

O objetivo desta prática é introduzir o conceito de agrupamento de dados. Para isso, o aluno aplicará a técnica *k-means* e observará como a variação dos atributos modifica os grupos gerados. Além disso, analisará como a qualidade dos dados influencia na técnica de agrupamento.

Para esta prática você vai precisar:

- RapidMiner
- Editor de texto com suporte para figuras (ex: Microsoft Office Word, LibreOffice Writer, LaTeX)

A seguir são apresentados 2 tópicos: agrupamento de dados e ruídos. Escreva um relatório cumprindo os requisitos dos tópicos.

O relatório deve estar no formato pdf para submissão.

1. **Agrupamento de dados:** O que é aprendizado de máquina? O que é aprendizado não supervisionado? Qual o objetivo de uma técnica de agrupamento de dados (*clustering*)? Se você aplicar uma técnica de agrupamento em um conjunto de dados, o que você espera como resultado?

Neste trabalho utilizaremos a base *iris*. Podemos aplicar técnicas de agrupamento nela, mas por que dizemos que *iris* não é um problema de agrupamento?

A seguir, aplique e discuta os itens abaixo.

- Aplique a técnica *k-means* sobre a base *iris*
 - Observe os resultados usando Scatter plot
 - Compare os grupos formados pelo *k-means* com o rótulo (*label*) da base *iris*
 - Altere os parâmetros do *k-means*. Observe e compare novamente.
2. **Ruídos:** O que é ruído em mineração de dados? Como ele pode acontecer?
 - Simule ruído em um dos atributos da base *iris* através do operador *Add Noise*
 - Faça um novo agrupamento utilizando *k-means*. Como o ruído influenciou no agrupamento dos dados?