

TRATAMENTO DE DADOS



MANOELA KOHLER

Prof.manoela@ica.ele.puc-rio.br

TÓPICOS

Análise exploratória

Pré-processamento

- Balanceamento
- Outliers
- Missing values
- Normalização
- Seleção de atributos (Filtros, Wrappers, PCA)

Associação:

- Apriori
- FP-Growth
- Eclat

Classificação:

- Regressão logística
- Support Vector Machine (SVM)
- Árvores de Decisão
- Random Forest
- Redes Neurais
- *K nearest neighbors*

Regressão

- Regressão linear simples
- Regressão linear múltipla
- Regressão não linear simples
- Regressão não linear múltipla

Agrupamento

- Particionamento (K-means, K-medoids)
- Hierárquico (DIANA, AGNES)
- Densidade (DBSCAN)

Séries Temporais

- Naive
- Média Móvel
- Amortecimento exponencial
- Auto-regressivo integrados de média móvel
- Auto regressivo não linear

Recapitulação

DEFINIÇÃO E CONCEITOS:

- Quando se procura DM no Google Images, tem-se:



DEFINIÇÃO E CONCEITOS: INTERDISCIPLINAR

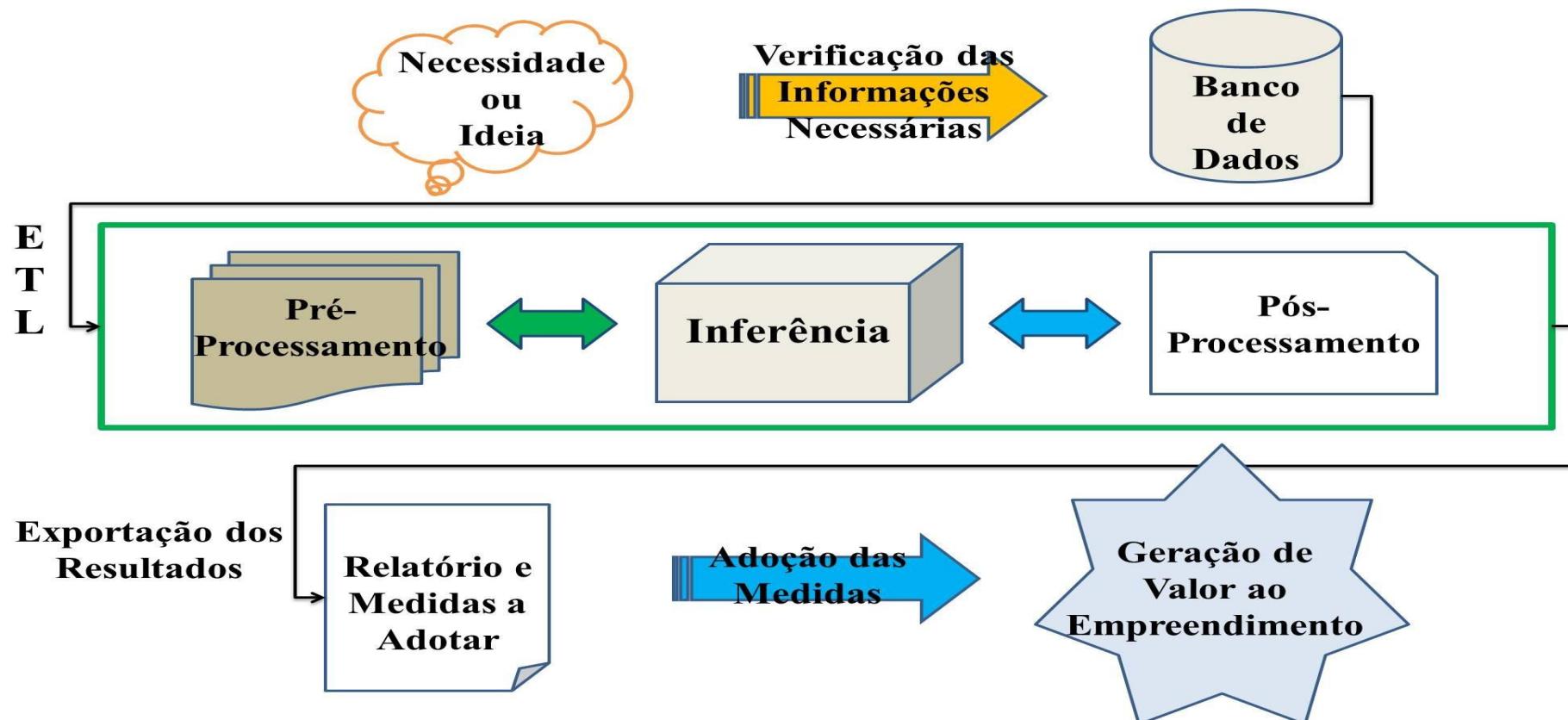


5 CLASSES DE PROBLEMAS DE DM



ETAPAS DE UM PROJETO DE DATA MINING

ESQUEMA BÁSICO DE UM PROJETO DE DM



Análise Exploratória de Dados

Tratamento de Dados

Missing Values

MISSING VALUES

- Muito comum no mundo real;
- Atributos novos que surgem com o passar do tempo e com a necessidade de novas informações por parte das empresas;
- Atributos não preenchidos por falta de obrigatoriedade;

OBS: Verificar se o número de registros com dados faltantes (para um atributo específico) por si só não justifica a remoção do atributo;

SUBSTITUIÇÃO PELA MÉDIA

Idade	Estado Civil	Nota	Atrito
35	Casado	9	Sim
53	?	7	Sim
68	Casado	10	Não
20	Solteiro	8	Não
29	Casado	?	Sim

$$\text{Média: } 9 + 7 + 10 + 8 / 4 = 8.5$$

SUBSTITUIÇÃO PELA MÉDIA

Idade	Estado Civil	Nota	Atrito
35	Casado	9	Sim
53	?	7	Sim
68	Casado	10	Não
20	Solteiro	8	Não
29	Casado	8.5	Sim

SUBSTITUIÇÃO PELA MÉDIA POR CLASSE

Idade	Estado Civil	Nota	Atrito
35	Casado	9	Sim
53	?	7	Sim
68	Casado	10	Não
20	Solteiro	8	Não
29	Casado	?	Sim

$$\text{Média: } 9 + 7 / 2 = 8$$

SUBSTITUIÇÃO PELA MÉDIA POR CLASSE

Idade	Estado Civil	Nota	Atrito
35	Casado	9	Sim
53	?	7	Sim
68	Casado	10	Não
20	Solteiro	8	Não
29	Casado	8	Sim

SUBSTITUIÇÃO PELO VALOR MAIS FREQUENTE

Idade	Estado Civil	Nota	Atrito
35	Casado	9	Sim
53	?	7	Sim
68	Casado	10	Não
20	Solteiro	8	Não
29	Casado	8	Sim

SUBSTITUIÇÃO PELO VALOR MAIS FREQUENTE

Idade	Estado Civil	Nota	Atrito
35	Casado	9	Sim
53	Casado	7	Sim
68	Casado	10	Não
20	Solteiro	8	Não
29	Casado	8	Sim

Normalização

NORMALIZAÇÃO

- Dar aos atributos pesos iguais;
- Diminuir o tempo de convergência dos algoritmos;

Redução de Dimensionalidade

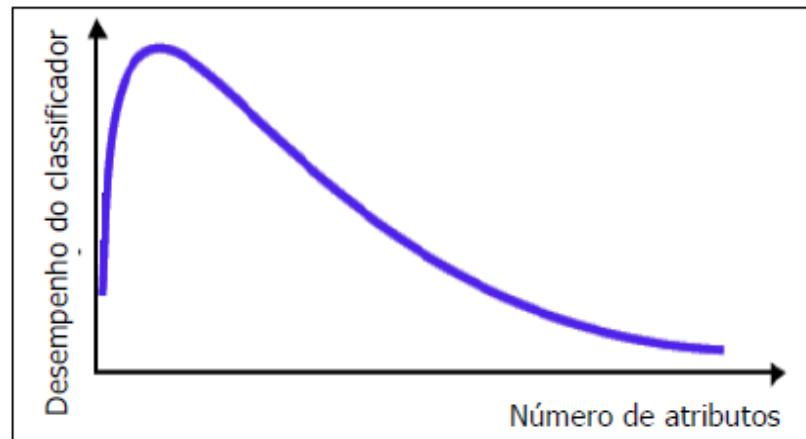
REDUÇÃO DE DIMENSIONALIDADE

Maldição da dimensionalidade (Curse of dimensionality)

- Termo que se refere a vários fenômenos que surgem na análise de dados em espaços com muitas dimensões (atributos);
- Muitas vezes com centenas ou milhares de dimensões;
- Basicamente, adicionar características não significa sempre melhora no desempenho de um classificador.

REDUÇÃO DE DIMENSIONALIDADE

- De modo geral, o desempenho de um classificador tende a se degradar a partir de um determinado nº de atributos
 - Mesmo que eles sejam atributos úteis



REDUÇÃO DE DIMENSIONALIDADE

Tarefas:

- Redução do número de atributos.

Objetivos

- Diminuir o custo do aprendizado;
- Aumentar a precisão do algoritmo;
- Gerar modelos compactos mais fáceis de interpretar.

REDUÇÃO DE DIMENSIONALIDADE

Em geral, espera-se que todos os atributos sejam relevantes porém nem sempre é possível garantir isso.

Além disso, alguns atributos são redundantes e assim poderiam ser eliminados.

Objetivo:

- Definir conjunto de atributos que sejam relevantes e não-redundantes

REDUÇÃO DE DIMENSIONALIDADE

➤ Abordagens

- Seleção X Agregação de Atributos

➤ Seleção de atributos:

- Escolha de um sub-conjunto de atributos relevantes dentre os atributos disponíveis
 - E.g., Filtros e Wrappers

➤ Agregação de atributos:

- Criação de novos atributos a partir da combinação dos atributos existentes
 - E.g., PCA

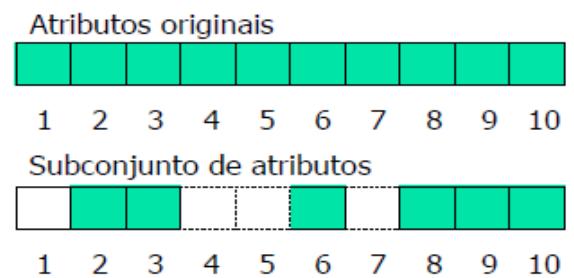
Seleção de Atributos

SELEÇÃO DE ATRIBUTOS

Filtros

Os métodos de seleção de filtro aplicam uma medida estatística para atribuir uma pontuação a cada atributo.

Os atributos são ranqueados pela pontuação e então podem ser selecionados para serem mantidos ou removidos do conjunto de dados.

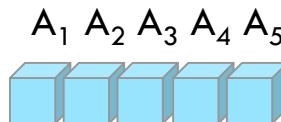


SELEÇÃO DE ATRIBUTOS

Wrapper

Os métodos Wrapper consideram a seleção de um conjunto de recursos como um **problema de busca**, onde diferentes combinações são preparadas, avaliadas e comparadas com outras combinações. Um modelo preditivo é usado para avaliar uma combinação de características e atribuir uma pontuação baseada na precisão do modelo.

Suponha uma base com 5 atributos:



Cromossomo: cada gene pode assumir um de dois valores: 0 ou 1.

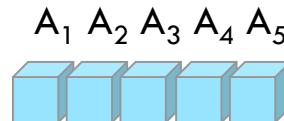
0 = Sem o respectivo atributo;
1 = Com o respectivo atributo;

SELEÇÃO DE ATRIBUTOS

Wrapper

Os métodos Wrapper consideram a seleção de um conjunto de recursos como um **problema de busca**, onde diferentes combinações são preparadas, avaliadas e comparadas com outras combinações. Um modelo preditivo é usado para avaliar uma combinação de características e atribuir uma pontuação baseada na precisão do modelo.

Suponha uma base com 5 atributos:



Cada indivíduo criado durante a evolução do AG é apresentado ao algoritmo classificador.

SELEÇÃO DE ATRIBUTOS

Wrapper

Os métodos Wrapper consideram a seleção de um conjunto de recursos como um **problema de busca**, onde diferentes combinações são preparadas, avaliadas e comparadas com outras combinações. Um modelo preditivo é usado para avaliar uma combinação de características e atribuir uma pontuação baseada na precisão do modelo.

Suponha uma base com 5 atributos:

A ₁	A ₂	A ₃	A ₄	A ₅
0	1	1	0	1

Toda a base com os três atributos escolhidos (A_2 , A_3 e A_5) é usada para treinar o algoritmo classificador e a função de avaliação pode ser, por exemplo, a acurácia de treinamento.



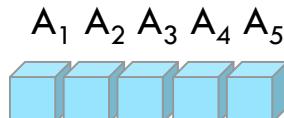
Acurácia =
30%

SELEÇÃO DE ATRIBUTOS

Wrapper

Os métodos Wrapper consideram a seleção de um conjunto de recursos como um **problema de busca**, onde diferentes combinações são preparadas, avaliadas e comparadas com outras combinações. Um modelo preditivo é usado para avaliar uma combinação de características e atribuir uma pontuação baseada na precisão do modelo.

Suponha uma base com 5 atributos:



O algoritmo genético evolui até chegar à resposta ótima.

Acurácia =
93%



11001

SELEÇÃO DE ATRIBUTOS

- Embarcados (Embedded)
 - Processo de seleção faz parte do algoritmo de aprendizado
 - Exemplo: Árvore de Decisão

Agregação de Atributos

AGREGAÇÃO DE ATRIBUTOS

- Exemplo: dois atributos → “massa” e “volume”
- Esses atributos podem ser agregados em um único atributo: “densidade”
 - $\text{densidade} = \text{massa} / \text{volume}$
- Nesse caso, não há perda de informação.

AGREGAÇÃO DE ATRIBUTOS

PCA

Consiste em transformar um conjunto de variáveis originais em outro conjunto de variáveis de mesma dimensão denominadas de componentes principais.

Os componentes principais apresentam propriedades importantes:

- Cada componente principal é uma combinação linear de todas as variáveis originais;
- Todos os componentes são ortogonais entre si, portanto não há informações redundantes;
- São estimados com o propósito de reter, em ordem de estimação, o máximo de informação, em termos da variação total contida nos dados.

Redução de massa de dados, com menor perda possível da informação.

Análise de componentes principais é completamente reversível, tornando-a uma ferramenta versátil e útil para redução de dados, compressão de dados, entre outros.

