- A API de resultados deve conter o nome das equipas, a data do jogo e todos os dados estatísticos relativamente a um jogo de futebol, equipa, jogador e treinador;
- O sistema deve ser intuitivo;
- O sistema deve ser de fácil compreensão;
- O sistema deve possuir filtros de forma automática filtrando pelo mês atual, aquando da análise do sucesso dos sites de prognósticos;
- O sistema deve permitir a introdução de filtros, na pesquisa de um prognóstico específico, de forma manual.



## 4 Implementação da solução

Neste capítulo, é detalhado todo o processo de implementação do projeto, analisando-se, individualmente, cada fase do mesmo, tanto ao nível da arquitetura como da sua codificação.

Inicialmente, são apresentadas as tecnologias e linguagens adotadas e, numa fase mais avançada, é descrita toda a implementação da solução.

Por último, procedeu-se a uma avaliação dos resultados obtidos.

### 4.1 Tecnologias Adotadas

Por forma a viabilizar a execução deste projeto, recorreu-se ao *Google Colab*, um serviço de nuvem gratuito disponibilizado pela Google, com o intuito de promover a pesquisa do ML e da IA.

Efetivamente, esta ferramenta, acessível através de um site, permite uma fácil iniciação do projeto, visto não necessitar de qualquer tipo de configuração, bastando, assim, programar de modo usual.

Não obstante, poder-se-ia, igualmente, ter utilizado o *Jupyter Notebook* para desenvolver o projeto de ML. No entanto, no caso deste trabalho, o *Google Colab* revelou-se mais vantajoso, apresentando um número acrescido de benefícios.

A Tabela 2 estabelece uma pequena comparação entre o *Jupyter Notebook* e o *Google Colab*.

Tabela 2 - *Google Colab vs Jupyter Notebook*

Característica	<i>Jupyter Notebook</i>	<i>Google Colab</i>
Baseado em Cloud	Não	Sim
Sincronização de Ficheiros	Não	Sim
Partilha de Ficheiros	Não	Sim
Instalação de Bibliotecas	Sim	Não

Através da análise desta tabela, é possível inferir que o *Google Colab* se revela mais satisfatório do que o *Jupyter Notebook* ao nível da partilha e sincronização de ficheiros, utilizando o *Google Drive* para efetuar essas mesmas tarefas.

Adicionalmente, apresenta como vantagem o facto de não necessitar da instalação de qualquer biblioteca, sendo suficiente proceder à sua importação, de forma a permitir a utilização das mesmas ao longo do código.

Posto isto, o *Jupyter Notebook* deve ser usado quando se trabalha com ficheiros/dados sensíveis, que possuam determinadas regras de proteção e segurança, e que, por esse motivo, não possam ser guardados nem usados em ambiente de *Cloud*.

## 4.2 Linguagens Adotadas

Para o desenvolvimento deste projeto, recorreu-se à linguagem de programação *Python*.

De facto, esta linguagem apresentou-se como sendo a solução mais viável para a incrementação deste projeto, caracterizando-se por uma fácil aprendizagem, pautada por uma sintaxe bastante intuitiva, que permitiu agilizar o processo de programar. Além disso, o facto de ser uma *Open Source* e possuir uma grande comunidade de programadores representou um fator preponderante, visto que, para qualquer questão hipotética associada ao desenvolvimento de um determinado projeto, permite consultar a comunidade e esclarecer as dúvidas através da partilha de ideias e transferência de conhecimento entre os programadores.

Por fim, a existência de um vasto leque de bibliotecas na área da ciência dos dados facilitou o desenvolvimento deste trabalho com algoritmos prontos a serem usados com simples linhas de código.

## 4.3 Bibliotecas utilizadas

Foram utilizadas as seguintes bibliotecas no desenvolvimento deste trabalho:

- *Http.client*:
  - Biblioteca usada para a realização de pedidos http a uma api;
- *Json*:
  - Dispõe de vários métodos e funções para lidar com dados no formato de json;
- *Pandas*:
  - Utilizada para a manipulação de tabelas e dados de forma rápida e intuitiva;
- *Sklearn*:
  - Biblioteca com diversos algoritmos disponíveis de classificação, regressão, *clustering*, pré-processamento, entre outros;
- *Numpy*:
  - Suporta o processamento de grandes quantidades de dados, disponibilizando uma grande coleção de funções matemáticas de alto nível.

## 4.4 Descrição da implementação

Para a implementação deste projeto, inicialmente, recuperaram-se todos os dados históricos existentes na fonte de dados da API.

Depois, esses dados foram guardados, em formato json, ao longo das diversas pastas da drive. A Figura 12 demonstra a organização das pastas que contêm o histórico guardado.

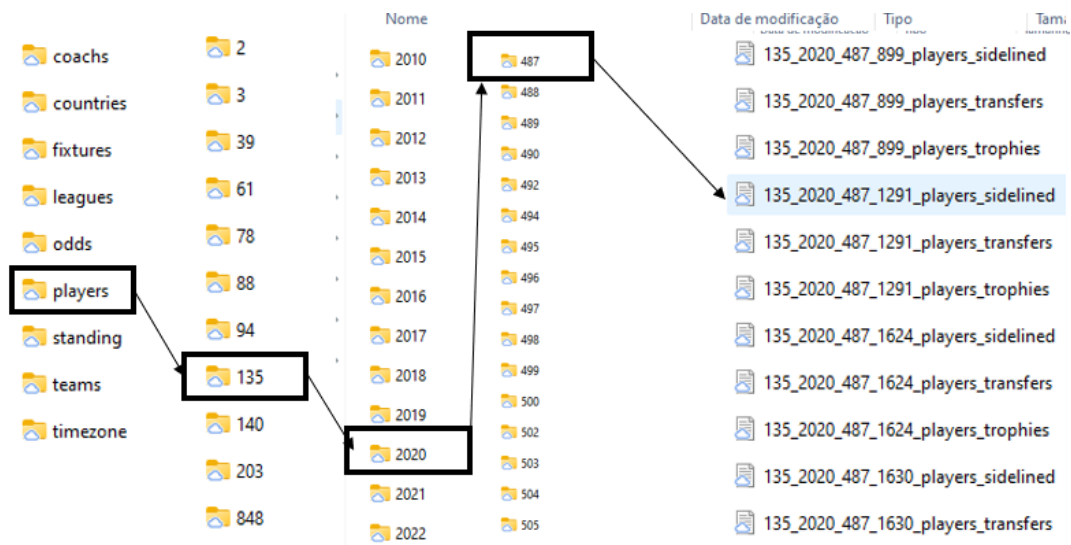


Figura 12 – Organização dos ficheiros

Inicialmente, os dados foram guardados em formato json. A Figura 13 - Exemplo de json ilustra um exemplo.

```
{'winner': {'id': 157, 'name': 'Bayern Munich', 'comment': 'Win or draw'},
'win_or_draw': True,
'under_over': None,
'goals': {'home': '-3.5', 'away': '-2.5'},
'advice': 'Double chance : Bayern Munich or draw',
'percent': {'home': '45%', 'draw': '45%', 'away': '10%'}}
```

