

O objetivo desta prática é aplicar regras de associação em um contexto biomolecular. Espera-se que o aluno consiga gerar regras de associação e analisá-las.

Para esta prática você vai precisar:

- RapidMiner
- Editor de texto com suporte para figuras (ex: Microsoft Office Word, LibreOffice Writer, LaTeX)
- Base de dados sobre respostas de uma bactéria sobre antibióticos

A seguir são apresentados 4 tópicos: entendendo a base, preparação, geração de regras de associação e análise.

O relatório deve estar no formato pdf para submissão.

Utilizamos regras de associação para encontrar relações interessantes entre as variáveis em um conjunto de dados. Nesta prática, procuraremos relações entre as variações de expressão gênica em uma bactéria sob efeito de antibióticos. Espera-se identificar genes que diminuem ou aumentam sua expressão juntos.

1. **Entendendo a base:** A fim de entender melhor a base de dados, responda os seguintes questionamentos
 - Qual o contexto em que a base de dados se encontra?
 - O que cada instância/exemplo representa?
 - O que os atributos representam?
2. **Preparação:** Cada variável possui valores 0 e 1. Estes valores são identificados como numéricos pelo RapidMiner. Queremos que eles sejam binários para podermos aplicar as regras de associação.
 - Utilizando o operador *Numerical to Binomial* converta os valores numéricos em binominais.
3. **Geração de regras de associação:** antes de gerarmos as regras, precisamos encontrar quais são os conjuntos de itens mais frequentes.
 - Encontre os conjuntos de itens mais frequentes utilizando o operador *FP-Growth*
 - Encontre as regras de associação utilizando o operador *Create Association Rules*

4. Análise:

- Escolha duas regras de associação com itens diferentes e faça uma análise. Quais as características dos genes envolvidos? Eles estão sofrendo *upregulation* ou *down-regulation*? Que tipo de proteína esses genes expressam? O que essas regras podem significar?