

# USO DE ÁRVORES DE DECISÃO PARA ANÁLISE DE DADOS NAS DISCIPLINAS DE LABORATÓRIO DE FÍSICA

Bruno Arthur Basso Silva<sup>1</sup>, Gabriela Molina Ciocci<sup>1</sup>, Eliane de Fátima Chinaglia<sup>2</sup>  
<sup>1,2</sup> Centro Universitário FEI

[brunobasso.contato@gmail.com](mailto:brunobasso.contato@gmail.com); [echinaglia@fei.edu.br](mailto:echinaglia@fei.edu.br)

**Resumo:** A formação em engenharia, tradicionalmente focada na técnica, precisa evoluir para atender às demandas da era tecnológica. A simulação e a comparação com dados experimentais, oferece uma nova abordagem para o ensino. Neste projeto, desenvolvemos uma árvore de decisão em Python capaz de determinar a massa de um objeto a partir do movimento de dois corpos em um trilho de ar, com o objetivo de demonstrar o potencial da inteligência artificial para alunos de engenharia.

## 1. Introdução

A rápida evolução da Inteligência Artificial (IA) exige uma atualização nos cursos de engenharia. Para acompanhar essa transformação, é fundamental que os futuros engenheiros desenvolvam competências em IA compreendendo seus fundamentos e aplicações. É necessário que os futuros engenheiros tenham uma compreensão básica de como os algoritmos de IA funcionam, a fim de empregá-los de maneira consciente, positiva e construtiva. Para isso, o Laboratório de Física irá contribuir no sentido de mostrar os conceitos básicos de uma máquina inteligente para capacitar os engenheiros. Essas tecnologias têm o potencial de aprimorar processos, otimizar a tomada de decisões e catalisar a inovação em uma vasta gama de domínios [1].

Este projeto visa introduzir os alunos de engenharia aos conceitos básicos de IA e Aprendizado de Máquina, com foco em agrupamento hierárquico e Árvores de Decisão, juntamente aos conceitos de Física básica. Um banco de dados será obtido a partir de um experimento com dois corpos acoplados em um trilho de ar, e uma Árvore de Decisão será treinada para identificar a massa de um dos corpos.

## 2. Metodologia

A primeira etapa para o desenvolvimento desse projeto consistiu na aquisição de dados para obter um banco de dados eficiente e confiável, com o intuito de treinar a IA. Como ilustrado na figura 1, foi utilizado um trilho de ar para a obtenção dos dados de aceleração de um sistema com 2 massas acopladas. Sensores ópticos posicionados estrategicamente permitiram a medida da velocidade de um dos corpos ( $m_1$ ) em função do tempo determinando assim a aceleração do sistema.

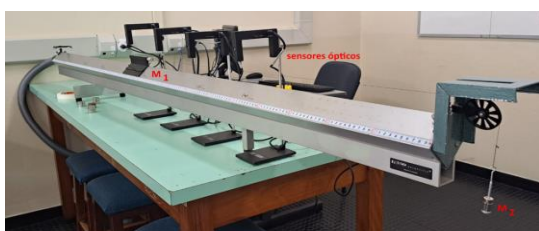


Figura 1 - Arranjo experimental do trilho de ar

Na tabela I está apresentada a variação de valores de massas utilizadas para a aquisição de dados.

Tabela I - Variação das massas utilizadas no experimento

$m_1$ (kg)	$m_2$ (kg)
0,209	0,004 a 0,013
0,309	0,014 a 0,019
0,409	0,021 a 0,031

Com o banco de dados em mãos, realizamos a análise de agrupamento hierárquico [2] com o objetivo de agrupar as massas  $m_2$ .

A Árvore de Decisão necessita de parâmetros de entrada para saber qual caminho será feito com a finalidade de encontrar a resposta [3]. Neste projeto vamos utilizar os seguintes parâmetros de entrada: a massa  $m_1$  e aceleração experimental. Como saída de classificação, esperamos identificar a massa  $m_2$ .

A programação do agrupamento hierárquico e da árvore de decisão está sendo realizada em Python.

## 3. Resultados

Para realizarmos uma análise exploratória do banco de dados, foi utilizado uma ferramenta gráfica chamada Box Plot, que nos ajuda a visualizar e analisar conjuntos de dados. Através de uma caixa, são apresentados a linha mediana, a forma da distribuição, tendência central, dispersão e presença de outliers de forma compacta e informativa [2]. O resultado para nosso banco de dados é apresentado na Figura 2.

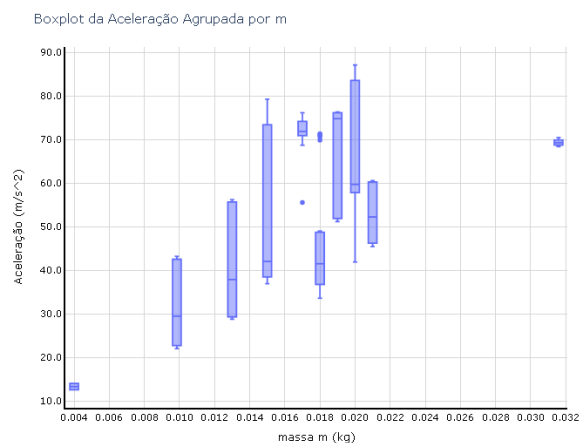


Figura 2 - Box Plot da aceleração em função do  $m_2$

Percebe-se uma grande dispersão dos valores de aceleração para as massas  $m_2$  entre os valores de 0,010 kg até 0,022 kg. Essa variabilidade dificultou que

as massas  $m_2$  fossem agrupadas de modo coerente. Como consequência, isso pode diminuir a acurácia na resposta da IA por se tratar de valores muito próximos. Assim, no prosseguimento desse projeto será necessário escolher somente dois valores de  $m_2$  no intervalo entre 0,010kg e 0,020kg, a fim de concentrar as informações sobre a aceleração.