Figura 13 - Exemplo de json que a API retorna

Após a leitura do *json*, este foi transformado num *dataframe*, recorrendo a uma função da biblioteca pandas que normaliza uma variável *json* numa tabela. A Figura 14 é representativa da transformação em questão.



Figura 14 – Normalização de uma variável json num dataframe

Logo de seguida, foram realizadas algumas transformações de colunas, tal como a que diz respeito às colunas de texto, que passaram a colunas numéricas, classificando as diversas opções. Desta forma, o modelo foi capaz de ler todas as colunas numéricas de um modo mais rápido e intuitivo.

Posteriormente, foram filtradas as colunas relevantes tais como:

- Para a equipa visitada:
  - Vitórias nos últimos 5 jogos em casa;
  - Derrotas nos últimos 5 jogos em casa;
  - Empates nos últimos 5 jogos em casa;
  - Golos marcados nos últimos 5 jogos em casa;
  - Golos sofridos nos últimos 5 jogos em casa;
  - Média de golos sofridos nos jogos em casa;
  - Média de golos marcados nos jogos em casa;
  - Percentagem de vitórias em casa;
  - Percentagem de derrotas em casa;
  - Percentagem de empates em casa;
  - Classificação na liga;
  - Número de vitórias frente à equipa visitante;
  - Número de derrotas frente à equipa visitante;
  - Número de empates frente à equipa visitante;
  - Número de golos marcados frente à equipa visitante;
  - Número de golos sofridos frente à equipa visitante;
  - Média de golos marcados frente à equipa visitante;
  - Média de golos sofridos frente à equipa visitante;
  - Vitórias nos últimos 5 jogos em casa, frente a frente, com a equipa visitante;
  - Derrotas nos últimos 5 jogos em casa, frente a frente, com a equipa visitante;
  - Empates nos últimos 5 jogos em casa, frente a frente, com a equipa visitante;
  - Média de golos marcados nos últimos 5 jogos em casa, frente a frente, com a equipa visitante;
  - Média de golos sofridos nos últimos 5 jogos em casa, frente a frente, com a equipa visitante;
  
- Para a equipa visitante:
  - Vitórias nos últimos 5 jogos em fora;
  - Derrotas nos últimos 5 jogos em fora;
  - Empates nos últimos 5 jogos em fora;
  - Golos marcados nos últimos 5 jogos em fora;
  - Golos sofridos nos últimos 5 jogos em fora;
  - Média de golos sofridos nos jogos em fora;
  - Média de golos marcados nos jogos em fora;
  - Percentagem de vitórias em fora;
  - Percentagem de derrotas em fora;
  - Percentagem de empates em fora;
  - Classificação na liga;
  - Número de vitórias frente à equipa visitada;
  - Número de derrotas frente à equipa visitada;

- Número de empates frente à equipa visitada;
- Número de golos marcados frente à equipa visitada;
- Número de golos sofridos frente à equipa visitada;
- Média de golos marcados frente à equipa visitada;
- Média de golos sofridos frente à equipa visitada;
- Vitórias nos últimos 5 jogos em casa, frente a frente, com a equipa visitada;
- Derrotas nos últimos 5 jogos em casa, frente a frente, com a equipa visitada;
- Empates nos últimos 5 jogos em casa, frente a frente, com a equipa visitada;
- Média de golos marcados nos últimos 5 jogos em casa, frente a frente, com a equipa visitada;
- Média de golos sofridos nos últimos 5 jogos em casa, frente a frente, com a equipa visitada;

Estas colunas serviram para ensinar e testar o modelo, como se pode visualizar na Figura 15.

```
fixtureClean = fixturePrediction[['fixtureId', 'league.id', 'league.name', 'teams.home.id', 'teams.home.name', 'teams.away.id', 'teams.away.name', 'predictions.winner',
'predictions.goals.home', 'predictions.goals.away', 'predictions.percent.home', 'predictions.percent.draw', 'predictions.percent.away', 'predictions.advice',
'teams.home.last_5.form', 'teams.home.last_5.att', 'teams.home.last_5.def', 'teams.home.last_5.goals.for.average', 'teams.home.last_5.goals.against.average',
'teams.away.last_5.form', 'teams.away.last_5.att', 'teams.away.last_5.def', 'teams.away.last_5.goals.for.average', 'teams.away.last_5.goals.against.average',
'teams.home.league.form', 'teams.home.league.fixtures.wins.home', 'teams.home.league.fixtures.wins.away',
'teams.home.league.fixtures.draws.home', 'teams.home.league.fixtures.draws.away', 'teams.home.league.fixtures.loses.home', 'teams.home.league.fixtures.loses.away',
'teams.home.league.goals.for.average.home', 'teams.home.league.goals.for.average.away',
'teams.away.league.form', 'teams.away.league.fixtures.wins.home', 'teams.away.league.fixtures.wins.away',
'teams.away.league.fixtures.draws.home', 'teams.away.league.fixtures.draws.away', 'teams.away.league.fixtures.loses.home', 'teams.away.league.fixtures.loses.away',
'teams.away.league.goals.for.average.home', 'teams.away.league.goals.for.average.away'
]]
```

Figura 15 – Filtragem das colunas a serem analisadas

O desenho da implementação pode ser traduzido através da Figura 16.