Component	Proportion of Variance	Cumulative Variance
PC 1	0.121	0.121
PC 2	0.078	0.199
PC 3	0.069	0.268
PC 4	0.062	0.330
PC 5	0.054	0.384
PC 6	0.048	0.431
PC 7	0.042	0.474
PC 8	0.039	0.512
PC 9	0.035	0.547
PC 10	0.034	0.581
PC 11	0.028	0.609
PC 12	0.026	0.635
PC 13	0.025	0.660
PC 14	0.024	0.685
PC 15	0.024	0.708
PC 16	0.023	0.731
PC 17	0.020	0.751
PC 18	0.019	0.770
PC 19	0.018	0.788
PC 20	0.018	0.805
PC 21	0.017	0.822
PC 22	0.016	0.838

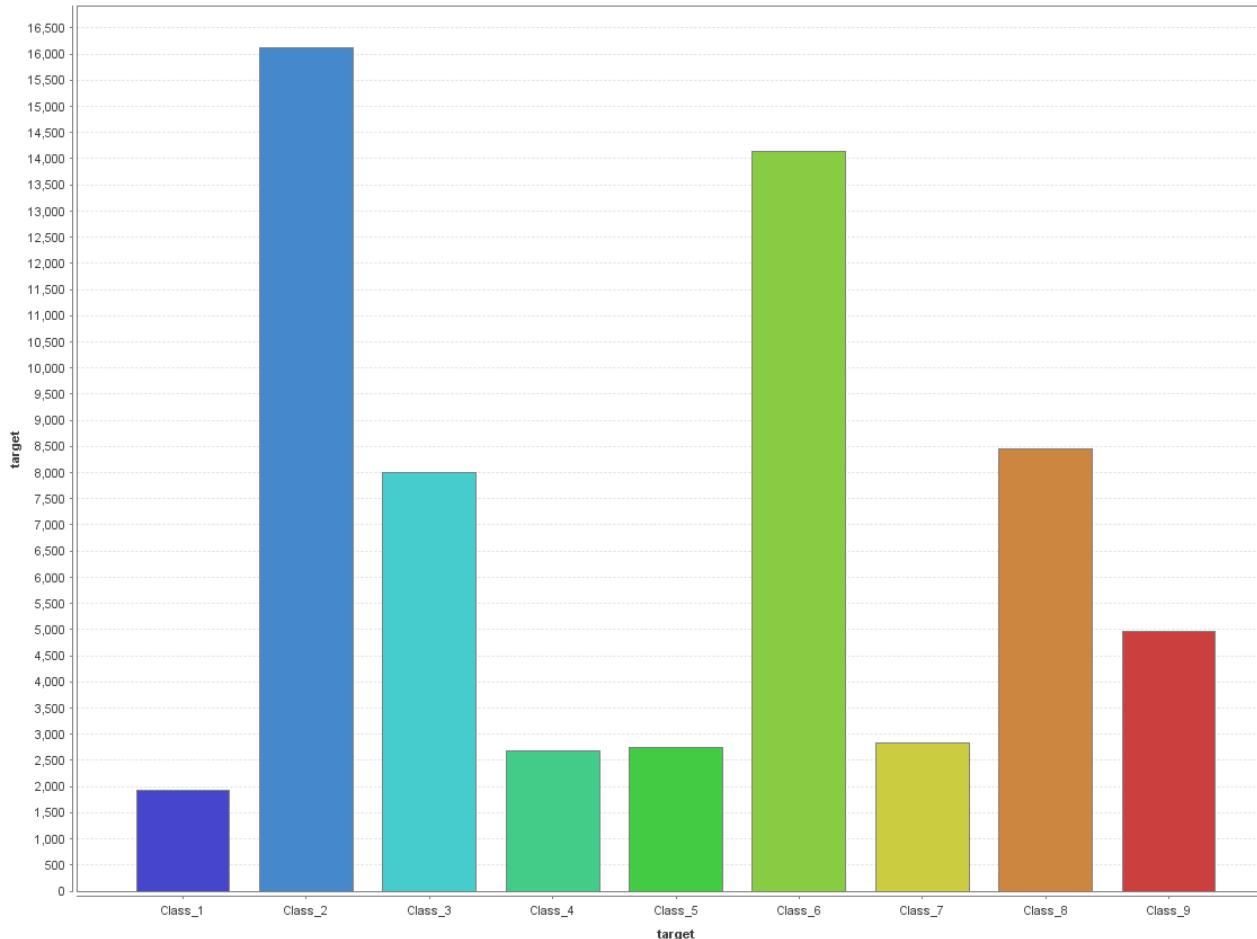
Component	Proportion of Variance	Cumulative Variance
PC 24	0.015	0.809
PC 25	0.014	0.883
PC 26	0.013	0.897
PC 27	0.012	0.908
PC 28	0.011	0.920
PC 29	0.011	0.931
PC 30	0.010	0.941
PC 31	0.010	0.951
PC 32	0.010	0.961
PC 33	0.008	0.969
PC 34	0.007	0.976
PC 35	0.006	0.982
PC 36		
PC 37		
PC 38		
PC 39		
PC 40		
PC 41		

Limiar de variância = 0.95

Resulta em 31 componentes principais
de um total de 49 atributos

Balanceamento de Dados

BALANCEAMENTO DE DADOS

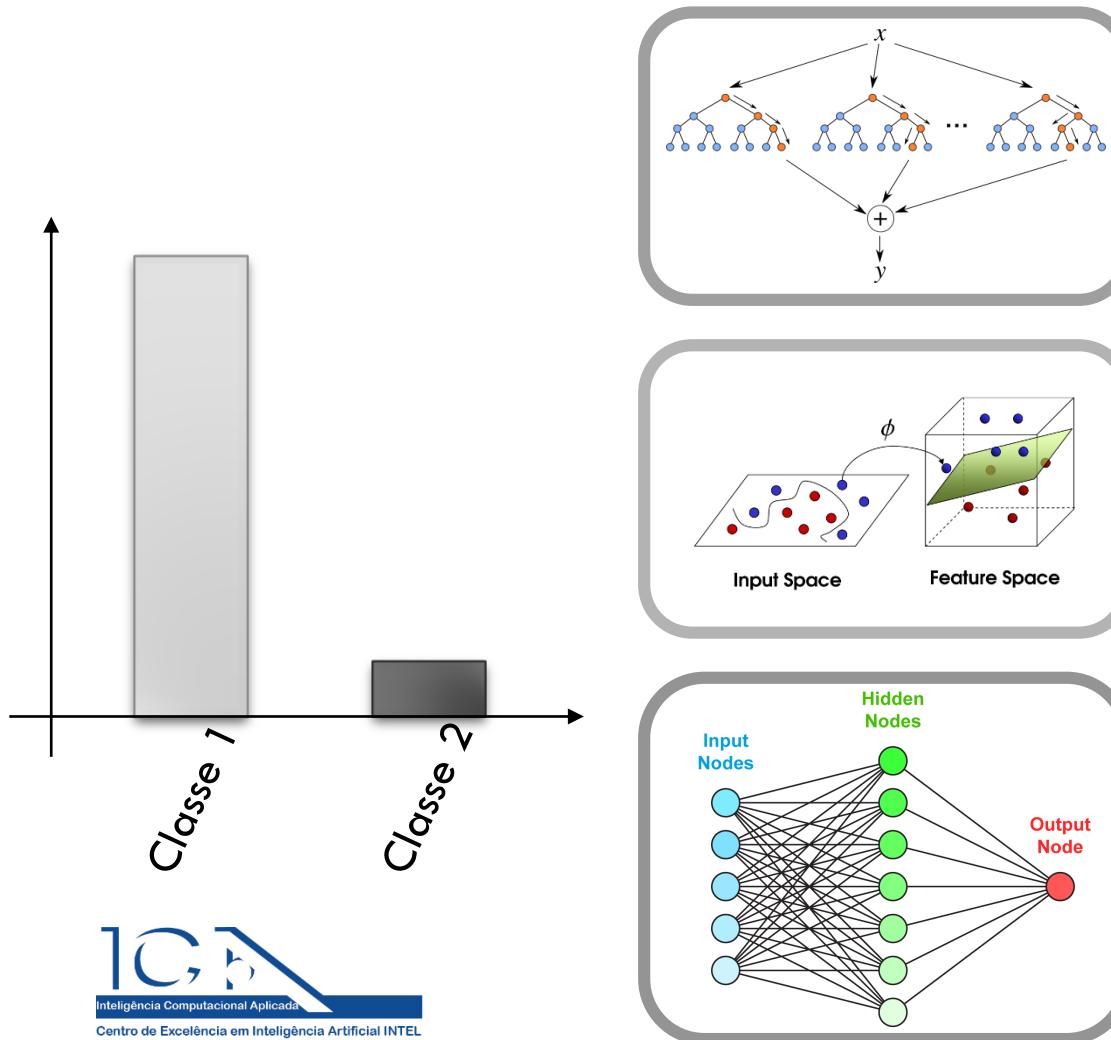


BALANCEAMENTO DE DADOS

A grande maioria das bases de dados não têm o mesmo número de instâncias em cada classe, mas quando essa diferença é pequena não há problemas.

Em algumas bases, esse comportamento é esperado. Por exemplo, em bases de dados que representam transações fraudulentas. A maioria das transações serão “Normais”, e uma pequena minoria de transações serão “Fraudulentas”.

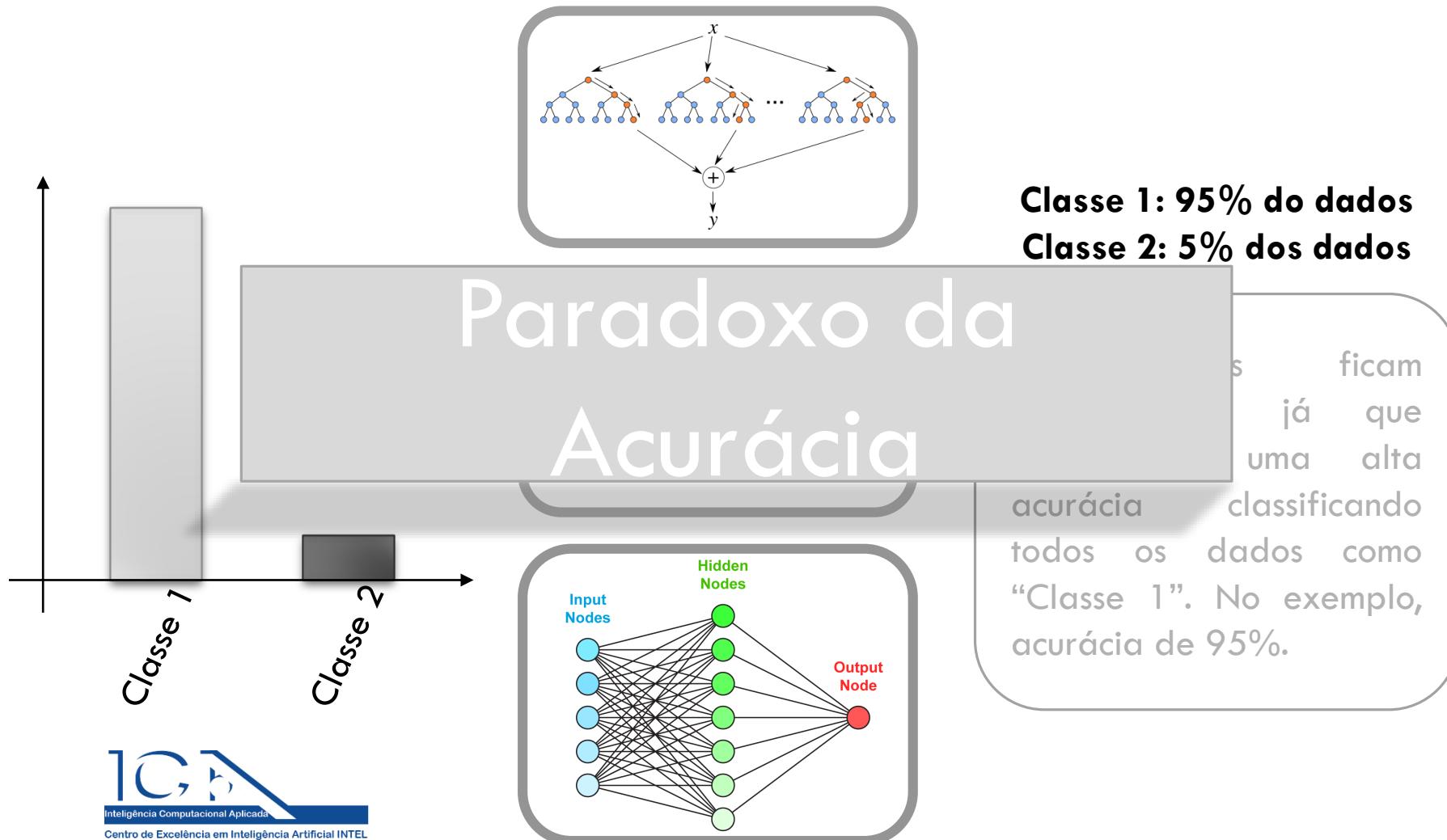
BALANCEAMENTO DE DADOS



Classe 1: 95% do dados
Classe 2: 5% dos dados

Classificadores ficam “preguiçosos”, já que consegue-se uma alta acurácia classificando todos os dados como “Classe 1”. No exemplo, acurácia de 95%.

BALANCEAMENTO DE DADOS



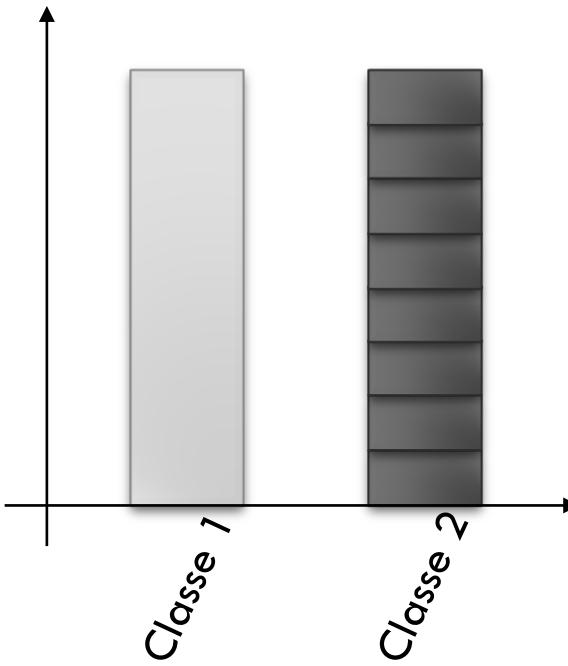
BALANCEAMENTO DE DADOS

1. Coletar mais dados
2. Mudar a métrica de performance

- Matriz de confusão
- Precisão
- Recall
- Kappa

BALANCEAMENTO DE DADOS

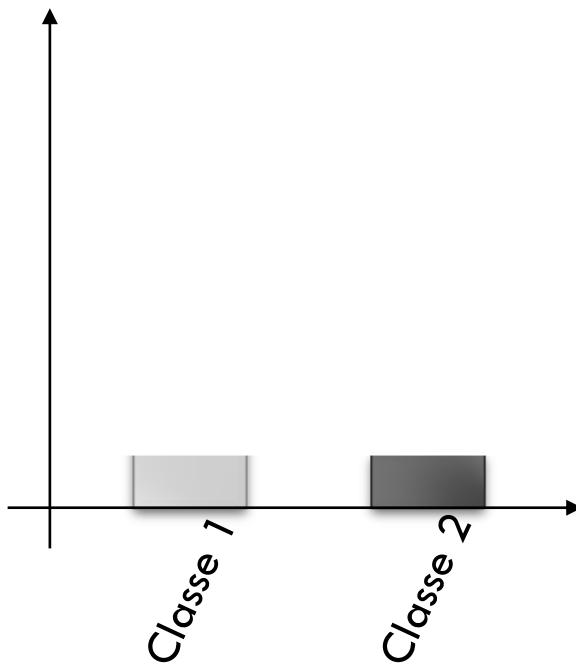
3. Reamostrar conjunto de dados
 - Over-sampling (aleatório)



BALANCEAMENTO DE DADOS

3. Reamostrar conjunto de dados

- Over-sampling (aleatório)
- Under-sampling (aleatório)