Após essa análise exploratória, utilizamos o método do Elbow que consiste em ajudar os cientistas de dados a selecionarem o número ideal de clusters, ajustado por um ponto de inflexão da curva, também conhecido como “cotovelo” [4]. A análise do gráfico apresentado na figura 3 indica que podemos usar 5 clusters para o agrupamento dos nossos dados.

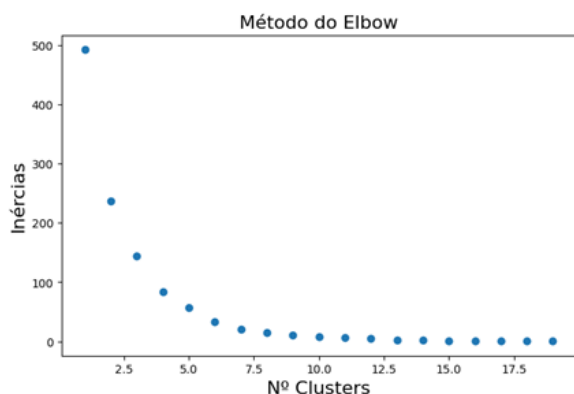


Figura 3 - Gráfico do método de Elbow

Sabendo a quantidade ideal de clusters, partimos para uma análise do agrupamento hierárquico cujo resultado é um dendrograma, um tipo de diagrama que mostra como os dados estão divididos em clusters.

Para compor o dendrograma é necessário usar métodos e métricas que definem os agrupamentos. Após vários testes, utilizamos o método *single* que se baseia no cálculo da distância mínima entre os pontos de duas observações distintas, juntamente com a métrica *euclidean*, que calcula a distância euclidiana entre eles [5].

A figura 4 nos mostra o dendrograma cujo eixo x se trata dos valores de  $m_2$  do banco de dados e o eixo y se refere aos valores dos cálculos da distância referente ao método escolhido.

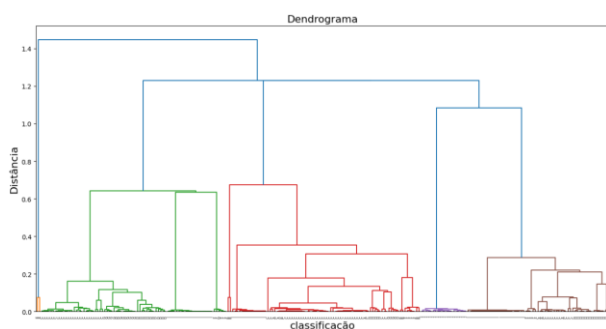


Figura 4 - Dendrograma do método *single* e métrica *euclidean*

Devido ao grande número de dados não é possível observar de modo adequado as informações da

classificação das massas  $m_2$  na figura. No entanto, analisando uma tabela com o resultado do agrupamento, percebe-se que as massas  $m_2$  não são agrupadas de modo coerente. Isso é resultado da dispersão de valores de aceleração apresentados na figura 2 para o intervalo de massa  $m_2$  entre 0,010kg e 0,020kg. Assim, o banco de dados ainda não está adequado para treinar a árvore de decisão.

#### 4. Conclusão

A aquisição de dados experimentais mostrou-se desafiadora, mas de extrema importância para um bom treinamento da árvore de decisão. Assim, nosso próximo passo é otimizar os valores de massa  $m_2$  que serão utilizados para a aquisição dos dados. Além disso, novos dados serão obtidos com o trilho de ar inclinado para termos mais um parâmetro a ser considerado na análise. Espera-se assim melhorar a qualidade do banco de dados de tal forma que variabilidade dos dados seja refletida nos processos de agrupamento e consequentemente no treinamento da árvore de decisão. Com o banco de dados aprimorado, teremos uma IA treinada para que em uma aula do laboratório de física 1, os alunos entrem com os valores da massa  $m_1$ , o ângulo de inclinação da rampa e a aceleração experimental e a IA classifique o valor da massa  $m_2$ , com uma certa acurácia.

#### 5. Referências

- [1] IBM, **AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the difference?**, 2023. Disponível em: <https://www.ibm.com/think/topics/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>. Acesso em: 09/08/2024.
- [2] FÁVERO. **MANUAL DE ANÁLISE DE DADOS ESTATÍSTICA E MODELAGEM MULTIVARIADA COM EXCEL, SPSS E STATA**, v.9.1 – v.9.2, p. 300 – 336, 2022. Acesso em: 09/08/2024.
- [3] KLOSTERMAN, Stephen. **PROJETOS DE CIÊNCIA DE DADOS COM PYTHON**. v.5. p – 207, 2020. Acesso em: 10/08/2024.
- [4] YELLOWBRICK. **ELBOW METHOD**. 2016. Disponível em: <https://www.scikit-yb.org/en/latest/api/cluster/elbow.html>. Acesso em: 11/08/2024.
- [5] PUIGBÒ, Pere. **DENDROUPGMA: A DENDROGRAM CONSTRUCTION UTILITY**. 2009. Disponível em: [https://usuaris.tinet.cat/debb/UPGMA/DendroUPGMA\\_Tut.pdf](https://usuaris.tinet.cat/debb/UPGMA/DendroUPGMA_Tut.pdf). Acesso em: 12/08/2024.

#### Agradecimentos

Ao Centro Universitário FEI pela disponibilidade de suas instalações e equipamentos e apoio financeiro.

<sup>1</sup>Aluno de IC do Centro Universitário FEI. Projeto com vigência de 11/2023 a 11/2023 PRO-BID004/23.