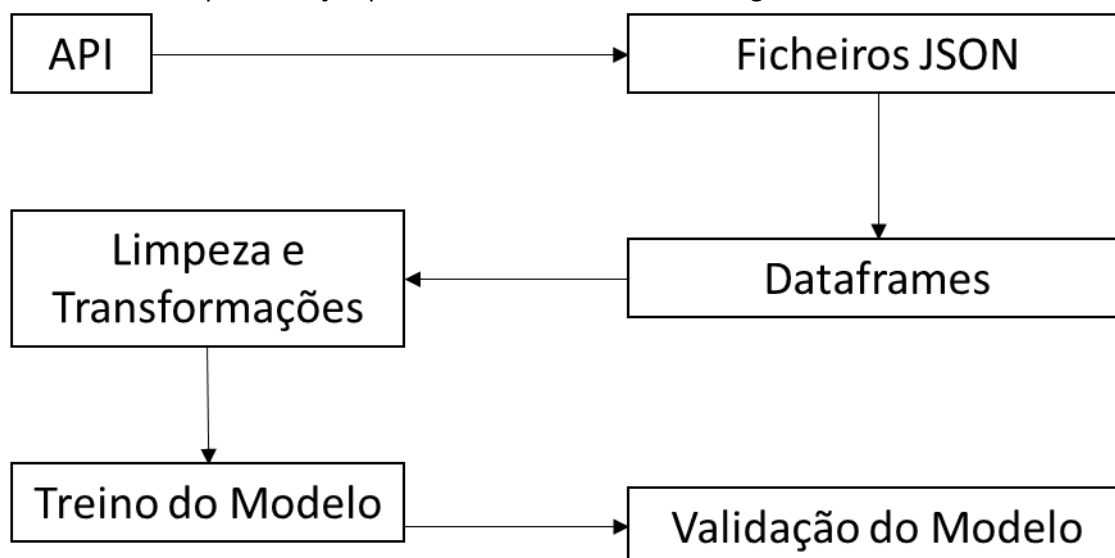


Figura 16 – Implementação da Solução

Ao observar a Figura 16, é possível constatar que os últimos dois passos realizados na implementação da solução do projeto foram o treino do modelo e a respetiva validação.

Para isso, num estágio inicial, foi usado o *RandomForestClassifier* para classificar se uma equipa num determinado jogo iria perder, empatar ou ganhar.

## 4.5 Avaliação da Solução

Com o propósito de avaliar a solução, foram realizados vários testes utilizando o *RandomForestClassifier*.

Neste sentido, a seguinte tabela mostra a percentagem de sucesso, na determinação de um jogo terminar empatado e/ou da equipa vencedora, mediante o número de dados que são fornecidos para teste, dados estes descritos na coluna “Informação”. Foi utilizada a liga dos campeões como liga de análise.

### 4.5.1 Métricas Usadas

Para avaliar a solução, recorreu-se a duas métricas:

- *Accuracy*;
- *Precision*.

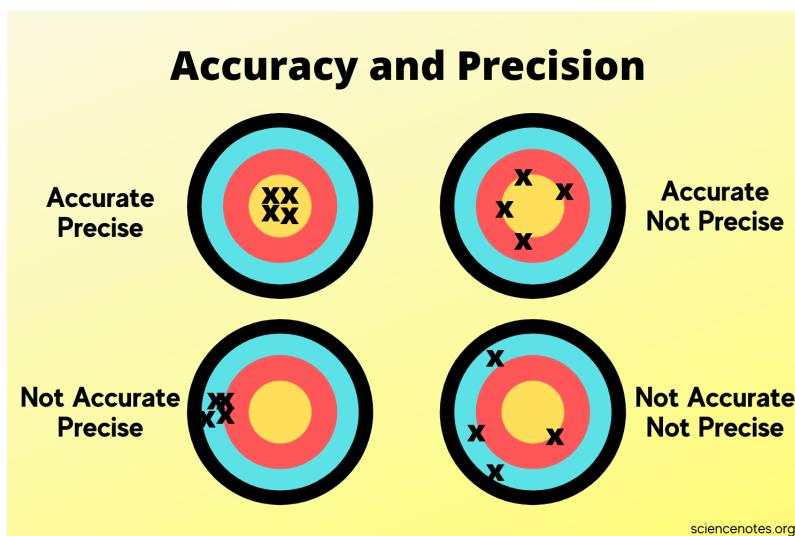


Figura 17 – *Accuracy and Precision* [42]

Através da Figura 17, é possível compreender a diferença entre estes dois conceitos, a *Accuracy* e a *Precision*.

De facto, a *Accuracy*, que corresponde à exatidão, representa a capacidade de atingir o alvo central, ou, por outras palavras, o quão próximos determinados valores obtidos se encontram de um certo valor de referência.

Por sua vez, a *Precision*, isto é, a precisão, relaciona-se com a proximidade dos dados, mesmo não sendo o objetivo principal. Ou seja, é referente à proximidade entre os diversos valores obtidos.

Posto isto, e para viabilizar uma análise mais detalhada da solução, será atribuída uma maior importância à *Accuracy*, visto que representará a probabilidade de atingir de forma certa o nosso “*target*”, ou seja, que equipa irá vencer o encontro.

#### 4.5.2 Liga dos Campeões

Tabela 3 – Resultados obtidos na Liga dos Campeões com 3 possíveis resultados de *output* usando RandomForestClassifier

Informação	Accuracy Score	Real vs Previsão			
<b>Dados de Teste de 2011 até 2021</b> 3 opções de resultado 0 → Vitória da equipa da casa 1 → Empate 2 → Vitória da equipa de fora	0.50467	predicted	0	1	2
		actual			
		0	0	31	12
		1	2	84	17
		2	6	40	26
<b>Dados de Teste de 2011 até 2020</b> 3 opções de resultado 0 → Vitória da equipa da casa 1 → Empate 2 → Vitória da equipa de fora	0.5076	predicted	0	1	2
		actual			
		0	0	59	19
		1	4	153	24
		2	5	84	48
<b>Dados de Teste de 2011 até 2019</b> 3 opções de resultado 0 → Vitória da equipa da casa 1 → Empate 2 → Vitória da equipa de fora	0.4950	predicted	0	1	2
		actual			
		0	2	99	25
		1	4	225	43
		2	6	129	73

Assim, é possível inferir que a percentagem de *accuracy* é muito baixa em todos os estudos, apresentando uma pequena melhoria quando o resultado de testes e de treino é repartido pelo ano de 2020.