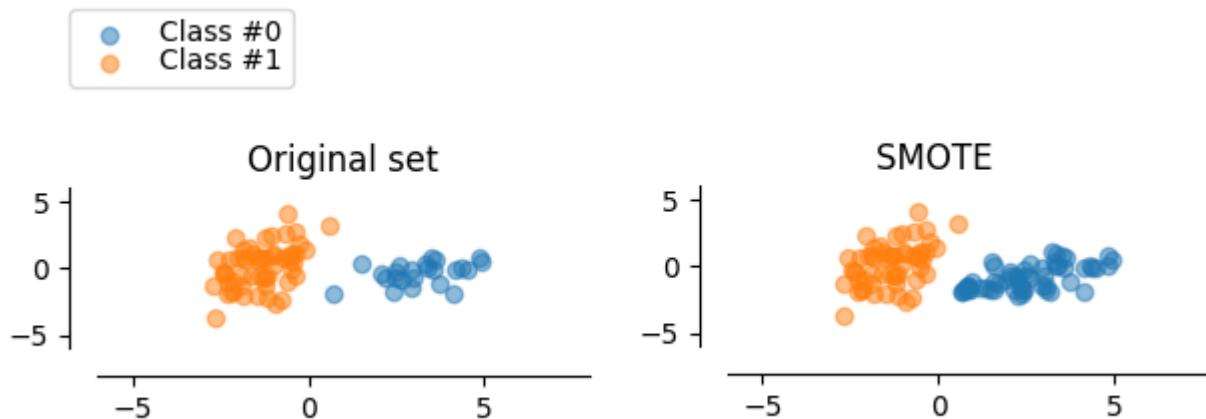


BALANCEAMENTO DE DADOS

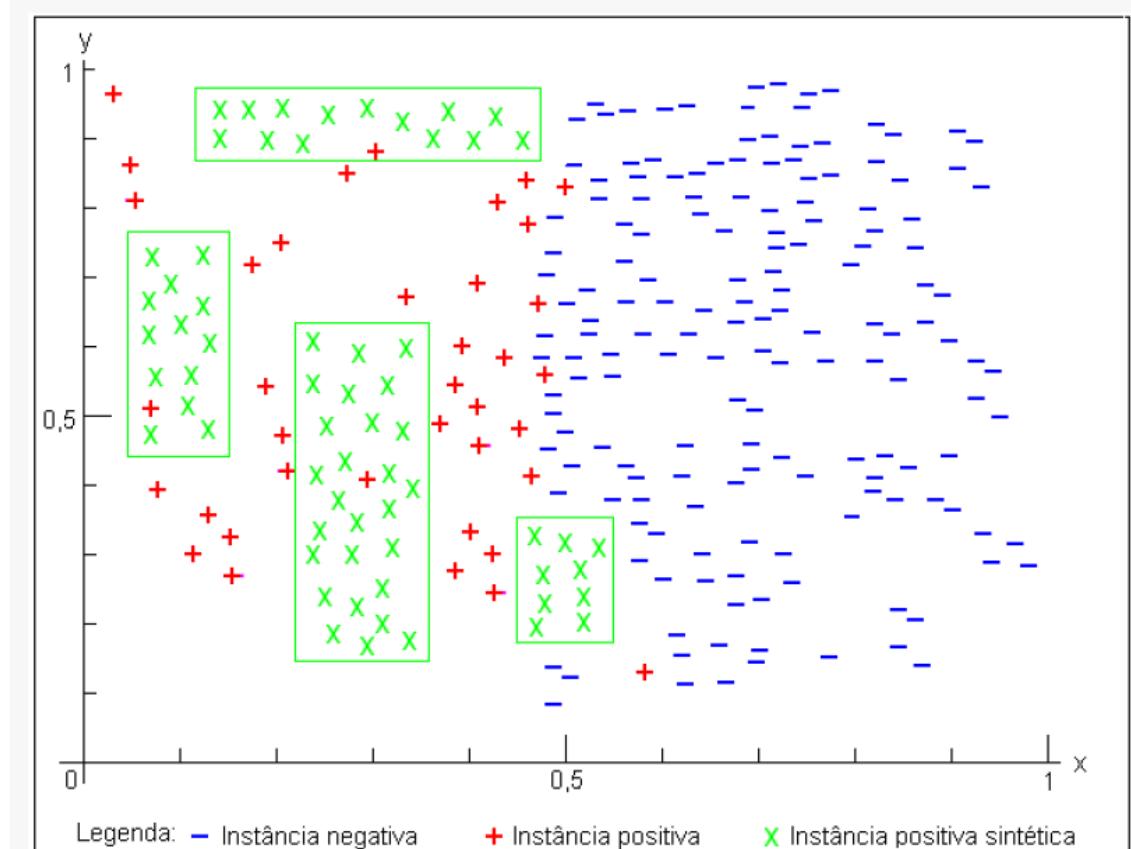
4. Gerar amostras aleatórias

Exemplo: SMOTE (Synthetic Minority Over-Sampling Technique)

Método de over-sampling que gera amostras sintéticas da classe rara (ao invés de criar cópias) perturbando um atributo por vez por um valor dentro da diferença entre os exemplos vizinhos.



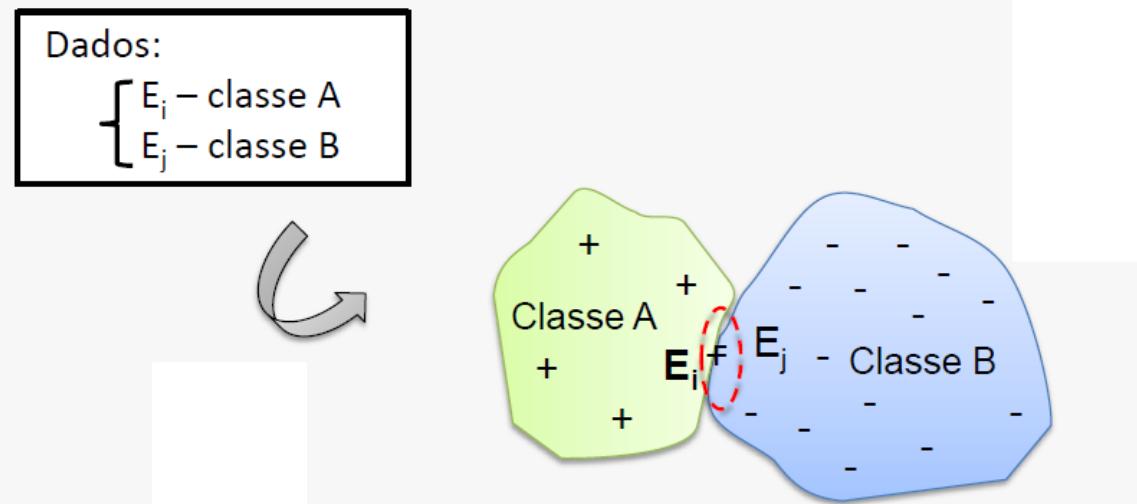
BALANCEAMENTO DE DADOS



Aplicação de regiões sintéticas de forma a balancear o conjunto de dados.

BALANCEAMENTO DE DADOS

- *Tomek Links* (Tomek, 1976)
 - Remove ruídos e pontos de fronteira;
 - Pode ser usado como
 - under-sampling (remove o ponto da classe predominante (-))
 - data cleaning (remove os dois pontos)

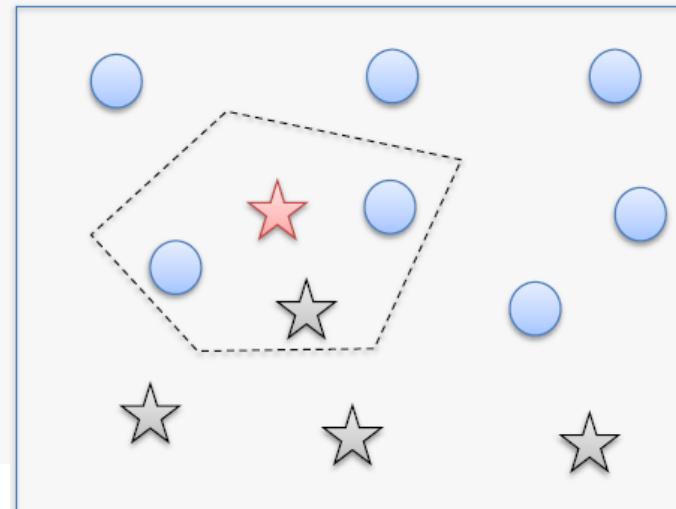


BALANCEAMENTO DE DADOS

Edited Nearest Neighbor (ENN) (Wilson, 1972)

Cleaning:

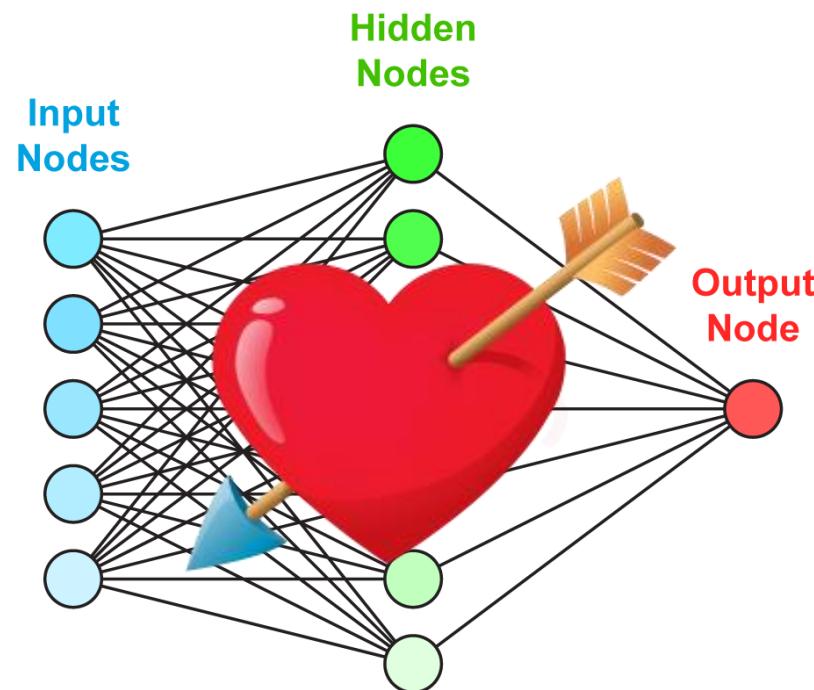
- dado um exemplo, procura os três vizinhos mais próximos
- Se o exemplo difere da classe de 2 dos 3 exemplos, o mesmo é removido.



BALANCEAMENTO DE DADOS

5. Testar em diferentes algoritmos

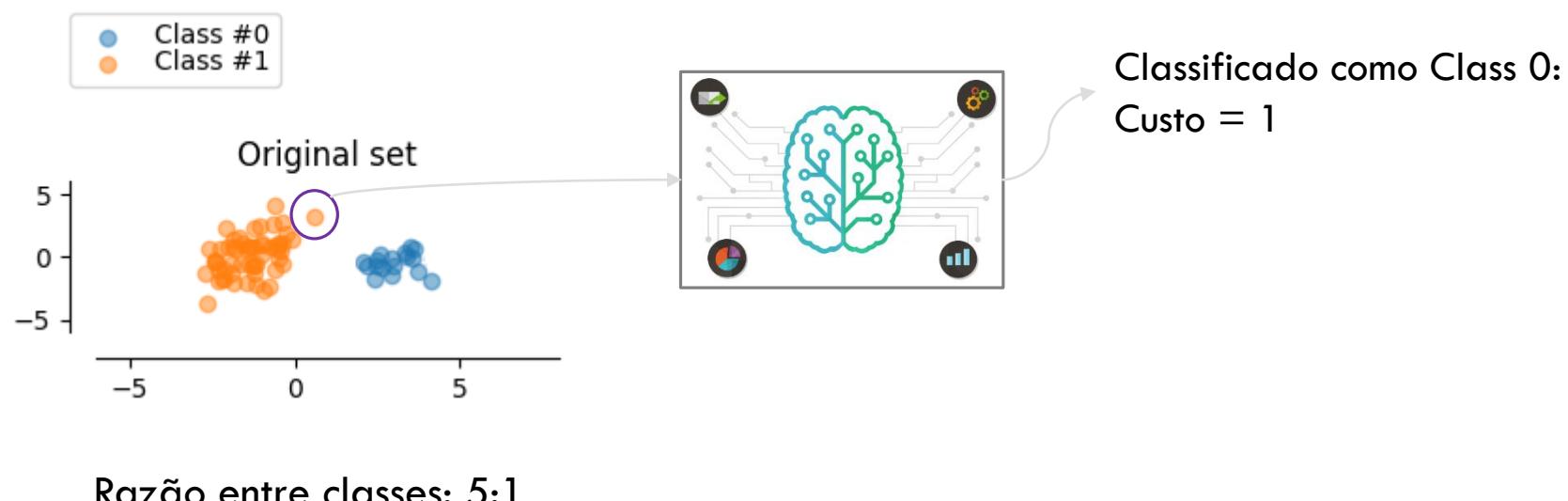
Não se deve ter um algoritmo favorito nesse caso. Tente diferentes algoritmos.



BALANCEAMENTO DE DADOS

6. Penalização

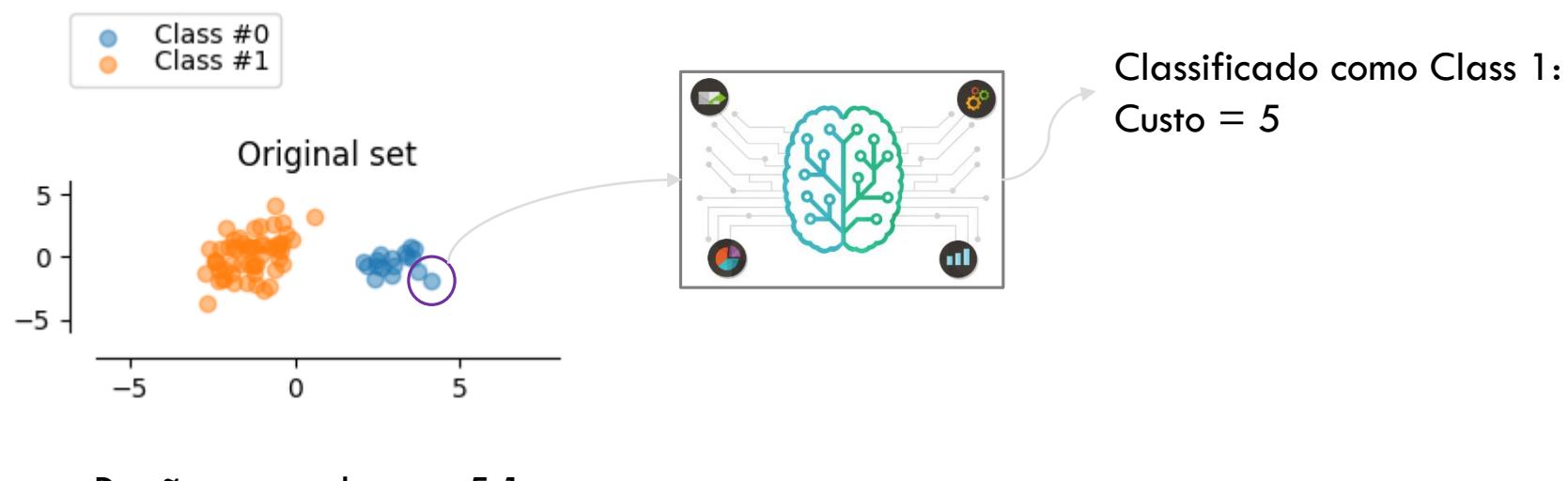
Custo adicional para classificação errada da classe rara.



BALANCEAMENTO DE DADOS

6. Penalização

Custo adicional para classificação errada da classe rara.



Outliers

OUTLIERS

Algoritmos de aprendizado de máquina são sensíveis à distribuição e intervalo dos dados. Outliers podem enviesar e induzir ao erro, resultando em um tempo de treinamento maior, menor acurácia e modelos piores.



1. Análise de valores extremos;
2. Métodos de proximidade
(e.g. k-means);
1. Métodos de projeção
(e.g. kohonen).

ESTUDO DE CASO

<http://archive.ics.uci.edu/ml/datasets/SECOM>

Um complexo processo de fabricação de semi-condutores moderno é normalmente feito sob vigilância consistente através do monitoramento de sinais, variáveis coletadas de sensores e pontos de medição de processo.

