



UNIVERSIDADE DO MINHO  
ESCOLA DE ENGENHARIA

---

# Sistemas Autónomos

---

Inteligência Ambiental e Sensorização

Inês Alves A81368  
José Pedro Pinto A80741  
Marta Ribeiro A82474

**MIEI - 4.º Ano - Sistemas Autónomos**

Braga, 22 de Março de 2020

## Conteúdo

<b>1</b>	<b>Introdução</b>	<b>2</b>
<b>2</b>	<b>Objetivos</b>	<b>3</b>
<b>3</b>	<b>Sensorização</b>	<b>3</b>
3.1	APIs Utilizadas . . . . .	4
<b>4</b>	<b>Tratamento de Dados</b>	<b>4</b>
<b>5</b>	<b>Modelo de Aprendizagem</b>	<b>6</b>
<b>6</b>	<b>Sistema Montado</b>	<b>6</b>
<b>7</b>	<b>Conclusões e Trabalho Futuro</b>	<b>8</b>

# 1 Introdução

Este projeto tem como principal objetivo o desenvolvimento de um sistema inteligente capaz de gerar conhecimento num determinado contexto de recolha de informação através de sensores. Tendo em conta que atualmente o mundo enfrenta um grave problema de saúde pública: a pandemia causada pelo vírus **COVID-19**, optamos por incluir essa temática na formulação do problema. Esta pandemia afeta já diversas áreas, pelo que se torna essencial analisar todos os efeitos associados. Para além disso, por ser um tema tão recente, atual e de constante mudança, revela-se um excelente desafio no que toca à aplicação de conhecimentos de análise e previsão nesta vertente, não só em termos de saúde pública, mas também no que diz respeito a efeitos secundários e a longo prazo que esta pandemia poderá trazer.

Posto isto, e percebendo que os efeitos colaterais deste tema são já bastante avassaladores, podemos observar que também o preço das criptomoedas foi abalado com esta crise. Sendo estes preços muito voláteis, analisar as suas flutuações é também de grande interesse económico e um apetecível caso de estudo. Já existem projetos com objetivo de monitorizar e prever os preços destas comodidades, no entanto, à luz dos recentes eventos, torna-se interessante desenvolver um sistema que tenha também em conta dados relativos à evolução da pandemia que enfrentamos.

Neste contexto, os sensores virtuais desempenham um papel fundamental para que seja possível recolher informação atual e fidedigna, permitindo, desta forma, que possamos gerar conhecimento e (possivelmente) oportunidades de investimento.

O presente relatório serve como apresentação do trabalho até agora desenvolvido e dos próximos passos que tomaremos no desenvolvimento deste sistema.

## 2 Objetivos

O objetivo prático do nosso sistema é ser capaz de fazer previsões sobre a variação dos preços de uma seleção de quatro criptomoedas (Bitcoin, Ethereum, Ripple e Litecoin) com base num modelo de Machine Learning que utilize dados relativos às flutuações deste mercado e à evolução da pandemia do COVID-19.

Assim, entendemos que é fundamental recolher, sistematicamente, informações relativas aos **preços** das criptomoedas, à **opinião pública atual** sobre as mesmas, bem como às **estatísticas da pandemia**. Para tal, o grupo optou por recorrer a sensores virtuais que nos permitem adquirir dados de qualidade, fundamentais para alimentar os modelos de aprendizagem implementados. Estes baseiam-se em chamadas a APIs através de pedidos HTTP.

## 3 Sensorização

Recorremos a três sensores virtuais, respetivamente destinados a obter informação sobre:

- **Preços das Criptomoedas**

Rácio de troca de cada uma das moedas escolhidas em relação ao dólar americano.

- **Estatísticas do COVID-19**

Para cada um dos 11 países escolhidos (10 mais afetados e Portugal), recolhemos:

- Número total de infetados
- Número de novos casos (casos que surgiram no dia desta recolha de dados)
- Número total de mortes

- **Opinião Pública**

Para cada criptomoeda, procuramos *tweets* numa seleção de *hashtags* do *Twitter* e procedemos a um simples processamento de linguagem natural que retorna o sentimento (positivo ou negativo) de cada tweet encontrado (segundo a API do twitter, são devolvidos os mais recentes e relevantes, tal como apareceria no *feed* de um utilizador. Assim, de acordo com o rácio de *tweets* positivos/negativos, classificamos cada medição da opinião pública como positiva, neutra ou negativa.

### 3.1 APIs Utilizadas

Para obter os preços das criptomoedas, utilizámos um plano grátis de um serviço chamado `coinapi.io`. Existem APIs sem custos, que não limitam o acesso para utilizadores de planos deste género, no entanto também não costumam ser tão fiáveis, portanto optámos por esta para já.