Após estes resultados, foi realizada uma simplificação no processo de classificação e, ao invés do algoritmo prever se ia ficar 0,1 ou 2, previu se ia ficar 0 no caso da equipa da casa ser a

vencedora e 1 no caso de ser empate ou a equipa de fora vencer. Os resultados foram bastantes melhores como podemos observar pela seguinte tabela.

Tabela 4 - Resultados obtidos na Liga dos Campeões com 2 possíveis resultados de *output* usando RandomForestClassifier

Informação	Accuracy Score	Precision Score	Real vs Previsão		
Dados de Teste de 2011 até 2021 2 opções de resultado	0.6376	0.62	predicted	0	1
			actual		
			0	77	38
			1	41	62
Dados de Teste de 2011 até 2020 2 opções de resultado	0.6237	0.5851	predicted	0	1
			actual		
			0	137	78
			1	71	110
Dados de Teste de 2011 até 2019 2 opções de resultado	0.6403	0.5925	predicted	0	1
			actual		
			0	215	119
			1	99	173

É possível observar que a amostra dos dados com a *Accuracy* mais elevada foi o ano de 2019. Já em relação à percentagem de *Precision* mais elevada, esta ocorreu com a divisão dos dados no ano de 2021.

#### 4.5.3 Liga Francesa

Seguidamente, foi realizada uma alteração ao estudo da liga em questão, visto que a liga dos campeões representa uma competição em que as equipas se defrontam poucas vezes e com um grande intervalo de tempo.

Tabela 5 - Resultados obtidos na Liga Francesa com 3 possíveis resultados de *output* usando RandomForestClassifier

Informação	Accuracy Score
Dados de Teste de 2011 até 2021 3 opções de resultado	0.5676
Dados de Teste de 2011 até 2020 3 opções de resultado	0.5714
Dados de Teste de 2011 até 2019 3 opções de resultado	0.5467

Tabela 6 - Resultados obtidos na Liga Francesa com 2 possíveis resultados de *output* usando RandomForestClassifier

Informação	Accuracy Score	Precision Score
Dados de Teste de 2011 até 2021 2 opções de resultado	0.7012	0.71
Dados de Teste de 2011 até 2020 2 opções de resultado	0.6965	0.69
Dados de Teste de 2011 até 2019 2 opções de resultado	0.66425	0.67

#### 4.5.4 Liga Portuguesa

Posteriormente, foi realizada um estudo à liga portuguesa.

Tabela 7 - Resultados obtidos na Liga Portuguesa com 3 possíveis resultados de *output* usando RandomForestClassifier

Informação	Accuracy Score
Dados de Teste de 2011 até 2021 3 opções de resultado	0.6024
Dados de Teste de 2011 até 2020 3 opções de resultado	0.5907
Dados de Teste de 2011 até 2019 3 opções de resultado	0.5757

Tabela 8 - Resultados obtidos na Liga Portuguesa com 2 possíveis resultados de *output* usando RandomForestClassifier

Informação	Accuracy Score	Precision Score
Dados de Teste de 2011 até 2021 2 opções de resultado	0.7412	0.7385
Dados de Teste de 2011 até 2020 2 opções de resultado	0.7365	0.7124
Dados de Teste de 2011 até 2019 2 opções de resultado	0.72425	0.7211

#### 4.5.5 Liga Inglesa

Por último foi realizado um estudo na liga inglesa, a liga de futebol mais popular em todo mundo.

Tabela 9 - Resultados obtidos na Liga Inglesa com 3 possíveis resultados de *output* usando RandomForestClassifier

Informação	Accuracy Score
Dados de Teste de 2011 até 2021 3 opções de resultado	0.6147
Dados de Teste de 2011 até 2020 3 opções de resultado	0.6012
Dados de Teste de 2011 até 2019 3 opções de resultado	0.5991

Tabela 10 - Resultados obtidos na Liga Inglesa com 2 possíveis resultados de *output* usando RandomForestClassifier

Informação	Accuracy Score	Precision Score
Dados de Teste de 2011 até 2021 2 opções de resultado	0.71	0.7087
Dados de Teste de 2011 até 2020 2 opções de resultado	0.6997	0.6912
Dados de Teste de 2011 até 2019 2 opções de resultado	0.6754	0.6894

#### 4.5.6 Análise Comparativa

Através da comparação dos resultados das Ligas Nacionais, a francesa, a inglesa e a portuguesa, com a Liga dos Campeões, conclui-se que o algoritmo usado consegue prever com maior *accuracy* os resultados dos jogos das ligas nacionais do que os jogos da liga dos campeões, tal como se pode observar, sinteticamente, na Tabela 11, Tabela 12 e Tabela 13.

Tabela 11 – Comparação entre a Liga dos Campeões e a Liga Francesa

Informação	Accuracy Score Liga dos Campeões	Accuracy Score Liga Francesa	Precision Score Liga dos Campeões	Precision Score Liga Francesa
Dados de Teste de 2011 até 2021 2 opções de resultado	0.6376	0.70124	0.62	0.71
Dados de Teste de 2011 até 2020 2 opções de resultado	0.6237	0.6965	0.5851	0.69
Dados de Teste de 2011 até 2019 2 opções de resultado	0.6403	0.66425	0.5925	0.67

Tabela 12 – Comparação entre a Liga dos Campeões e a Liga Portuguesa

Informação	Accuracy Score Liga dos Campeões	Accuracy Score Liga Portuguesa	Precision Score Liga dos Campeões	Precision Score Liga Portuguesa
Dados de Teste de 2011 até 2021 2 opções de resultado	0.6376	0.7412	0.7385	0.71
Dados de Teste de 2011 até 2020 2 opções de resultado	0.6237	0.7365	0.7124	0.69
Dados de Teste de 2011 até 2019 2 opções de resultado	0.6403	0.72425	0.7211	0.67

Tabela 13 – Comparação entre a Liga dos Campeões e a Liga Inglesa

Informação	Accuracy Score Liga dos Campeões	Accuracy Score Liga Inglesa	Precision Score Liga dos Campeões	Precision Score Liga Inglesa
Dados de Teste de 2011 até 2021 2 opções de resultado	0.6376	0.71	0.7087	0.71
Dados de Teste de 2011 até 2020 2 opções de resultado	0.6237	0.6997	0.5851	0.6912
Dados de Teste de 2011 até 2019 2 opções de resultado	0.6403	0.6754	0.5925	0.6894

Para além do facto de se analisar que as ligas nacionais apresentam uma maior probabilidade de *accuracy*, devido ao facto de serem ligas que se realizam de forma mais regular e em que as equipas se defrontam um maior número de vezes, também é possível observar que entre ligas não existe um grande diferencial de probabilidade de *accuracy*, mas destaca-se a liga portuguesa que apresenta uma maior *accuracy*, demonstrando que os jogos neste liga apresentam-se de mais fácil compreensão pelo algoritmo.