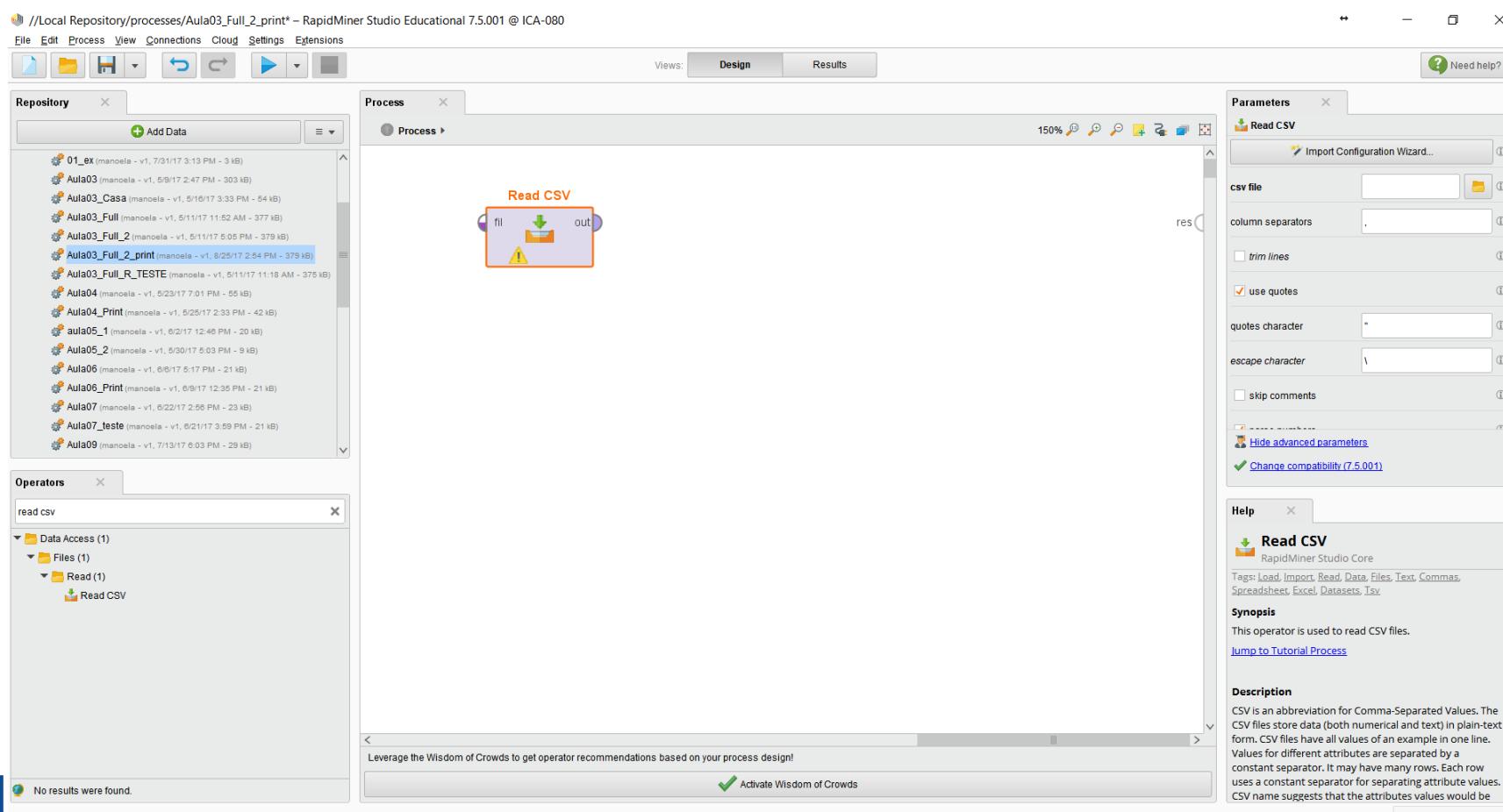
No entanto, nem todos estes sinais são igualmente valiosos num sistema de monitoramento específico. Os sinais medidos contêm uma combinação de informações úteis, informações irrelevantes, bem como ruído.

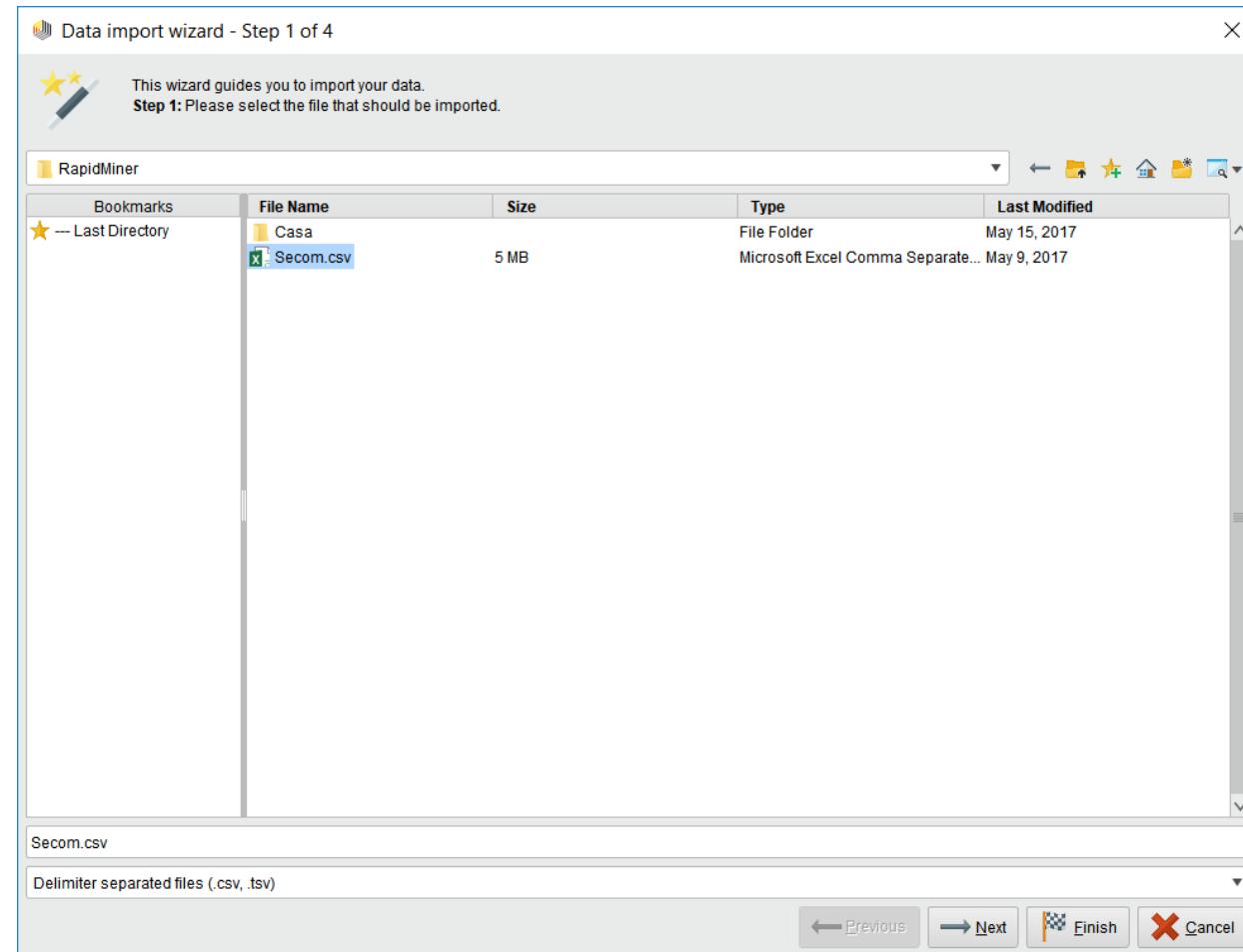
Engenheiros tipicamente têm um número muito maior de sinais do que são realmente necessários. Se considerarmos cada tipo de sinal como uma característica, então a seleção da característica pode ser aplicada para identificar os sinais mais relevantes.

Os Engenheiros de Processo podem então usar estes sinais para determinar os fatores chave para o processo de produção. Isso poderá trazer benefícios para o processo, como redução do tempo e diminuição nos custos de produção.

Base de dados SECOM: 1567 exemplos e 591 atributos

IMPORTAR A BASE





 Data import wizard - Step 2 of 4

This wizard guides you to import your data.
Step 2: Please specify how the file should be parsed and how columns are separated.

File Reading

File Encoding: windows-1252

Trim Lines

Skip Comments: #

Column Separation

Comma "," Space
 Semicolon ";" Tab
 Regular Expression: ,\S*\|\S*

Escape Character: \

Use Quotes: "

3030.93	2564	2187.7333	1411.1265	1.3602	100	97.6133	0.1242	1.5005	0.0162	-0.0034	0.9455	202.438
3095.78	2465.14	2230.4222	1463.6606	0.8294	100	102.3433	0.1247	1.4966	-5.00E-04	-0.0148	0.9627	200.547
2932.61	2559.94	2186.4111	1698.0172	1.5102	100	95.4878	0.1241	1.4436	0.0041	0.0013	0.9615	202.017
2988.72	2479.9	2199.0333	909.7926	1.3204	100	104.2367	0.1217	1.4882	-0.0124	-0.0033	0.9629	201.848
3032.24	2502.87	2233.3667	1326.52	1.5334	100	100.3967	0.1235	1.5031	-0.0031	-0.0072	0.9569	201.942
2946.25	2432.84	2233.3667	1326.52	1.5334	100	100.3967	0.1235	1.5287	0.0167	0.0055	0.9699	200.472
3030.27	2430.12	2230.4222	1463.6606	0.8294	100	102.3433	0.1247	1.5816	-0.027	0.0105	0.9591	202.090
3058.88	2690.15	2248.9	1004.4692	0.7884	100	106.24	0.1185	1.5153	0.0157	7.00E-04	0.9481	202.417
2967.68	2600.47	2248.9	1004.4692	0.7884	100	106.24	0.1185	1.5358	0.0111	-0.0066	0.9494	202.454
3016.11	2428.37	2248.9	1004.4692	0.7884	100	106.24	0.1185	1.5381	0.0159	0.0049	0.944	202.599
2994.05	2548.21	2195.1222	1046.1468	1.3204	100	103.34	0.1223	1.5144	-0.019	0.0013	0.9433	201.712
2928.84	2479.4	2196.2111	1605.7578	0.9959	100	97.9156	0.1257	1.469	0.017	-0.0154	0.9445	202.126
2900.27	2557.1	2185.1022	1610.1100	1.0001	100	102.34	0.1200	1.501	0.0050	0.0010	0.9505	200.100

Row, Column **Error** **Original value** **Message**

Previous Next Finish Cancel

 Data import wizard - Step 3 of 4 X

 This wizard guides you to import your data.

Step 3: In RapidMiner Studio, each attribute can be annotated. The most important annotation of an attribute is its name - a row with this annotation defines the names of the attributes. If your data does not contain attribute names, do not set this property. If further annotations are contained in the rows of your data file, you can assign them here.

Annotat...	att1	att2	att3	att4	att5	att6	att7	att8	att9	att10	att11	att12
-	3030.93	2564	2187.7333	1411.1265	1.3602	100	97.6133	0.1242	1.5005	0.0162	-0.0034	0.9455
-	3095.78	2465.14	2230.4222	1463.6606	0.8294	100	102.3433	0.1247	1.4966	-5.00E-04	-0.0148	0.9627
-	2932.61	2559.94	2186.4111	1698.0172	1.5102	100	95.4878	0.1241	1.4436	0.0041	0.0013	0.9615
-	2988.72	2479.9	2199.0333	909.7926	1.3204	100	104.2367	0.1217	1.4882	-0.0124	-0.0033	0.9629
-	3032.24	2502.87	2233.3667	1326.52	1.5334	100	100.3967	0.1235	1.5031	-0.0031	-0.0072	0.9569
-	2946.25	2432.84	2233.3667	1326.52	1.5334	100	100.3967	0.1235	1.5287	0.0167	0.0055	0.9699
-	3030.27	2430.12	2230.4222	1463.6606	0.8294	100	102.3433	0.1247	1.5816	-0.027	0.0105	0.9591
-	3058.88	2690.15	2248.9	1004.4692	0.7884	100	106.24	0.1185	1.5153	0.0157	7.00E-04	0.9481
-	2967.68	2600.47	2248.9	1004.4692	0.7884	100	106.24	0.1185	1.5358	0.0111	-0.0066	0.9494
-	3016.11	2428.37	2248.9	1004.4692	0.7884	100	106.24	0.1185	1.5381	0.0159	0.0049	0.944
-	2994.05	2548.21	2195.1222	1046.1468	1.3204	100	103.34	0.1223	1.5144	-0.019	0.0013	0.9433
-	2928.84	2479.4	2196.2111	1605.7578	0.9959	100	97.9156	0.1257	1.469	0.017	-0.0154	0.9445
-	2920.07	2507.4	2195.1222	1046.1468	1.3204	100	103.34	0.1223	1.531	-0.0259	0.0216	0.9595
-	3051.44	2529.27	2184.4333	877.6266	1.4668	100	107.8711	0.124	1.5236	-0.0209	-0.0031	0.9441
-	2963.97	2629.48	2224.6222	947.7739	1.2924	100	104.8489	0.1197	1.4474	0.0144	-0.0119	0.9582
-	2988.31	2546.26	2224.6222	947.7739	1.2924	100	104.8489	0.1197	1.5465	0.025	-0.0024	0.9616
-	3028.02	2560.87	2270.2556	1258.4558	1.395	100	104.8078	0.1207	1.4368	0.015	-0.0037	0.9623
-	3032.73	2517.79	2270.2556	1258.4558	1.395	100	104.8078	0.1207	1.5537	0.022	-0.0027	0.9613
-	3040.34	2501.16	2207.3889	962.5317	1.2043	100	104.0311	0.121	1.5481	-0.0367	0.0014	0.9634
-	2988.3	2519.05	2208.8556	1157.7224	1.5509	100	107.8022	0.1233	1.5362	-0.0259	-0.0179	0.9614
-	2987.32	2528.81						0.1195	1.6343	-0.0263	0.0116	0.9587

<
>
 Previous
 Next
 Finish
 Cancel

 Data import wizard - Step 4 of 4

This wizard guides you to import your data.
Step 4: RapidMiner Studio uses strongly typed attributes. In this step, you can define the data types of your attributes. Furthermore, RapidMiner Studio assigns roles to the attributes, defining what they can be used for by the individual operators. These roles can be also defined here. Finally, you can rename attributes or deselect them entirely.

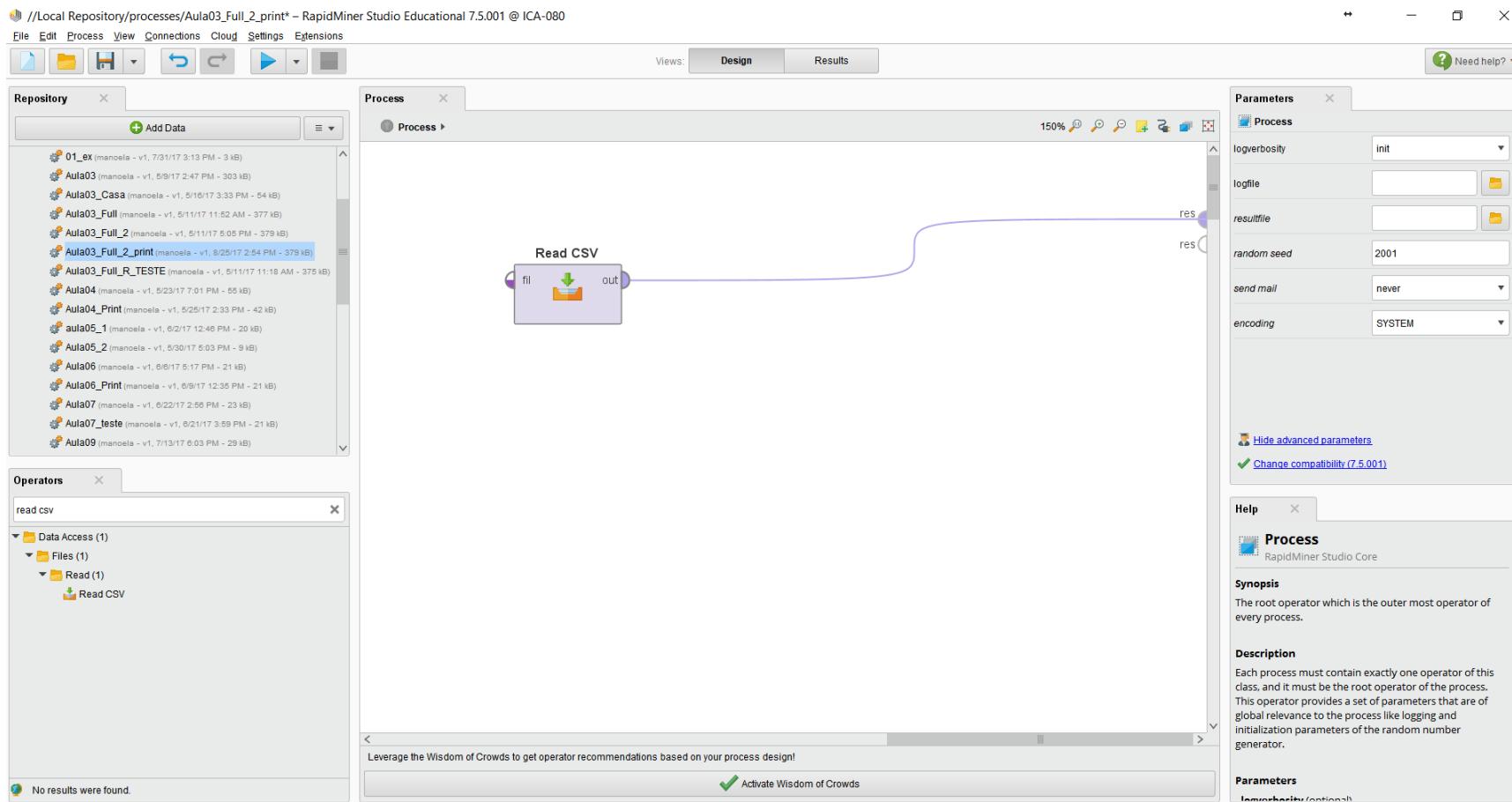
Reload data Date format

Preview uses only first 100 rows.

