

Atividade 1

Aprendizado de máquina e inteligência artificial em física

Considere a base de dados contida no arquivo `Stars.csv`, disponível para download na página da disciplina no Moodle, sobre as características observadas de um conjunto de 240 estrelas.

Elabore um código em Python para o processamento desses dados de forma a responder as seguintes questões:

- 1) Como os dados devem ser preparados para o processamento?
- 2) Quais as variáveis devem ser reescaladas e por qual método?
- 3) Aplique a redução de variáveis pelo método PCA e determine o número de componentes necessárias para se explicar, pelo menos, 90% da variância dos dados.
- 4) Olhando as três primeiras componentes principais do PCA, quais as variáveis da tabela são mais importantes para o agrupamento dos dados?
- 5) Utilize o método de agrupamento hierárquico para agrupar as estrelas de acordo com suas características. Qual o número de agrupamentos ideal para análise desses dados?
- 6) Utilize o método KMeans para agrupar as estrelas de acordo com suas características. Qual o número de agrupamentos ideal para análise desses dados?
- 7) Utilize o método DBSCAN para agrupar as estrelas de acordo com suas características. Qual o número de agrupamentos obtido? Comente sobre a presença ou não de *outliers*.
- 8) Faça uma análise sobre a performance geral dos algoritmos utilizados nos itens de 5 a 7 comparando com as categorias definidas manualmente e que podem ser encontradas no arquivo `Categorias.csv`.