# 5 Conclusão

Esta capítulo tem como intuito a identificação e exposição dos objetivos alcançados, assim como de todos os outros que não se puderam realizar, sendo apresentada, logo de seguida, a lista de limitações presentes no decorrer do projeto.

Por fim, é exposto o trabalho a realizar no futuro, terminando com uma apreciação que possui uma crítica sobre o desenvolvimento do projeto, numa visão mais académica.

## 5.1 Objetivos concretizados

Com este projeto, foi possível implementar um processo de previsão de resultados em jogos de futebol recorrendo a modelos de IA. O sistema foi capaz de prever se uma equipa iria ganhar, empatar ou perder, bem como numa outra versão se a equipa iria ganhar ou pelo menos não perder.

Apesar disso, não foi possível concretizar todos os objetivos inicialmente delineados. Assim sendo, apresenta-se, de seguida, os objetivos concretizados:

- Levantar o estado arte relacionado com extração e tratamento de dados, modelos de previsão baseados em IA, e soluções para previsão no âmbito de apostas desportivas;
- Implementar um processo de extração e tratamento de dados;
- Implementar um modelo de IA para previsão da equipa vencedora de um jogo de futebol;
- Analisar os resultados e retirar valor de negócio com a informação gerada pelo modelo de IA;
- Discutir e reportar as principais conclusões tiradas através deste estudo.

## 5.2 Trabalho Futuro e Limitações

A realização deste trabalho, tal como expectável, permitiu pôr em prática alguns dos fundamentos teóricos inicialmente abordados.

Contudo, e tal como anteriormente mencionado, não foi possível cumprir com todas as finalidades deste projeto. Assim, como forma de suprir as lacunas inerentes ao mesmo, surge a necessidade de, em pesquisas/estudos posteriores, proceder à implementação dos algoritmos de CNN e LSTM para a realização de comparações de desempenho. De igual forma, será necessário proceder a um estudo mais aprofundado dos dados para que se consiga extrair a máxima informação do conjunto de dados obtido.

Para finalizar, a análise dos resultados e das *features* que possuam um papel preponderante na aprendizagem automática também se revelará um fator importante para acrescer valor a pesquisas futuras neste domínio.

### **5.3 Apreciação final**

O presente projeto permitiu estudar, de uma forma mais aprofundada, o mundo das apostas desportivas, bem como a utilização de técnicas de IA para a obtenção de informação relevante, na imensidão de dados existentes, associada a várias partidas de futebol.

Deste modo, foi possível explorar várias abordagens realizadas por diversos autores em trabalhos relacionados, o que desempenhou um papel crucial na aquisição de conhecimentos úteis para a realização deste projeto.

De facto, e citando o meu relatório de estágio de 2019, “Na fase final deste projeto é possível perceber todo o meu entusiasmo. A sua consecução permitiu o desenvolvimento de capacidades e conhecimentos que poderão ser úteis em domínios da minha atividade profissional, possibilitando ainda o desenvolvimento de técnicas ligadas à inteligência artificial, o que me fascina ainda mais.” [43].

Com efeito, a aplicação de técnicas de IA ao mundo do futebol sempre se apresentou como uma aspiração para mim, devido ao meu fascínio por este desporto, além de que, ao ingressar na licenciatura em Engenharia Informática, o meu entusiasmo pelo ramo da inteligência artificial se revelou, desde logo, notório.

Para terminar, e de um modo geral, estou bastante satisfeito com o desenvolvimento deste trabalho e com as aprendizagens daqui resultantes que me permitiram consolidar inúmeros conceitos teóricos abordados nos últimos dois anos deste ciclo de estudos. Assim, e apesar de estar consciente dos aspetos a aprimorar e da imensidão de conhecimentos ainda por apreender, estou confiante de estar na presença de um importante marco da minha vida académica e profissional. Este assegurar-me-á, indubitavelmente, um futuro promissor pautado, agora, por um número acrescido de possibilidades profissionais.