	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
79	att580	att581	att582	att583	att584	att585	att586	att587	att588	att589	att590	att591
attribute	real	real	real	real	real	real	real	real	real	real	real	binomi...
	?	?	?	0.500	0.012	0.004	2.363	?	?	?	?	-1
010	0.020	0.006	208.204	0.502	0.022	0.005	4.445	0.010	0.020	0.006	208.204	-1
058	0.048	0.015	82.860	0.496	0.016	0.004	3.175	0.058	0.048	0.015	82.860	1
020	0.015	0.004	73.843	0.499	0.010	0.003	2.054	0.020	0.015	0.004	73.843	-1
?	?	?	0.480	0.477	0.104	99.303	0.020	0.015	0.004	73.843	-1	
034	0.015	0.005	44.008	0.495	0.019	0.004	3.828	0.034	0.015	0.005	44.008	-1
?	?	?	0.501	0.014	0.004	2.852	0.034	0.015	0.005	44.008	-1	
020	0.019	0.006	95.031	0.498	0.011	0.003	2.126	0.020	0.019	0.006	95.031	-1
011	0.012	0.004	111.653	0.499	0.017	0.005	3.446	0.011	0.012	0.004	111.653	-1
021	0.019	0.007	90.229	0.497	0.015	0.004	3.069	0.021	0.019	0.007	90.229	-1
035	0.021	0.007	57.812	0.492	0.016	0.004	3.212	0.035	0.021	0.007	57.812	1

<  0 errors. Ignore errors Show only errors >

Row, Column	Error	Original value	Message



//Local Repository/processes/Aula03_Full_2_print* – RapidMiner Studio Educational 7.5.001 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History ExampleSet (Read CSV)

Repository Add Data

ExampleSet (1567 examples, 1 special attribute, 590 regular attributes)

Filter (1567 / 1,567 examples): all

Row No.	att591	att1	att2	att3	att4	att5	att6	att7	att8	att9	att10	att11	att12
1	-1	3030.930	2564	2187.733	1411.127	1.360	100	97.613	0.124	1.500	0.016	-0.003	0.946
2	-1	3095.780	2465.140	2230.422	1463.661	0.829	100	102.343	0.125	1.497	-0.000	-0.015	0.963
3	1	2932.610	2559.940	2186.411	1698.017	1.510	100	95.488	0.124	1.444	0.004	0.001	0.962
4	-1	2988.720	2479.900	2199.033	909.793	1.320	100	104.237	0.122	1.488	-0.012	-0.003	0.963
5	-1	3032.240	2502.870	2233.367	1326.520	1.533	100	100.397	0.123	1.503	-0.003	-0.007	0.957
6	-1	2948.250	2432.840	2233.367	1326.520	1.533	100	100.397	0.123	1.529	0.017	0.005	0.970
7	-1	3030.270	2430.120	2230.422	1463.661	0.829	100	102.343	0.125	1.582	-0.027	0.011	0.959
8	-1	3058.880	2690.150	2248.900	1004.469	0.788	100	106.240	0.118	1.515	0.016	0.001	0.948
9	-1	2967.680	2600.470	2248.900	1004.469	0.788	100	106.240	0.118	1.536	0.011	-0.007	0.949
10	-1	3016.110	2428.370	2248.900	1004.469	0.788	100	106.240	0.118	1.538	0.016	0.005	0.944
11	1	2994.050	2548.210	2195.122	1046.147	1.320	100	103.340	0.122	1.514	-0.019	0.001	0.943
12	1	2928.840	2479.400	2196.211	1605.758	0.996	100	97.916	0.126	1.469	0.017	-0.015	0.945
13	-1	2920.070	2507.400	2195.122	1046.147	1.320	100	103.340	0.122	1.531	-0.026	0.022	0.960
14	-1	3051.440	2529.270	2184.433	877.627	1.467	100	107.871	0.124	1.524	-0.021	-0.003	0.944
15	1	2963.970	2629.480	2224.622	947.774	1.292	100	104.849	0.120	1.447	0.014	-0.012	0.958
16	-1	2988.310	2546.260	2224.622	947.774	1.292	100	104.849	0.120	1.546	0.025	-0.002	0.962
17	-1	3028.020	2560.870	2270.256	1258.456	1.395	100	104.808	0.121	1.437	0.015	-0.004	0.962
18	-1	3032.730	2517.790	2270.256	1258.456	1.395	100	104.808	0.121	1.554	0.022	-0.003	0.961
19	-1	3040.340	2501.160	2207.389	962.532	1.204	100	104.031	0.121	1.548	-0.037	0.001	0.963
20	-1	2988.300	2519.050	2208.856	1157.722	1.551	100	107.802	0.123	1.536	-0.026	-0.018	0.961
21	-1	2987.320	2528.810	?	?	?	?	?	0.119	1.634	-0.026	0.012	0.959
22	-1	?	2481.850	2207.389	962.532	1.204	100	104.031	0.121	1.556	0.000	-0.004	0.962
23	-1	3002.270	2497.450	2207.389	962.532	1.204	100	104.031	0.121	1.546	0.019	-0.011	0.949
24	1	2884.740	2514.540	2160.367	899.949	1.402	100	105.498	0.124	1.558	-0.032	-0.014	0.964
25	-1	3010.410	2632.800	2203.900	1116.413	1.264	100	102.273	0.120	1.423	0.019	0.007	0.977
26	-1	2979.740	2446.560	2257.167	1437.957	1.492	100	106.340	0.120	1.514	0.002	0.006	0.958

//Local Repository/processes/Aula03_Full_2_print* – RapidMiner Studio Educational 7.5.001 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History ExampleSet (Replace Missing Values)

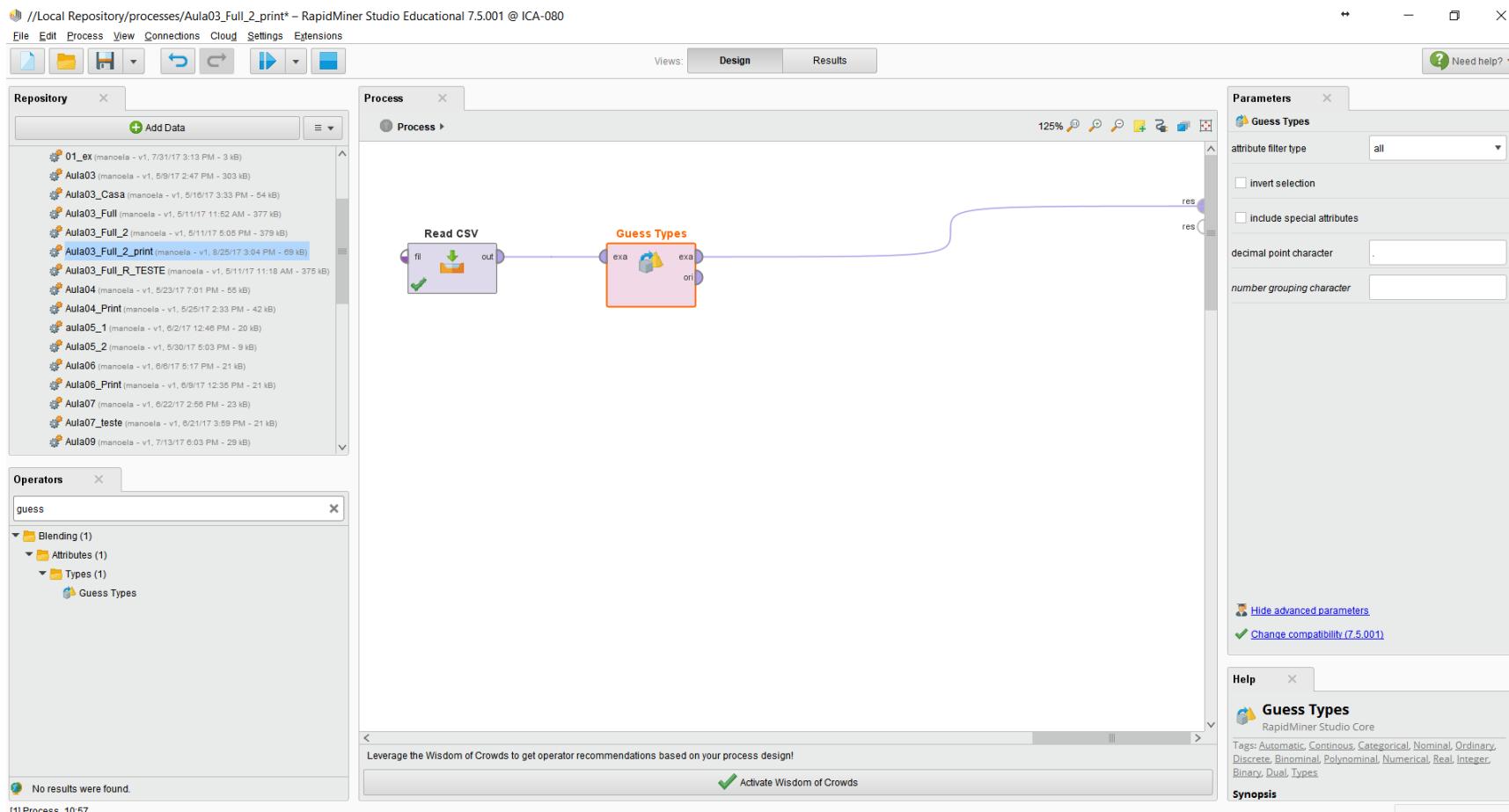
Name	Type	Missing	Statistics	
att537	Integer	0	0	0
att538	Integer	0	Min 0 Max 0	Average 0
att539	Integer	0	Min 0 Max 0	Average 0
att86	Polynomial	0	Least 0.1184 (1) Most 0.1139 (1350)	Values 0.1139 (1350), 0.1148 (8), ...[95 more]
att110	Polynomial	0	Least 0.9842 (1) Most 0.9815 (1037)	Values 0.9815 (1037), 0.9811 (18), ...[94 more]
att111	Polynomial	0	Least 99.9961 (1) Most 100.9209 (1020)	Values 100.9209 (1020), 100.2517 (2), ...[544 more]
att112	Polynomial	0	Least 236.9546 (1) Most 230.8541 (1020)	Values 230.8541 (1020), 230.0055 (2), ...[543 more]
att221	Polynomial	0	Least 0.024 (1) Most 0.0071 (1350)	Values 0.0071 (1350), 0.007 (7), ...[67 more]
att245	Polynomial	0	Least 4.00E-04 (1) Most 0.0014 (1050)	Values 0.0014 (1050), 9.00E-04 (31), ...[64 more]
att246	Polynomial	0	Least 99.9022 (1) Most 0.9406 (1020)	Values 0.9406 (1020), 0.7043 (2), ...[537 more]
att247	Polynomial	0	Least 9.9184 (1) Most 3.5774 (1020)	Values 3.5774 (1020), 3.3372 (2), ...[545 more]
att359	Polynomial	0	Least 0.0047 (1) Most 0.0023 (1371)	Values 0.0023 (1371), 0.0024 (30), ...[18 more]
att383	Polynomial	0	Least 0.6271 (1) Most 5.00E-04 (1103)	Values 5.00E-04 (1103), 3.00E-04 (85), ...[27 more]

Showing attributes 1 - 591

Filter (591 / 591 attributes): Search for Attributes

Repository

- 01_DecisionTree (mancela - v1, 5/2/17 2:04 PM - 2 kB)
- 01_ex (mancela - v1, 7/31/17 3:13 PM - 3 kB)
- Aula03 (mancela - v1, 5/9/17 2:47 PM - 303 kB)
- Aula03_Casa (mancela - v1, 5/16/17 3:33 PM - 54 kB)
- Aula03_Full (mancela - v1, 5/11/17 11:52 AM - 377 kB)
- Aula03_Full_2 (mancela - v1, 5/11/17 5:05 PM - 379 kB)
- Aula03_Full_2_print (mancela - v1, 8/25/17 3:04 PM - 89 kB)
- Aula03_Full_R_TESTE (mancela - v1, 5/11/17 11:18 AM - 1 kB)
- Aula04 (mancela - v1, 5/23/17 7:01 PM - 55 kB)
- Aula04_Print (mancela - v1, 5/25/17 2:33 PM - 42 kB)
- aula05_1 (mancela - v1, 6/2/17 12:46 PM - 20 kB)
- Aula05_2 (mancela - v1, 5/30/17 5:03 PM - 9 kB)
- Aula06 (mancela - v1, 6/6/17 5:17 PM - 21 kB)
- Aula06_Print (mancela - v1, 6/9/17 12:35 PM - 21 kB)
- Aula07 (mancela - v1, 6/22/17 2:25 PM - 23 kB)
- Aula07_Teste (mancela - v1, 6/21/17 3:59 PM - 21 kB)
- Aula09 (mancela - v1, 7/13/17 6:03 PM - 29 kB)
- Aula09_2 (mancela - v1, 7/6/17 9:49 AM - 48 kB)
- Aula09_3 (mancela - v1, 7/6/17 12:52 PM - 17 kB)
- aulaBi (mancela - v1, 7/13/17 6:22 PM - 29 kB)
- Copy of Aula07 (mancela - v1, 6/21/17 6:33 PM - 34 kB)
- Copy of Aula09 (mancela - v1, 7/12/17 5:32 PM - 33 kB)
- IBM_AttributeSelection (mancela - v1, 5/3/17 10:58 AM - 1 kB)
- IBM_AttributeSelection (mancela - v1, 5/4/17 12:12 PM - 1 kB)
- IBM_AttributeSelection_2 (mancela - v1, 5/4/17 12:12 PM - 1 kB)
- IBM_JG_WeightSelection (mancela - v1, 5/4/17 12:18 PM - 1 kB)
- IBM_Limpeza (mancela - v1, 5/2/17 4:07 PM - 8 kB)
- IBM_Limpeza_NN_SVM (mancela - v1, 5/2/17 4:34 PM - 2 kB)
- IBM_MissingValues (mancela - v1, 5/4/17 12:12 PM - 13 kB)
- IBM_NeuralNet_DecisionTree (mancela - v1, 5/2/17 3:38 PM - 1 kB)
- New Text Document (mancela - v1, 5/30/17 5:48 PM - 6 kB)
- R (mancela - v1, 5/10/17 6:24 PM - 45 kB)
- teste (mancela - v1, 5/2/17 2:05 PM - 5 kB)
- teste_wrapper (mancela - v1, 5/5/17 2:28 PM - 3 kB)



