**APRENDIZADO DE MÁQUINA**

***Curso de especialização em Bancos de Dados***

***Curso de especialização em Business Intelligence***

***Curso de especialização em Big Data***

**Prof. Hugo de Paula**

**ATIVIDADE EM LABORATÓRIO 3**

**BOOSTING**

*.*

# OBJETIVOS

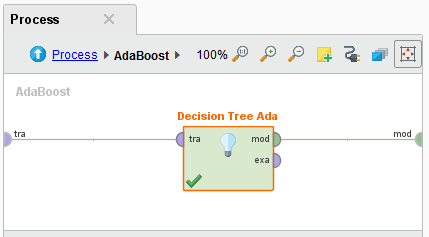
1. A base de dados SONAR foi utilizada por Gorman e Sejnowski, em seu trabalho sobre redes neurais, para classificar sinais de sonar. A tarefa consistia em treinar uma rede neural para diferenciar entre um cilindro de metal – potencial mina terrestre – de uma rocha cilíndrica.
   1. Cada registro da base possui 60 atributos numéricos entre 0 e 1.0, e uma classe.
   2. São 11 padrões para minas e 97 padrões para rocha.
2. Este trabalho irá comparar o uso de redes neurais e de técnicas de *boosting* para a classificação de sinais de sonar.

# PREPARAÇÃO DOS DADOS

1. A base se encontra na planilha ***Atividade 4 - Sonar.xlsx***.
2. Use o operador ***Set Role*** para definir o campo ***Classe*** como do tipo ***label***.

# MODELAGEM

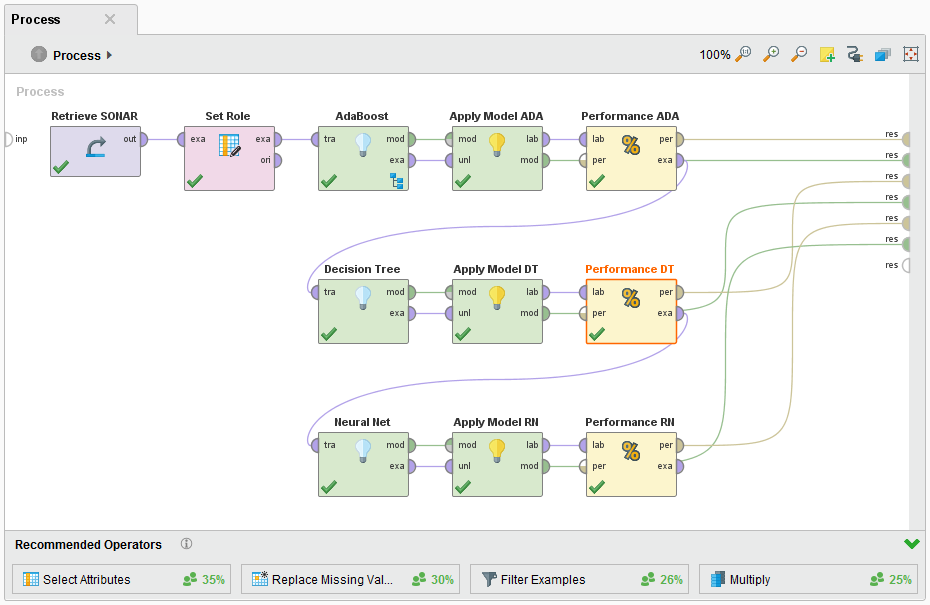
1. Vamos construir três modelos:
   1. Árvore de decisão, com o operador ***Modelling → Predictive → Trees → Decision Tree***. Pode deixar ***Parameters*** com os valores inalterados.
   2. AdaBoost, com o operador ***Modelling → Predictive → Ensembles → AdaBoost.*** Configure o parâmetro iterations do AdaBoost para 5. Esse parâmetro definirá quantos modelos serão gerados. Abra o operador AdaBoost (duplo clique) e adicione uma árvore de decisão internamente a ele.



* 1. Redes neurais, com o operador ***Modelling → Predictive → Neural Nets → Neural Net.*** Pode deixar ***Parameters*** com os valores inalterados.

# APLICANDO O MODELO E AVALIANDO A PERFORMANCE

1. Selecione o operador ***Scoring → Apply Model***. Adicione três operadores, um para cada modelo.
2. Conecte cada um dos modelos (***AdaBoost***, ***Decision Tree***¸ ***Neural Net***) na entrada ***mod*** de um operador ***Apply Model***.
3. Vamos analisar a performance dos modelos. Selecione o operador ***Validation → Performance → Performance***. Adicione três operadores, um para cada modelo.
4. Conecte cada um dos operadores ***Apply Model*** em um operador ***Performance***.
5. Conecte a base de dados em cascata nas entradas ***unl*** (***unlabelled data***) de cada operador ***Apply Model***.



# EXERCÍCIO

1. Execute o processo.
2. Explore as saídas dos modelos.
   1. Descreva a topologia da rede que foi treinada. Compare o conjunto de regras da árvore de decisão com os modelos previstos pelo AdaBoost.
   2. Verifique e compare a matriz de confusão e as medidas de performance: acurácia, precisão e revocação, de cada modelo.
   3. Analise a interpretabilidade dos modelos. Altere as opções de ***prunning*** e ***preprunning*** no AdaBoost para criar árvores mais simples, mas que tenham igual poder de previsão.
3. Escreva um documento mostrando o processo realizado, explicando os operadores e seus parâmetros, e analisando os resultados obtidos.
   1. Tente se limitar a uma ou duas páginas.
   2. Utilize imagens da tela do ***RapidMiner*** para mostrar os operadores, os dados antes e após o processamento.
   3. Este documento não precisa de capa e nem ser formatado em nenhum formato específico de trabalho acadêmico (estilo ABNT).
   4. Basta ter o título **ATIVIDADE 4**, e conter o nome do autor do trabalho.