# Bibliografia

- [1] “Sports Betting Market Size, Share & Trends Analysis By Platform, By Type, By Sports Type (Football, Basketball, Baseball, Horse Racing, Cricket, Hockey, Others), By Region, And Segment Forecasts, 2022 - 2030,” 2022. [Online]. Available: <https://www.grandviewresearch.com/industry-analysis/sports-betting-market-report>.
- [2] Serviço de Regulação e Inspeção de Jogos, “Registo da atividade de jogo online em Portugal,” 2022. [Online]. Available: [https://www.srij.turismodeportugal.pt/fotos/editor2/estatisticas/estatisticas\\_online\\_2t\\_2022.pdf](https://www.srij.turismodeportugal.pt/fotos/editor2/estatisticas/estatisticas_online_2t_2022.pdf).
- [3] Pedro Morais, “OS ERROS A EVITAR E OS 8 CONSELHOS PARA BEM PRINCIPIAR!,” 2016. <https://www.pedromoraistrader.com/os-erros-a-evitar-e-os-8-conselhos-para-bem-principiar/>.
- [4] G. Somasundaram, A. Shrivastava, and E. E. Services, *Armazenamento e Gerenciamento de Informações: Como armazenar, gerenciar e proteger informações digitais*. 2009.
- [5] Lúcio Colangelo Filho, *Gestão da qualidade de dados*. 1998.
- [6] C. Sezões, J. Oliveira, and M. Baptista, *Business Intelligence*. 2006.
- [7] D. B. DOS SANTOS, “Visualização de dados estruturados e não estruturados da área da saúde,” Universidade Estadual Paulista “Júlio de Mesquita Filho,” 2021.
- [8] L. Zhang, Z. Qi, and F. Meng, “A Review on the Construction of Business Intelligence System Based on Unstructured Image Data,” *Procedia Comput. Sci.*, vol. 199, pp. 392–398, 2021, doi: 10.1016/j.procs.2022.01.048.
- [9] M. Souibgui, F. Atigui, S. Zammali, S. Cherfi, and S. Ben Yahia, “Data quality in ETL process: A preliminary study,” *Procedia Comput. Sci.*, vol. 159, pp. 676–687, 2019, doi: 10.1016/j.procs.2019.09.223.
- [10] Stitch, “ETL Database.” <https://www.stitchdata.com/etldatabase/>.
- [11] C. Ramos, J. C. Augusto, and D. Shapiro, “Ambient intelligencethe next step for artificial intelligence,” *IEEE Intell. Syst.*, vol. 23, no. 2, pp. 15–18, 2008, doi: 10.1109/MIS.2008.19.
- [12] J. M. F. Cardona, “Inteligência artificial na qualidade de dados,” Universidade Nova de Lisboa, 2020.
- [13] G. K. S. Bharatwaja Namatherdhala, Noman Mazher, “ARTIFICIAL INTELLIGENCE IN PRODUCT MANAGEMENT: SYSTEMATIC REVIEW,” 2022.
- [14] A. M. Turing, *Computing Machinery and Intelligence*. 1950.
- [15] S. V Sohan Lal Solanki, Saneya Pandrowala, Abhirup Nayak, Manish Bhandare, Reshma P Ambulkar and Shrikhande, “Artificial intelligence in perioperative management of

major gastrointestinal surgeries,” *World J. W J G Gastroenterol.*, vol. 27, 2021, doi: 10.3748/wjg.v27.i21.2758.

- [16] Stuart Russell and Peter Norvig, *Artificial Intelligence A Modern Approach*. 2022.
- [17] R. J. da Silva, “A INTELIGÊNCIA ARTIFICIAL NO CONTEXTO DA CIÊNCIA DA INFORMAÇÃO: UMA ANÁLISE DE DOMÍNIO,” Universidade do Porto, 2021.
- [18] I. Saidani, A. Ouni, and M. W. Mkaouer, “Improving the prediction of continuous integration build failures using deep learning,” 2022. doi: <https://doi.org/10.1007/s10515-021-00319-5>.
- [19] B. Namatherdhala, N. Mazher, and G. K. Sriram, “ARTIFICIAL INTELLIGENCE IN PRODUCT MANAGEMENT: SYSTEMATIC REVIEW,” *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 04, no. 07, 2022, [Online]. Available: <http://www.irjmets.com/>.
- [20] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*, Second. O’Reilly Media, Inc, 2019.
- [21] A. L. Samuel, “Some Studies in Machine Learning Using the Game of Checkers,” *IBM J.*, 1959.
- [22] T. M. Mitchell, *Machine Learning*. McGraw-Hill Science/Engineering/Math, 1997.
- [23] vask3k, “The map of the machine learning world,” [Online]. Available: [https://vas3k.com/blog/machine\\_learning/](https://vas3k.com/blog/machine_learning/).
- [24] A. OSENI, N. MOUSTAFA, H. JANICKE, P. LIU, Z. TARI, and A. VASILAKOS, “Security and Privacy for Artificial Intelligence: Opportunities and Challenges,” *J. ACM*, vol. 37, 2020.
- [25] A. Wolfewicz, “Deep Learning vs. Machine Learning – What’s The Difference?,” 2022. <https://levity.ai/blog/difference-machine-learning-deep-learning>.
- [26] Aladár Kollár, “Betting models using AI: a review on ANN, SVM, and Markov chain,” 2021. [Online]. Available: <https://mpra.ub.uni-muenchen.de/106821/>.
- [27] M. O. Okwu and L. K. Tartibu, “Artificial Neural Network,” 2021.
- [28] V. S. dos Santos, “Neurônio.” <https://mundoeducacao.uol.com.br/biologia/neuronios.htm>.
- [29] O. Davydova, “7 Types of Artificial Neural Networks for Natural Language Processing,” 2017. <https://www.kdnuggets.com/2017/10/7-types-artificial-neural-networks-natural-language-processing.html>.
- [30] S. J. Hochreiter S, *Long short-term memory*. 1997.
- [31] F. Landi, L. Baraldi, M. Cornia, and R. Cucchiara, “Working Memory Connections for LSTM,” 2021. doi: <https://doi.org/10.1016/j.neunet.2021.08.030>.
- [32] D. S. Academy, “Arquitetura de Redes Neurais Long Short Term Memory (LSTM),” 2022. <https://www.deeplearningbook.com.br/arquitetura-de-redes-neurais-long-short-term-memory/>.