//Local Repository/processes/Aula03_Full_2* – RapidMiner Studio Educational 7.5.001 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

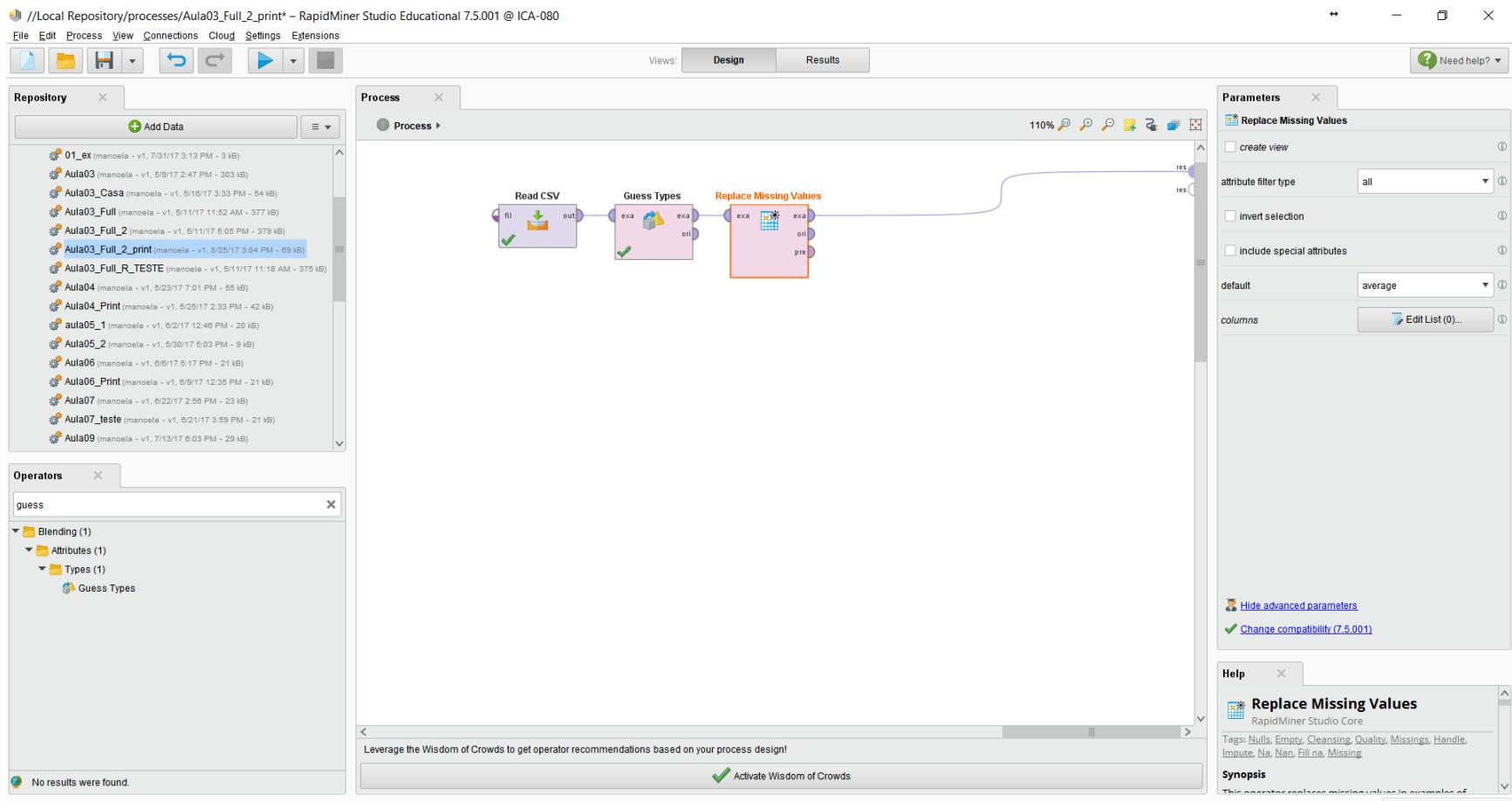
Result History ExampleSet (Guess Types (0))

Name	Type	Missing	Statistics	
att55	Integer	9	Min 0 Max 0 Average 0	
att56	Integer	9	Min 0 Max 0 Average 0	
att57	Integer	9	Min 0 Max 0 Average 0	
att58	Integer	9	Min 0 Max 0 Average 0	
att59	Integer	9	Min 0 Max 0 Average 0	
att1	Real	6	Min 2743.240 Max 3356.350 Average 3014.453	
att2	Real	7	Min 2158.750 Max 2846.440 Average 2495.850	
att3	Real	14	Min 2060.660 Max 2315.267 Average 2200.547	
att4	Real	14	Min 0 Max 3715.042 Average 1396.377	
att5	Real	14	Min 0.681 Max 1114.537 Average 4.197	
att7	Real	14	Min 82.131 Max 129.252 Average 101.113	
att8	Real	9	Min 0 Max 0.129 Average 0.122	
att9	Real	2	Min 1.191 Max 1.656 Average 1.463	

Showing attributes 1 - 591 Examples: 1,567 Special Attributes: 1 Regular Attributes: 590

Repository

- processes (mancela)
 - 01_DecisionTree (mancela - v1, 5/2/17 2:04 PM - 2 kB)
 - 01_ex (mancela - v1, 7/31/17 3:13 PM - 3 kB)
 - Aula03 (mancela - v1, 5/9/17 2:47 PM - 303 kB)
 - Aula03_Casa (mancela - v1, 5/16/17 3:33 PM - 54 kB)
 - Aula03_Full (mancela - v1, 5/11/17 11:52 AM - 377 kB)
 - Aula03_Full_2 (mancela - v1, 5/11/17 5:05 PM - 379 kB)**
 - Aula03_Full_R_TESTE (mancela - v1, 5/11/17 11:18 AM - 379 kB)
 - Aula04 (mancela - v1, 5/23/17 7:01 PM - 55 kB)
 - Aula04_Print (mancela - v1, 5/25/17 2:33 PM - 42 kB)
 - aula05_1 (mancela - v1, 6/2/17 12:46 PM - 20 kB)
 - Aula05_2 (mancela - v1, 5/30/17 5:03 PM - 9 kB)
 - Aula06 (mancela - v1, 6/6/17 5:17 PM - 21 kB)
 - Aula06_Print (mancela - v1, 6/9/17 12:35 PM - 21 kB)
 - aula07 (mancela - v1, 6/22/17 2:58 PM - 23 kB)
 - Aula07_teste (mancela - v1, 6/21/17 3:59 PM - 21 kB)**
 - Aula09 (mancela - v1, 7/3/17 6:03 PM - 29 kB)
 - Aula09_2 (mancela - v1, 7/6/17 9:49 AM - 48 kB)
 - Aula09_3 (mancela - v1, 7/6/17 12:52 PM - 17 kB)
 - aulaBi (mancela - v1, 7/13/17 6:22 PM - 29 kB)
 - Copy of Aula07 (mancela - v1, 6/21/17 6:33 PM - 34 kB)
 - Copy of Aula09 (mancela - v1, 7/12/17 5:32 PM - 33 kB)
 - IBM_AttributeSelection (mancela - v1, 5/3/17 10:58 AM - 1 kB)
 - IBM_AttributeSelection (mancela - v1, 5/4/17 12:12 PM - 1 kB)
 - IBM_AttributeSelection_2 (mancela - v1, 5/4/17 12:12 PM - 1 kB)
 - IBM_I_G_WeightSelection (mancela - v1, 5/4/17 12:18 PM - 1 kB)
 - IBM_Limpeza (mancela - v1, 5/2/17 4:07 PM - 8 kB)
 - IBM_Limpeza_NN_SVM (mancela - v1, 5/2/17 4:34 PM - 2 kB)
 - IBM_MissingValues (mancela - v1, 5/4/17 12:12 PM - 13 kB)
 - IBM_NeuralNet_DecisionTree (mancela - v1, 5/2/17 3:38 PM - 13 kB)
 - New Text Document (mancela - v1, 5/30/17 5:48 PM - 6 kB)
 - R (mancela - v1, 5/10/17 6:24 PM - 45 kB)
 - teste (mancela - v1, 5/2/17 2:05 PM - 5 kB)
 - teste_wrapper (mancela - v1, 5/6/17 2:28 PM - 3 kB)



//Local Repository/processes/Aula03_Full_2_print* – RapidMiner Studio Educational 7.5.001 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

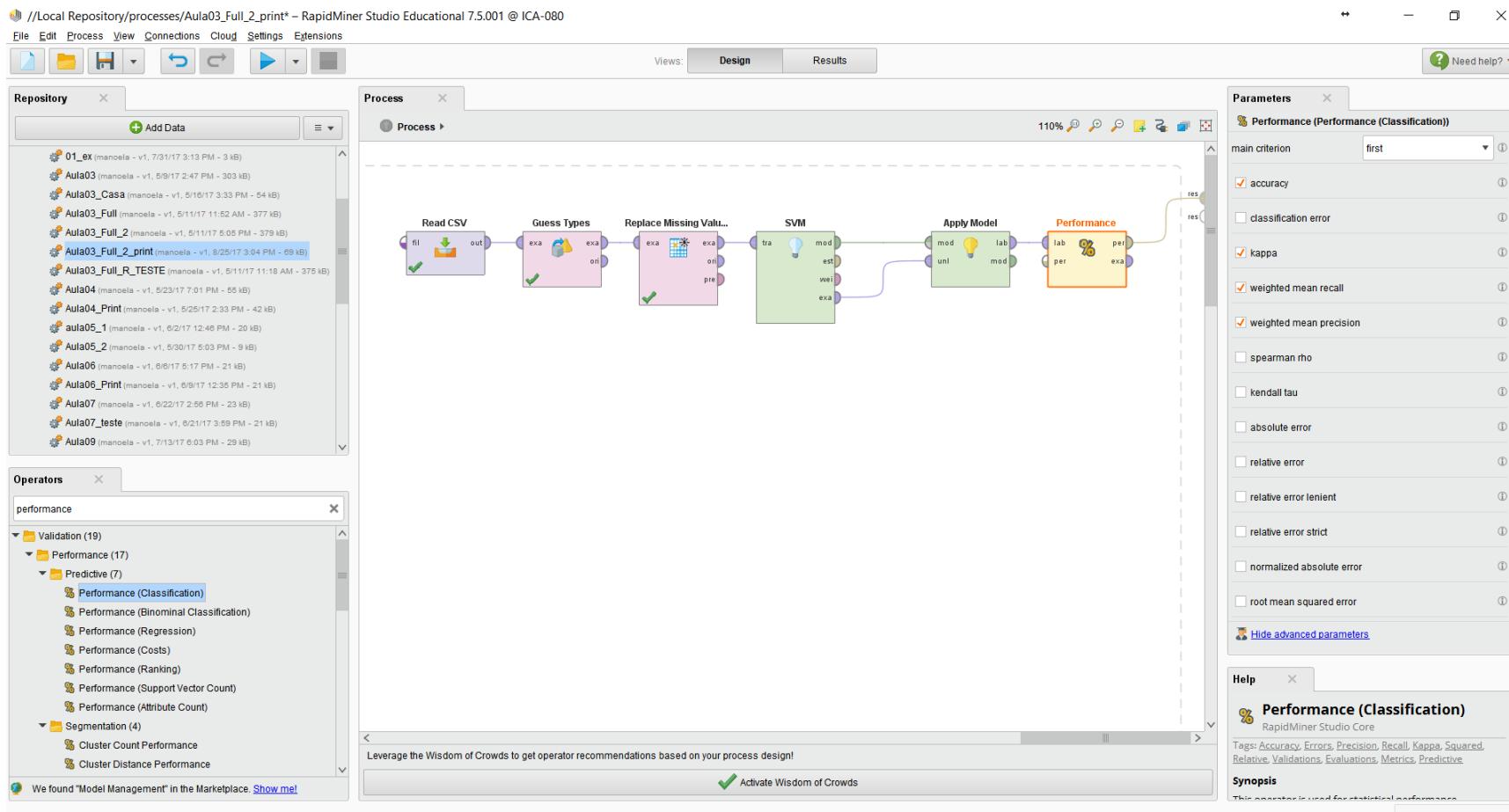
Result History ExampleSet (Replace Missing Values)

Name	Type	Missing	Statistics		
Label att591	Binomial	0	Least 1 (104)	Most -1 (1463)	Values -1 (1463), 1 (104)
att1	Real	0	Min 2743.240	Max 3356.350	Average 3014.453
att2	Real	0	Min 2158.750	Max 2846.440	Average 2495.850
att3	Real	0	Min 2060.660	Max 2315.267	Average 2200.547
att4	Real	0	Min 0	Max 3715.042	Average 1396.377
att5	Real	0	Min 0.681	Max 1114.537	Average 4.197
att6	Integer	0	Min 100	Max 100	Average 100
att7	Real	0	Min 82.131	Max 129.252	Average 101.113
att8	Real	0	Min 0	Max 0.129	Average 0.122
att9	Real	0	Min 1.191	Max 1.656	Average 1.463
att10	Real	0	Min -0.053	Max 0.075	Average -0.001
att11	Real	0	Min -0.035	Max 0.053	Average 0.000
att12	Real	0	Min 0.655	Max 0.985	Average 0.964

Showing attributes 1 - 591 Examples: 1.567 Special Attributes: 1 Regular Attributes: 590

Repository

- 01_DecisionTree (mancela - v1, 5/2/17 2:04 PM - 2 kB)
- 01_ex (mancela - v1, 7/31/17 3:13 PM - 3 kB)
- Aula03 (mancela - v1, 5/9/17 2:47 PM - 303 kB)
- Aula03_Casa (mancela - v1, 5/16/17 3:33 PM - 54 kB)
- Aula03_Full (mancela - v1, 5/11/17 11:52 AM - 377 kB)
- Aula03_Full_2 (mancela - v1, 5/11/17 5:05 PM - 379 kB)
- Aula03_Full_2_print (mancela - v1, 8/25/17 3:04 PM - 89 kB)
- Aula03_Full_R_TESTE (mancela - v1, 5/11/17 11:18 AM - 1 kB)
- Aula04 (mancela - v1, 5/23/17 7:01 PM - 55 kB)
- Aula04_Print (mancela - v1, 5/25/17 2:33 PM - 42 kB)
- aula05_1 (mancela - v1, 6/2/17 12:46 PM - 20 kB)
- Aula05_2 (mancela - v1, 5/30/17 5:03 PM - 9 kB)
- Aula06 (mancela - v1, 6/6/17 5:17 PM - 21 kB)
- Aula06_Print (mancela - v1, 6/9/17 12:35 PM - 21 kB)
- Aula07 (mancela - v1, 6/22/17 2:25 PM - 23 kB)
- Aula07_Test (mancela - v1, 6/21/17 3:59 PM - 21 kB)
- Aula09 (mancela - v1, 7/13/17 6:03 PM - 29 kB)
- Aula09_2 (mancela - v1, 7/6/17 9:49 AM - 48 kB)
- Aula09_3 (mancela - v1, 7/6/17 12:52 PM - 17 kB)
- aulaBi (mancela - v1, 7/13/17 6:22 PM - 29 kB)
- Copy of Aula07 (mancela - v1, 6/21/17 6:33 PM - 34 kB)
- Copy of Aula09 (mancela - v1, 7/12/17 5:32 PM - 33 kB)
- IBM_AttributeSelection (mancela - v1, 5/3/17 10:58 AM - 1 kB)
- IBM_AttributeSelection (mancela - v1, 5/4/17 12:12 PM - 1 kB)
- IBM_AttributeSelection_2 (mancela - v1, 5/4/17 12:12 PM - 1 kB)
- IBM_JG_WeightSelection (mancela - v1, 5/4/17 12:18 PM - 1 kB)
- IBM_Limpeza (mancela - v1, 5/2/17 4:07 PM - 8 kB)
- IBM_Limpeza_NN_SVM (mancela - v1, 5/2/17 4:34 PM - 2 kB)
- IBM_MissingValues (mancela - v1, 5/4/17 12:12 PM - 13 kB)
- IBM_NeuralNet_DecisionTree (mancela - v1, 5/2/17 3:38 PM - 1 kB)
- New Text Document (mancela - v1, 5/30/17 5:48 PM - 6 kB)
- R (mancela - v1, 5/10/17 6:24 PM - 45 kB)
- teste (mancela - v1, 5/2/17 2:05 PM - 5 kB)
- teste_wrapper (mancela - v1, 5/5/17 2:28 PM - 3 kB)