No caso dos dados relativos ao COVID, existe neste momento um grande esforço por parte de diversas entidades no sentido de disponibilizar informação atualizada de forma gratuita para que mais pessoas possam ajudar a combater esta crise. Assim, não foi difícil de encontrar uma API gratuita e fiável, recorrendo ao *site* `rapidapi.com`, que se especializa em listar APIs com todo o tipo de finalidades.

Para obter os *tweets*, o *Twitter* disponibiliza uma API e um SDK multiplataforma que facilita imenso esse tipo de tarefa. Para usufruir deste serviço apenas necessitámos de criar uma conta de *developer* para ter acesso a um conjunto de credenciais que nos permitem fazer os pedidos de informação.

## 4 Tratamento de Dados

De forma a que, numa fase posterior de desenvolvimento, seja possível processar os dados para obter a variação do preço das criptomoedas, nesta fase geramos um *dataset* com as informações recolhidas através dos sensores virtuais descritos anteriormente. De reforçar que, para avaliar de uma forma generalizada os tweets, foi implementada uma função simples de *nlp* que faz um balanço positivo (1), neutro (0.5) ou negativo (0). Assim sendo cada registo do *dataset* possui a seguinte estrutura:

- **time** - data a que o registo foi obtido;
- **BTC\_PRICE\_USD** - preço da Bitcoin;
- **ETH\_PRICE** - preço da Ethereum;
- **XRP\_PRICE\_USD** - preço da Ripple;
- **LTC\_PRICE\_USD** - preço da Litecoin;
- Para cada um dos países monitorizados:
  - **Country\_total\_cases** - número total de casos no país em questão;

- **Country\_new\_cases** - número de novos casos no país em questão;
- **Country\_total\_deaths** - número total de mortes no país em questão;
- **BTC\_twitter** - opinião pública relativa à bitcoin;
- **ETH\_twitter** - opinião pública relativa à ethereum;
- **XRP\_twitter** - opinião pública relativa à ripple;
- **LTC\_twitter** - opinião pública relativa à litecoin;

Futuramente, este *dataset* "cru" será processado de modo a torná-lo mais adequado ao tipo de aprendizagem que pretendemos utilizar (por exemplo introduzindo novas colunas, calculadas a partir das já existentes e relevantes para os modelos utilizados). Por exemplo, se usarmos *supervised* ou *reinforcement learning*, provavelmente incluiremos uma coluna que corresponda à variação observada em cada registo (e que pode ser obtida através de uma subtração entre valores de dois registos).

## 5 Modelo de Aprendizagem

Nesta fase ainda é complicado de dizer com certeza modelo e o algoritmo que utilizaremos para levar a cabo a aprendizagem e as previsões propriamente ditas.

O estado da arte, na previsão da variação de preços de criptomoedas, envolve geralmente na utilização de modelos centrados em RNNs (*Recurrent Neural Networks*), dada a sua provada eficácia com dados de carácter sequencial. Uma possível abordagem neste sentido requeriria um cuidadoso tratamento dos dados produzidos pelos sensores mencionados anteriormente.

Há outras abordagens mais simples que podemos adotar, fazendo regressões com outros tipos de modelos (e.g. *Gradient Boosting*, *Random Forest*...).