- [33] E. Kang, "Long Short-Term Memory (LSTM): Concept," *medium*, 2017. <https://medium.com/@kangeugine/long-short-term-memory-lstm-concept-cb3283934359>.
- [34] S.P. Leo Kumar, "Knowledge-based expert system in manufacturing planning: state-of-the-art review.," <https://doi.org/10.1080/00207543.2018.1424372>, 2018.
- [35] I. Reis, D. Baron, and S. Shahaf, "Probabilistic Random Forest: A Machine Learning Algorithm for Noisy Data Sets," 2018, doi: <https://doi.org/10.3847/1538-3881/aaf101>.
- [36] Nasiba Mahdi Abdulkareem and A. M. Abdulazeez, "Machine Learning Classification Based on Radom Forest Algorithm: A Review," *Int. J. Sci. Bus.*, vol. 5, pp. 128–142, 2021.
- [37] J. Boksem, "Finding Ways to Improve the Prediction Accuracy of a Model that Predicts the Outcome of a Football Match using Machine Learning," University of Twente, 2022.
- [38] C. Game, "Dos campos de batalha para as casas de apostas online."
- [39] C. Ferro, "Cada jornada de futebol movimentou 340 milhões em apostas," *Diário de Notícias*, 2018, [Online]. Available: <https://www.dn.pt/portugal/interior/cada-jornada-de-futebol--movimentou-340-milhoes-em-apostas-9080898.html>.
- [40] Editorial, "Porque as Odds se Alteram?," *Apostas Casino Brasil*, 2022. <https://apostascasinobrasil.com/porque-as-odds-se-alteram/>.
- [41] ARQSI-ISEP, "Introdução à Arquitetura de Software," 2019.
- [42] Todd Helmenstine, "What Is the Difference Between Accuracy and Precision?," *Science Notes*, 2021. <https://sciencenotes.org/what-is-the-difference-between-accuracy-and-precision/>.
- [43] P. Alves, "Business Intelligence em Apostas Desportivas," 2019.
- [44] A. Fernandes, "O que é API? Entenda de uma maneira simples," 2018. <https://vertigo.com.br/o-que-e-api-entenda-de-uma-maneira-simples/>.
- [45] SmartBear, "API Endpoints." <https://smartbear.com/learn/performance-monitoring/api-endpoints/>.
- [46] Alexandre, "O que é JSON," 2011. <https://www.devmedia.com.br/o-que-e-json/23166>.
- [47] ESOFT-ISEP, "Engenharia de Requisitos, Análise e Design," 2018.
- [48] Knowledge, "O que é a User Story?," 2019. <https://www.knowledge21.com.br/sobreagilidade/user-stories/o-que-e-user-story/>.
- [49] G. Primo, "User Stories o que são? Como usar?," 2019. <https://blog.myscrumhalf.com/user-stories-o-que-sao-como-usar/>.



# Anexos

## Anexo 1 Glossário dos Requisitos

Tabela 14 - Glossário de Requisitos

Termo/Expressão	Significado/Descrição
<b>Prognóstico</b>	Palpite sobre um resultado esperado de um determinado jogo.
<b>Sites de Apostas</b>	Local onde se efetuam as apostas monetárias.
<b>ODD</b>	Valor decimal que representa a probabilidade de um acontecimento ocorrer, e de quanto percentualmente é que uma pessoa lucra ao apostar naquela ODD. Por exemplo, uma ODD de 1,33 significa que a probabilidade de aquele acontecimento ocorrer é de 75% (1/1.33), e que ao apostar 1 euro se lucra 0.33 em caso de sucesso.
<b>Sucesso do Prognóstico</b>	Identifica o estado do prognóstico, se este foi ganha, devolvido, perdido, ou se ainda está à espera do resultado.
<b>Equipa Visitada</b>	Equipa que joga na sua casa, normalmente no seu estádio.
<b>Equipa Visitante</b>	Equipa que joga fora de casa.
<b>G</b>	Golos marcados.
<b>BTS</b>	Both Teams Score, traduzindo, ambas as equipas marcam.
<b>W</b>	Winner, traduzindo, Vencedor
<b>RC</b>	Serve para identificar o mercado de prognóstico "Resultado Correto"
<b>HT</b>	Half-Time que significa intervalo.
<b>FT</b>	Full-Time que significa tempo total de jogo.
<b>ET</b>	Extra-Time que significa prolongamento.
<b>P</b>	Penalties.

## Anexo 2 Extração dos Dados da API

É possível observar o código de extração através do link: <https://github.com/pedrocbt21/pedrocbt21/blob/main/ApiSports.ipynb>.

Nesse link é apresentado o processo de ligação à API bem como a extração e um primeiro tratamento aos dados recebidos da mesma.

### Standings

Deve ser atualizado uma vez por dia na season atual

<https://www.api-football.com/documentation-v3#tag/Standings>

```
InÃ [A ~
def getSeasonStandings(season, leagueId):
    seasonStandingsUrl = "standings?season=" + str(season) + "&league=" + str(leagueId)
    jsonLeagueSeasonStandings = getApiSports(seasonStandingsUrl)
    path = "standing/league/" + str(leagueId) + "/" + str(season)
    filename = str(leagueId) + "_" + str(season) + "_season_standings"
    saveJson(path, filename, jsonLeagueSeasonStandings)
    return jsonLeagueSeasonStandings

InÃ [A ~
def getAllSeasonStandings(listOfSeasonLeague):
    listOfSeasonLeagueStandings = []

    for x in listOfSeasonLeague:
        teamSeasonLeagueJson = json.loads(x)
        jsonLeagueSeasonStandingsResp = getSeasonStandings(teamSeasonLeagueJson["season"], teamSeasonLeagueJson["league"])

        listOfSeasonLeagueStandings.append(str(jsonLeagueSeasonStandingsResp["response"]))
    saveJson("standing", "allSeasonLeagueStanding", listOfSeasonLeagueStandings)
    return listOfSeasonLeagueStandings

InÃ [A ~
#ListofSeasonLeague = getAllSeasonLeagueInformation(jsonLeagues)
#ListofSeasonLeagueStandings = getAllSeasonStandings(ListofSeasonLeague)
listOfSeasonLeagueStandings = importJson("standing/allSeasonLeagueStanding.json")
```

### Fixtures

Update Frequency: This endpoint is updated every 15 seconds.

Recommended Calls: 1 call per minute for the leagues, teams, fixtures who have at least one fixture in progress otherwise 1 call per day.

<https://www.api-football.com/documentation-v3#tag/Fixtures>

```
InÃ [A ~
def getFixturesLive(leagueId):
    fixturesRoundsLiveUrl = f"fixtures?live=all&league={str(leagueId)}"
    jsonFixturesLive = getApiSports(fixturesRoundsLiveUrl)
    return jsonFixturesLive

InÃ [A ~
def getFixtures(season, leagueId):
    fixturesRoundsUrl = f"fixtures?season={str(season)}&league={str(leagueId)}"
    jsonFixtures = getApiSports(fixturesRoundsUrl)
    path = f"fixtures/league/{str(leagueId)}/{str(season)}"
    filename = f"{str(leagueId)}_{str(season)}_fixtures"
    saveJson(path, filename, jsonFixtures)
    return jsonFixtures

InÃ [A ~
def getAllFixtures(listOfSeasonLeague):
    listOfSeasonLeagueFixtures = []

    for x in listOfSeasonLeague:
        fixtureSeasonLeagueJson = json.loads(x)
        jsonSeasonLeagueFixturesResp = getFixtures(fixtureSeasonLeagueJson["season"], fixtureSeasonLeagueJson["league"])
        for y in jsonSeasonLeagueFixturesResp["response"]:
            jsonSeasonLeagueFixtures = "{\n\"fixture\": " + str(y["\"fixture\"]["\"id\"]") + " ,\n\"league\": " + str(y["\"league\"]["\"id\"]") + " ,\n\"season\": " + str(y["\"leag

        listOfSeasonLeagueFixtures.append(str(jsonSeasonLeagueFixtures))

    saveJson("fixtures", "allSeasonLeagueFixtures", listOfSeasonLeagueFixtures)
    return listOfSeasonLeagueFixtures

InÃ [A ~
#ListofSeasonLeagueFixtures = getAllFixtures(ListofSeasonLeague)
listOfSeasonLeague = getAllSeasonLeagueInformation(jsonLeagues)
listOfSeasonLeagueFixtures = importJson("fixtures/allSeasonLeagueFixtures.json")

InÃ [A ~
#jsonFixturesLive = getFixturesLive(169) #Liga chinesa teste
```