//Local Repository/processes/Aula03_Full_2_print* – RapidMiner Studio Educational 7.5.001 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History PerformanceVector (Performance)

Criterion accuracy

accuracy: 93.43%

	true -1	true 1	class precision
pred. -1	1463	103	93.42%
pred. 1	0	1	100.00%
class recall	100.00%	0.96%	

Description Annotations

Repository

- 01_DecisionTree (mancela - v1, 5/2/17 2:04 PM - 2 kB)
- 01_ex (mancela - v1, 7/31/17 3:13 PM - 3 kB)
- Aula03 (mancela - v1, 5/9/17 2:47 PM - 303 kB)
- Aula03_Casa (mancela - v1, 5/16/17 3:33 PM - 54 kB)
- Aula03_Full (mancela - v1, 5/11/17 11:52 AM - 377 kB)
- Aula03_Full_2 (mancela - v1, 5/11/17 5:05 PM - 379 kB)
- Aula03_Full_2_print (mancela - v1, 8/25/17 3:04 PM - 89 kB)
- Aula03_Full_R_TESTE (mancela - v1, 5/11/17 11:18 AM - 1 kB)
- Aula04 (mancela - v1, 5/23/17 7:01 PM - 55 kB)
- Aula04_Print (mancela - v1, 5/25/17 2:33 PM - 42 kB)
- aula05_1 (mancela - v1, 6/2/17 12:46 PM - 20 kB)
- Aula05_2 (mancela - v1, 5/30/17 5:03 PM - 9 kB)
- Aula06 (mancela - v1, 6/6/17 5:17 PM - 21 kB)
- Aula06_Print (mancela - v1, 6/9/17 12:35 PM - 21 kB)
- Aula07 (mancela - v1, 6/22/17 2:55 PM - 23 kB)
- Aula07_Test (mancela - v1, 6/21/17 3:59 PM - 21 kB)
- Aula09 (mancela - v1, 7/13/17 6:03 PM - 29 kB)
- Aula09_2 (mancela - v1, 7/6/17 9:49 AM - 48 kB)
- Aula09_3 (mancela - v1, 7/6/17 12:52 PM - 17 kB)
- aulaBi (mancela - v1, 7/13/17 6:22 PM - 29 kB)
- Copy of Aula07 (mancela - v1, 6/21/17 6:33 PM - 34 kB)
- Copy of Aula09 (mancela - v1, 7/12/17 5:32 PM - 33 kB)
- IBM_AttributeSelection (mancela - v1, 5/3/17 10:58 AM - 1 kB)
- IBM_AttributeSelection_2 (mancela - v1, 5/4/17 12:12 PM - 1 kB)
- IBM_AttributeSelection_3 (mancela - v1, 5/4/17 12:12 PM - 1 kB)
- IBM_G_WeightSelection (mancela - v1, 5/4/17 12:18 PM - 1 kB)
- IBM_Limpeza (mancela - v1, 5/2/17 4:07 PM - 8 kB)
- IBM_Limpeza_NN_SVM (mancela - v1, 5/2/17 4:34 PM - 2 kB)
- IBM_MissingValues (mancela - v1, 5/4/17 12:12 PM - 13 kB)
- IBM_NeuralNet_DecisionTree (mancela - v1, 5/2/17 3:38 PM - 1 kB)
- New Text Document (mancela - v1, 5/30/17 5:48 PM - 6 kB)
- R (mancela - v1, 5/10/17 6:24 PM - 45 kB)
- teste (mancela - v1, 5/2/17 2:05 PM - 5 kB)
- teste_wrapper (mancela - v1, 5/5/17 2:28 PM - 3 kB)

//Local Repository/processes/Aula03_Full_2_print* – RapidMiner Studio Educational 7.5.001 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History PerformanceVector (Performance)

PerformanceVector

```
PerformanceVector:  
accuracy: 93.43%  
ConfusionMatrix:  
True: -1 1  
-1: 1463 103  
1: 0 1  
kappa: 0.018  
ConfusionMatrix:  
True: -1 1  
-1: 1463 103  
1: 0 1  
weighted_mean_recall: 50.48%, weights: 1, 1  
ConfusionMatrix:  
True: -1 1  
-1: 1463 103  
1: 0 1  
weighted_mean_precision: 96.71%, weights: 1, 1  
ConfusionMatrix:  
True: -1 1  
-1: 1463 103  
1: 0 1
```

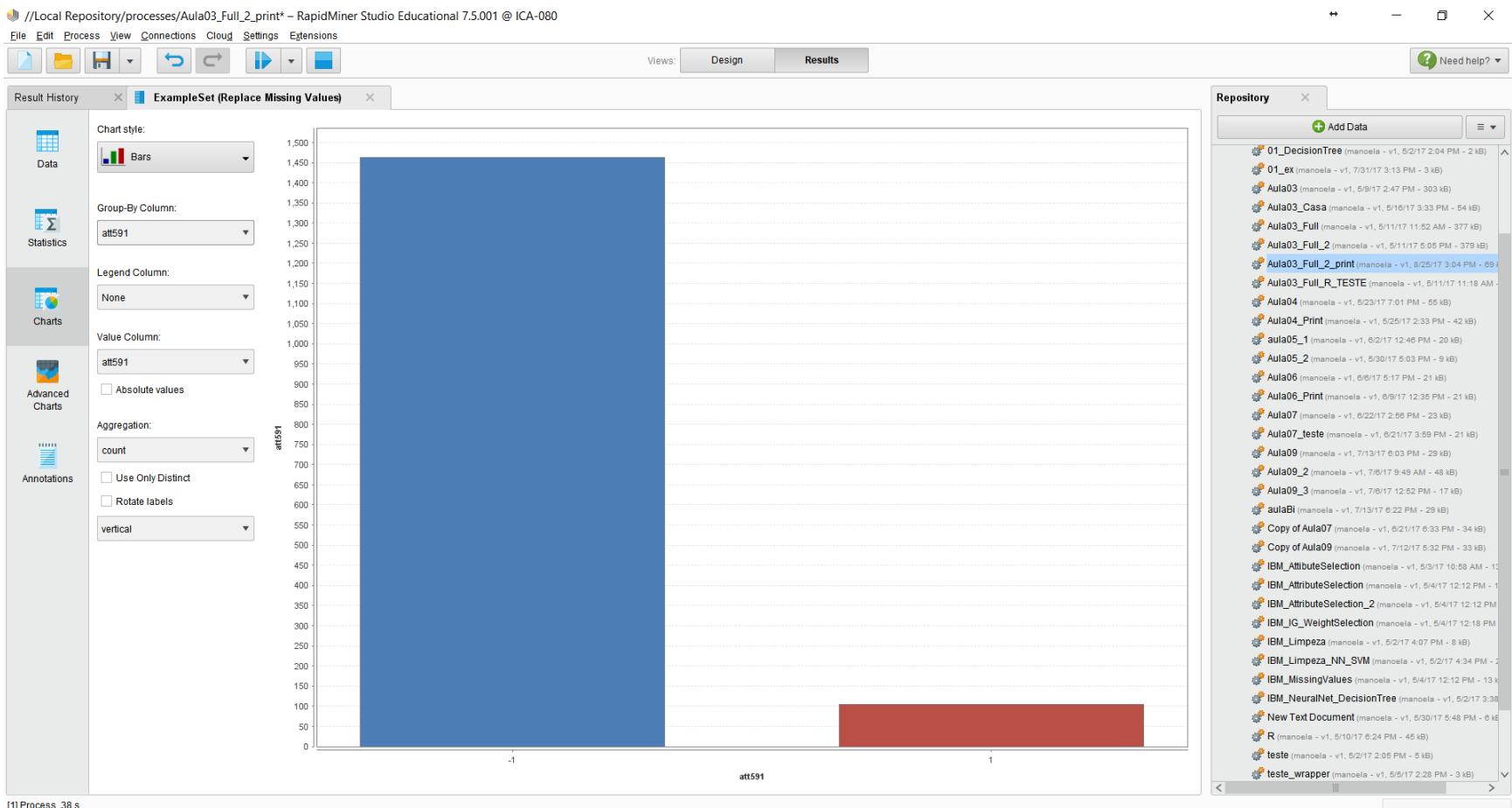
Annotations

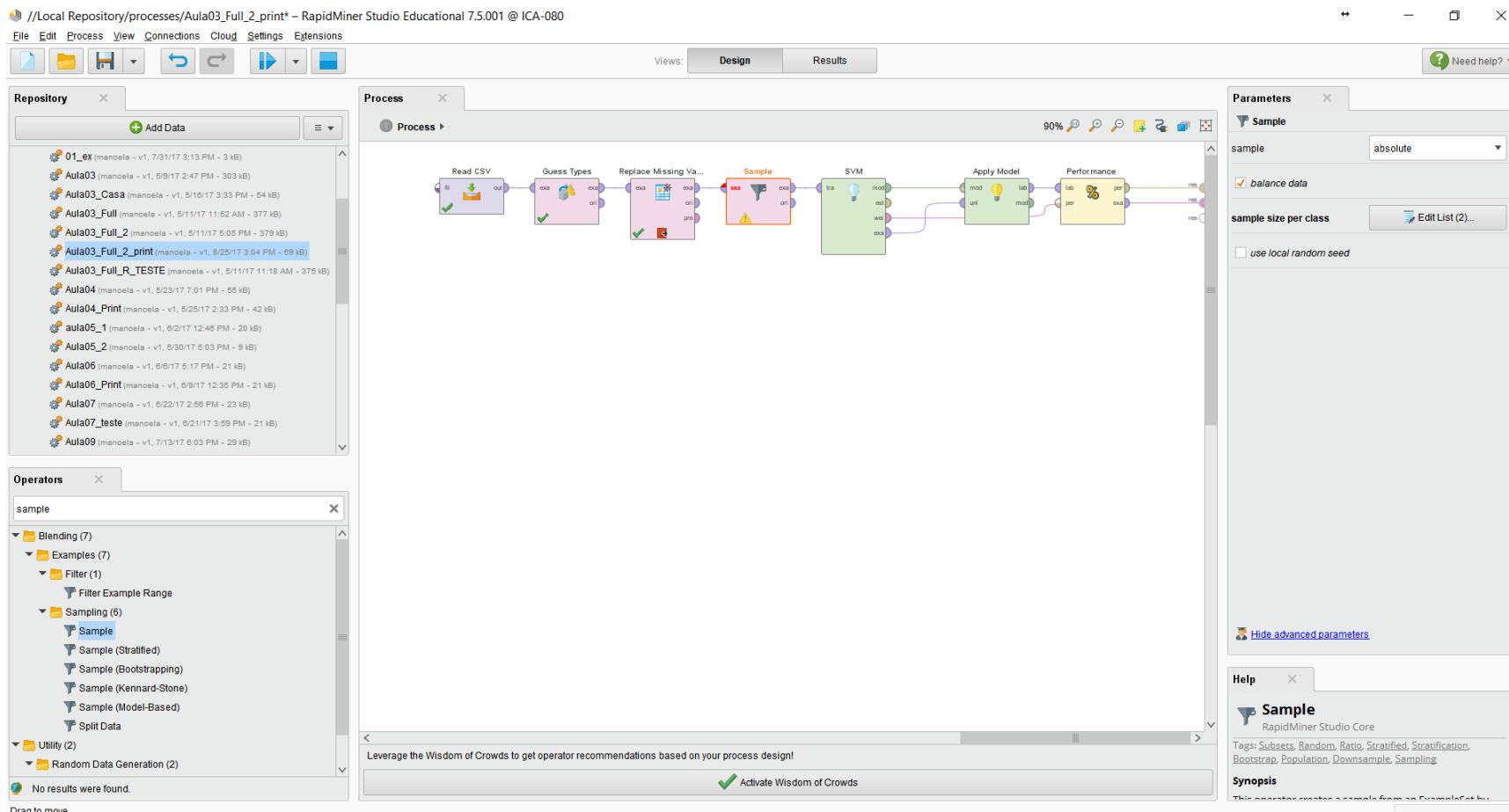
Repository

- 01_DecisionTree (mancela - v1, 5/2/17 2:04 PM - 2 kB)
- 01_ex (mancela - v1, 7/31/17 3:13 PM - 3 kB)
- Aula03 (mancela - v1, 5/9/17 2:47 PM - 303 kB)
- Aula03_Casa (mancela - v1, 5/16/17 3:33 PM - 54 kB)
- Aula03_Full (mancela - v1, 5/11/17 11:52 AM - 377 kB)
- Aula03_Full_2 (mancela - v1, 5/11/17 5:05 PM - 379 kB)
- Aula03_Full_2_print (mancela - v1, 8/25/17 3:04 PM - 89 kB)
- Aula03_Full_R_TESTE (mancela - v1, 5/11/17 11:18 AM - 1 kB)
- Aula04 (mancela - v1, 5/23/17 7:01 PM - 55 kB)
- Aula04_Print (mancela - v1, 5/25/17 2:33 PM - 42 kB)
- aula05_1 (mancela - v1, 6/2/17 12:46 PM - 20 kB)
- Aula05_2 (mancela - v1, 5/30/17 5:03 PM - 9 kB)
- Aula06 (mancela - v1, 6/6/17 5:17 PM - 21 kB)
- Aula06_Print (mancela - v1, 6/9/17 12:35 PM - 21 kB)
- Aula07 (mancela - v1, 6/22/17 2:52 PM - 23 kB)
- Aula07_Teste (mancela - v1, 6/21/17 3:59 PM - 21 kB)
- Aula09 (mancela - v1, 7/13/17 6:03 PM - 29 kB)
- Aula09_2 (mancela - v1, 7/6/17 9:49 AM - 48 kB)
- Aula09_3 (mancela - v1, 7/6/17 12:52 PM - 17 kB)
- aulaBi (mancela - v1, 7/13/17 6:22 PM - 29 kB)
- Copy of Aula07 (mancela - v1, 6/21/17 6:33 PM - 34 kB)
- Copy of Aula09 (mancela - v1, 7/12/17 5:32 PM - 33 kB)
- IBM_AttributeSelection (mancela - v1, 5/3/17 10:58 AM - 1 kB)
- IBM_AttributeSelection_2 (mancela - v1, 5/4/17 12:12 PM - 1 kB)
- IBM_AttributeSelection_3 (mancela - v1, 5/4/17 12:12 PM - 1 kB)
- IBM_G_WeightSelection (mancela - v1, 5/4/17 12:18 PM - 1 kB)
- IBM_Limpeza (mancela - v1, 5/2/17 4:07 PM - 8 kB)
- IBM_Limpeza_NN_SVM (mancela - v1, 5/2/17 4:34 PM - 2 kB)
- IBM_MissingValues (mancela - v1, 5/4/17 12:12 PM - 13 kB)
- IBM_NeuralNet_DecisionTree (mancela - v1, 5/2/17 3:38 PM - 1 kB)
- New Text Document (mancela - v1, 5/30/17 5:48 PM - 6 kB)
- R (mancela - v1, 5/10/17 6:24 PM - 45 kB)
- teste (mancela - v1, 5/2/17 2:05 PM - 5 kB)
- teste_wrapper (mancela - v1, 5/5/17 2:28 PM - 3 kB)