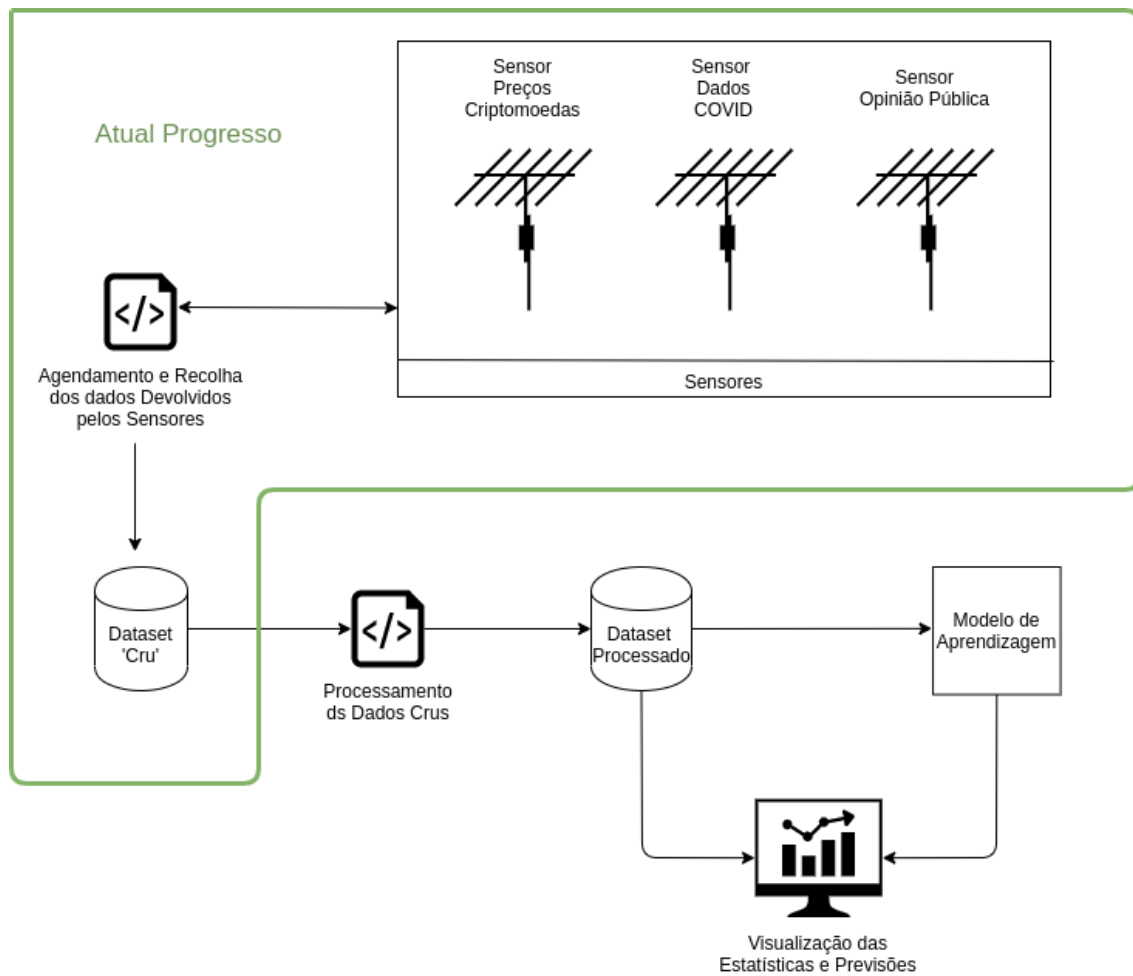
Uma opção ainda mais simples, menos ambiciosa, mas provavelmente ainda extremamente útil, seria adotar um modelo de classificação binária numa abordagem *One vs Rest* (para cada criptomoeda construir um modelo de classificação binária que preveja se o preço sobe ou desce).

Na próxima fase já teremos selecionado uma estratégia adequada ao contexto deste projeto.

## 6 Sistema Montado

Neste momento, temos já uma ideia da estrutura final do sistema que desenvolvemos. A seguinte figura representa essa estrutura, e a parte dela que de momento já desenvolvemos.

Apesar de poder aparentar que uma grande parte do sistema já se encontra implementada, ainda há importantes decisões a tomar, que podem afetar a maneira como queremos que os sensores (cujo funcionamento influencia todos os restantes componentes) recolham a informação.



**Figura 1.** Sistema a desenvolver



## 7 Conclusões e Trabalho Futuro

Em suma, até agora concluímos que o facto de trabalharmos com sensores virtuais permite-nos adotar para os mesmos um alto nível de abstração (obtendo informação mais cada vez mais próxima de conhecimento) com relativa facilidade. Esta escolha, no entanto, vem com um custo, que é o de perder o controlo sobre parte do funcionamento destes sensores (por exemplo, e se o *Twitter* pretende beneficiar o mercado das criptomoedas, dando menos prioridade aos *tweets* que não favoreçam essa agenda nos resultados de pesquisa? Isto prejudicaria a veracidade da informação retornada pelo nosso sensor de opinião pública, prejudicando o nosso sistema). Assim, quando trabalhamos com sensores virtuais, há que averiguar a maneira como os dados estão a ser recolhidos e prever situações em que determinadas origens de informação possam não ser fidedignas.

Futuramente, planeamos continuar o desenvolvimento deste sistema seguindo, por esta ordem, as seguintes tarefas:

- Escolher um modelo de aprendizagem (equilibrando a sua simplicidade e a utilidade das suas previsões)
- Desenvolver um processo para enriquecer o *dataset* que já estamos a construir e torná-lo adequado para ser consumido pelo modelo de aprendizagem
- *Fine-tuning* do funcionamento dos sensores considerando as novas implementações
- Desenvolvimento de uma interface para visualizar estatísticas alimentadas pelos dados recolhidos e pelo conhecimento produzido pelo modelo de aprendizagem