Figura 18 - Parte do código Python de Extração dos Dados da API

## Anexo 3 Tratamento e Análise dos Resultados

É possível observar o código de tratamento e de análise dos resultados através do link: <https://github.com/pedrocbt21/pedrocbt21/blob/main/SportsTreatment.ipynb>.

Nesse link é apresentado o processo de tratamento dos dados, treino e teste do modelo e análise de resultados.

### Tratamento dos dados

#### Equipas

```
InÃ [A -
leagueIdfilter = 203
filename = f"{filePath}/fixtures/league/203/2011/509613/203_2011_509613_fixtures_predictions.json"

#f"{filePath}/fixtures/league/848/2021/714875/848_2021_714875_fixtures_predictions.json"
#f"{filePath}/fixtures/league/39/2011/193817/39_2011_193817_fixtures_predictions.json"
#f"{filePath}/fixtures/league/203/2011/509613/203_2011_509613_fixtures_predictions.json"
# Opening JSON file
f = open(filename)

# returns JSON object as
# a dictionary
data = json.load(f)

#df = pd.read_json(filename)
#lstoOfSeasonLeagueFixtures

InÃ [A -
#data["response"][0]

InÃ [A -
fixturePrediction = pd.json_normalize(data["response"],max_level=6)
fixtureId = data["parameters"]["fixture"]
fixturePrediction["fixtureId"] = fixtureId
#fixturePrediction

InÃ [A -
fixtureClean = fixturePrediction[["fixtureId", 'league.id', 'league.name', 'teams.home.id', 'teams.home.name', 'teams.away.id', 'teams.away.name', 'pred
'predictions.goals.home', 'predictions.goals.away', 'predictions.percent.home', 'predictions.percent.draw', 'predictions.percent.away', 'predictions
'teams.home.last_5_form', 'teams.home.last_5_att', 'teams.home.last_5_def', 'teams.home.last_5.goals.for.average', 'teams.home.last_5.goals.against.a
'teams.away.last_5_form', 'teams.away.last_5_att', 'teams.away.last_5_def', 'teams.away.last_5.goals.for.average', 'teams.away.last_5.goals.against.a
'teams.home.league.form', 'teams.home.league.fixtures.wins.home', 'teams.home.league.fixtures.wins.away',
'teams.home.league.fixtures.draws.home', 'teams.home.league.fixtures.draws.away', 'teams.home.league.fixtures.loses.home', 'teams.home.league.fixtures
'teams.home.league.goals.for.average.home', 'teams.home.league.goals.for.average.away',
'teams.away.league.form', 'teams.away.league.fixtures.wins.home', 'teams.away.league.fixtures.wins.away',
'teams.away.league.fixtures.draws.home', 'teams.away.league.fixtures.draws.away', 'teams.away.league.fixtures.loses.home', 'teams.away.league.fixtures
'teams.away.league.goals.for.average.home', 'teams.away.league.goals.for.average.away'
]]

InÃ [A -
#fixtureClean

InÃ [A -
def ImportPredictionsDataFromJsonFiles(league,season,fixture,df_fixturePredictions):
    filename = f"{filePath}/fixtures/league/{str(league)}/{str(season)}/{str(fixture)}/{str(league)}_{str(season)}_{str(fixture)}_fixtures_predictions.j
    # Opening JSON file
    f = open(filename)
    # returns JSON object as a dictionary
    data = json.load(f)
    fixturePrediction = pd.json_normalize(data["response"],max_level=6)
    fixturePrediction["fixtureId"] = data["parameters"]["fixture"]
    fixtureClean = fixturePrediction[["fixtureId", 'league.id', 'league.name', 'league.season', 'teams.home.id', 'teams.home.name', 'teams.away.id', 'tear
'predictions.goals.home', 'predictions.goals.away', 'predictions.percent.home', 'predictions.percent.draw', 'predictions.percent.away', 'predict
'teams.home.last_5_form', 'teams.home.last_5_att', 'teams.home.last_5_def', 'teams.home.last_5.goals.for.average', 'teams.home.last_5.goals.agai
'teams.away.last_5_form', 'teams.away.last_5_att', 'teams.away.last_5_def', 'teams.away.last_5.goals.for.average', 'teams.away.last_5.goals.agai
'teams.home.league.form', 'teams.home.league.fixtures.wins.home', 'teams.home.league.fixtures.wins.away',
'teams.home.league.fixtures.draws.home', 'teams.home.league.fixtures.draws.away', 'teams.home.league.fixtures.loses.home', 'teams.home.league.fix
'teams.home.league.goals.for.average.home', 'teams.home.league.goals.for.average.away',
'teams.away.league.form', 'teams.away.league.fixtures.wins.home', 'teams.away.league.fixtures.wins.away',
'teams.away.league.fixtures.draws.home', 'teams.away.league.fixtures.draws.away', 'teams.away.league.fixtures.loses.home', 'teams.away.league.fix
'teams.away.league.goals.for.average.home', 'teams.away.league.goals.for.average.away'
    ]
    df_fixturePredictions = df_fixturePredictions.append(fixtureClean, ignore_index=True)
    return df_fixturePredictions
```

Figura 19 - Parte do código Python de Tratamento dos Dados da API