

O objetivo desta prática é introduzir o aluno sobre o tratamento de dados e exercitar a análise exploratória de um conjunto de dados. Para isso, espera-se que o aluno consiga, utilizando observações estatísticas e apresentações gráficas, encontrar relações entre atributos em um conjunto de dados e imputar valores ausentes.

Para esta prática você vai precisar:

- RapidMiner
- Editor de texto com suporte para figuras (ex: Microsoft Office Word, LibreOffice Writer, LaTeX)
- IMDB 5000 Movie Dataset (<https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset>)

A seguir são apresentados 4 tópicos: entendimento da base; matriz de correlação; valores ausentes e visualização. Escreva um relatório cumprindo os requisitos dos tópicos.

O relatório deve estar no formato pdf para submissão.

1. **Entendimento da base:** a fim de entender a base de dados, descreva o que cada um dos atributos representa.
2. **Correlação:**
 - O que significa a correlação entre dois atributos? O que é uma matriz de correlação? Quais atributos da base IMBD Movie Dataset fariam sentido em uma matriz de correlação?
 - Gere uma matriz de correlação entre os atributos. Quais atributos estão mais correlacionados? Gere um gráfico que mostre uma relação entre esses atributos.
3. **Valores ausentes:**
 - O que são valores ausentes? Por quais motivos uma base pode possuir valores ausentes? Quais técnicas são comuns para preencher valores ausentes?
 - Verifique quais atributos da base IMDB Movie possuem valores ausentes. tente preencher os valores de algum atributo. Reporte o que fez.