RESULTADOS

SVM	Acurácia	Kappa	Recall	Precisão
Missing	93.43%	0.018	50.48%	96.71%



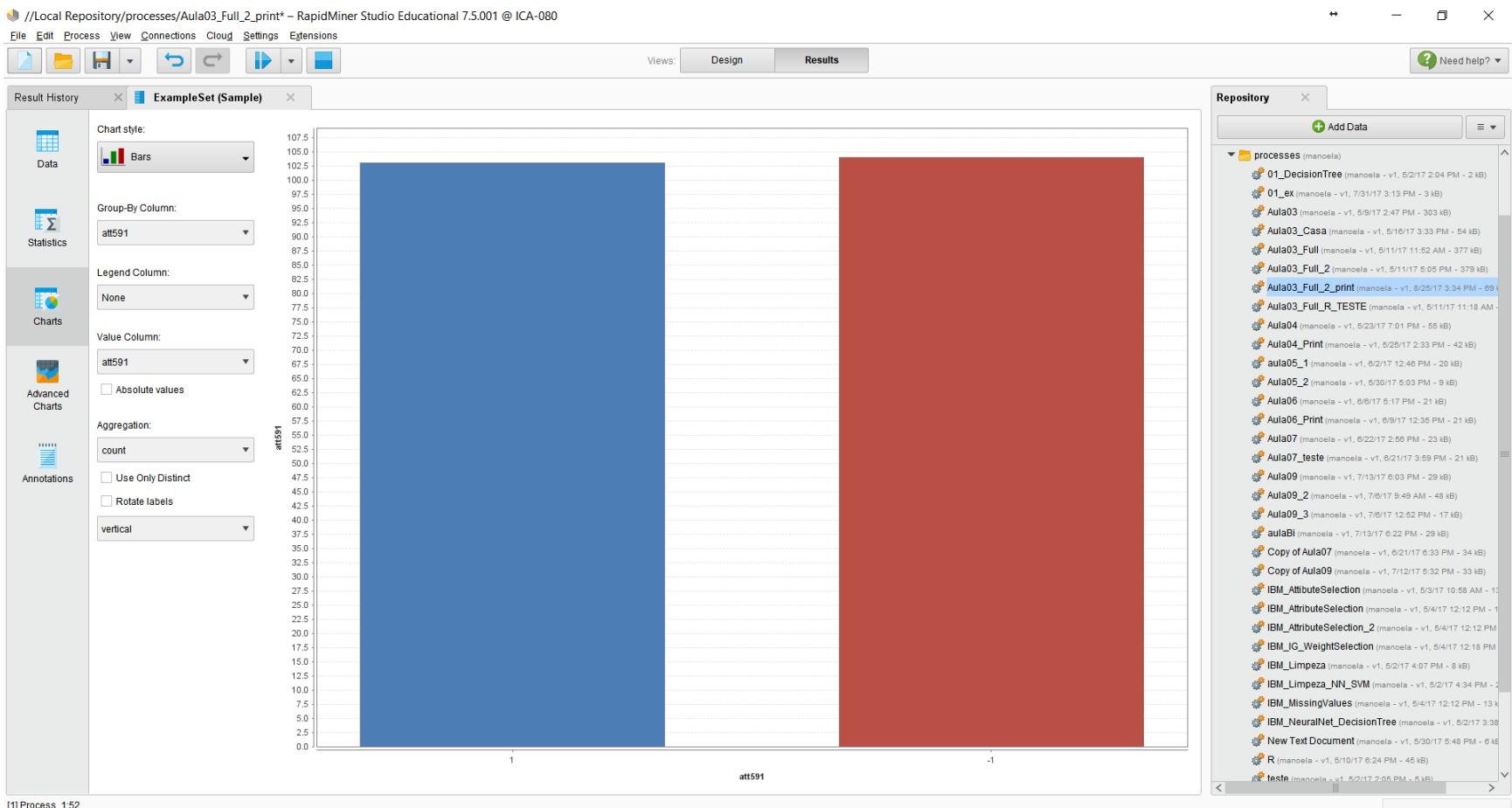


RAPIDMINER

The image shows two windows from the RapidMiner software. On the left is the main 'Parameters' dialog, specifically the 'Sample' section. It includes fields for 'sample' (set to 'absolute'), 'balance data' (checked), 'sample size per class' (with a button to 'Edit List (2)...'), and 'use local random seed' (unchecked). A green box highlights the 'balance data' checkbox and its associated row. An arrow points from this highlighted area to the right window, which is a modal titled 'Edit Parameter List: sample size per class'. This modal contains a table with two entries:

class	size
-1	104
1	104

At the bottom of the modal are buttons for 'Add Entry' (green plus icon), 'Remove Entry' (red minus icon), 'Apply' (green checkmark icon), and 'Cancel' (red X icon).



//Local Repository/processes/Aula03_Full_2_print* – RapidMiner Studio Educational 7.5.001 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History PerformanceVector (Performance)

Criterion accuracy

Table View Plot View

accuracy: 80.29%

	true -1	true 1	class precision
pred. -1	85	22	79.44%
pred. 1	19	82	81.19%
class recall	81.73%	78.85%	

Description Annotations

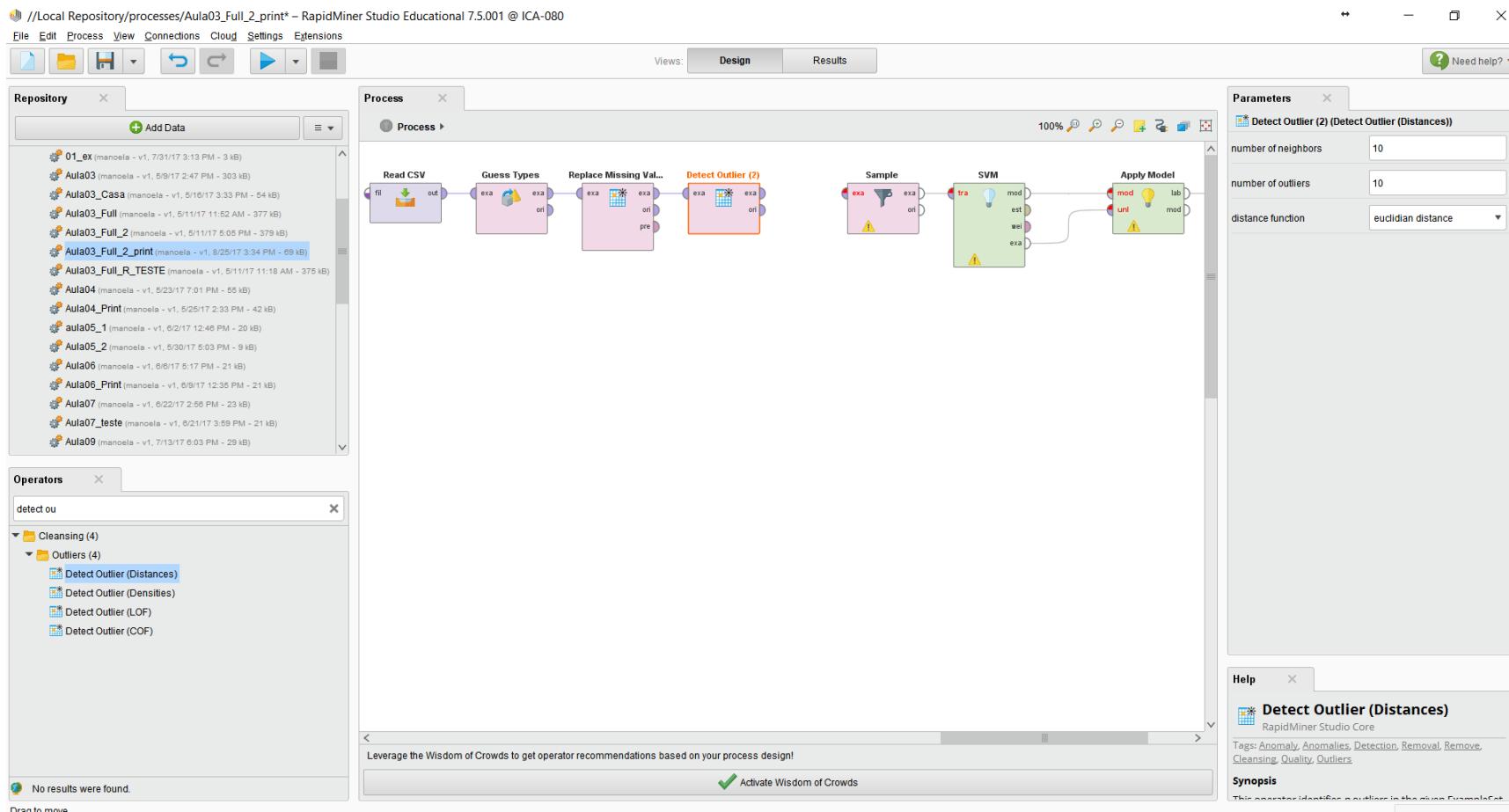
Repository

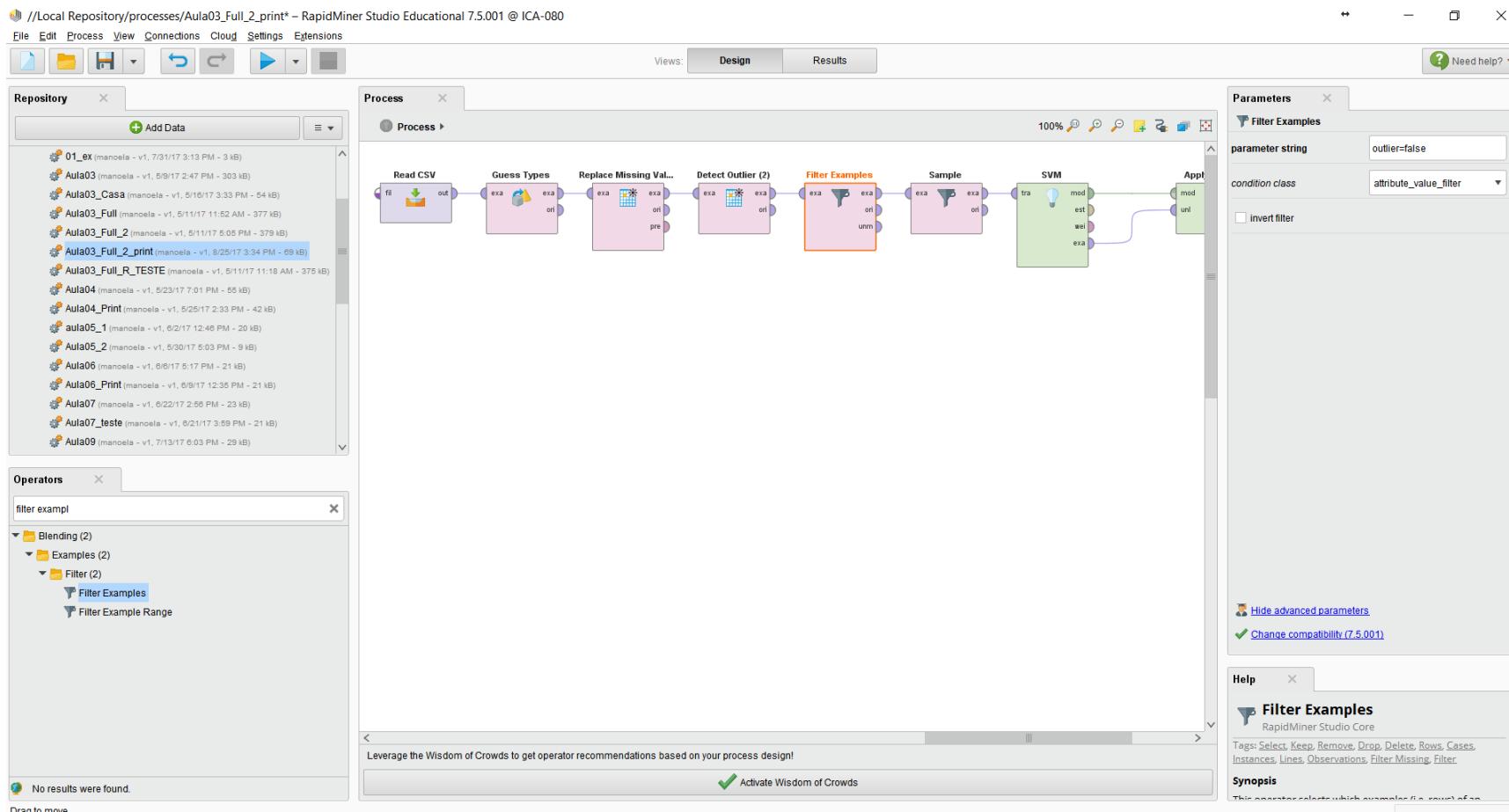
Add Data

- 01_DecisionTree (mancela - v1, 5/2/17 2:04 PM - 2 kB)
- 01_ex (mancela - v1, 7/31/17 3:13 PM - 3 kB)
- Aula03 (mancela - v1, 5/9/17 2:47 PM - 303 kB)
- Aula03_Casa (mancela - v1, 5/16/17 3:33 PM - 54 kB)
- Aula03_Full (mancela - v1, 5/11/17 11:52 AM - 377 kB)
- Aula03_Full_2 (mancela - v1, 5/11/17 5:05 PM - 379 kB)
- Aula03_Full_2_print (mancela - v1, 8/25/17 3:04 PM - 89 kB)
- Aula03_Full_R_TESTE (mancela - v1, 5/11/17 11:18 AM - 1 kB)
- Aula04 (mancela - v1, 5/23/17 7:01 PM - 55 kB)
- Aula04_Print (mancela - v1, 5/25/17 2:33 PM - 42 kB)
- aula05_1 (mancela - v1, 6/2/17 12:46 PM - 20 kB)
- Aula05_2 (mancela - v1, 5/30/17 5:03 PM - 9 kB)
- Aula06 (mancela - v1, 6/6/17 5:17 PM - 21 kB)
- Aula06_Print (mancela - v1, 6/9/17 12:35 PM - 21 kB)
- Aula07 (mancela - v1, 6/22/17 2:25 PM - 23 kB)
- Aula07_Test (mancela - v1, 6/21/17 3:59 PM - 21 kB)
- Aula09 (mancela - v1, 7/13/17 6:03 PM - 29 kB)
- Aula09_2 (mancela - v1, 7/6/17 9:49 AM - 48 kB)
- Aula09_3 (mancela - v1, 7/6/17 12:52 PM - 17 kB)
- aulaBi (mancela - v1, 7/13/17 6:22 PM - 29 kB)
- Copy of Aula07 (mancela - v1, 6/21/17 6:33 PM - 34 kB)
- Copy of Aula09 (mancela - v1, 7/12/17 5:32 PM - 33 kB)
- IBM_AttributeSelection (mancela - v1, 5/3/17 10:58 AM - 1 kB)
- IBM_AttributeSelection (mancela - v1, 5/4/17 12:12 PM - 1 kB)
- IBM_AttributeSelection_2 (mancela - v1, 5/4/17 12:12 PM - 1 kB)
- IBM_JG_WeightSelection (mancela - v1, 5/4/17 12:18 PM - 1 kB)
- IBM_Limpeza (mancela - v1, 5/2/17 4:07 PM - 8 kB)
- IBM_Limpeza_NN_SVM (mancela - v1, 5/2/17 4:34 PM - 2 kB)
- IBM_MissingValues (mancela - v1, 5/4/17 12:12 PM - 13 kB)
- IBM_NeuralNet_DecisionTree (mancela - v1, 5/2/17 3:38 PM - 1 kB)
- New Text Document (mancela - v1, 5/30/17 5:48 PM - 6 kB)
- R (mancela - v1, 5/10/17 6:24 PM - 45 kB)
- teste (mancela - v1, 5/2/17 2:05 PM - 5 kB)
- teste_wrapper (mancela - v1, 5/5/17 2:28 PM - 3 kB)

RESULTADOS

SVM	Acurácia	Kappa	Recall	Precisão
Missing	93.43%	0.018	50.48%	96.71%
+Balanceamento	80.29%	0.606	80.29%	80.31%





File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History ExampleSet (Detect Outlier (2))

ExampleSet (1567 examples, 2 special attributes, 590 regular attributes)

Filter (1567 / 1,567 examples): all

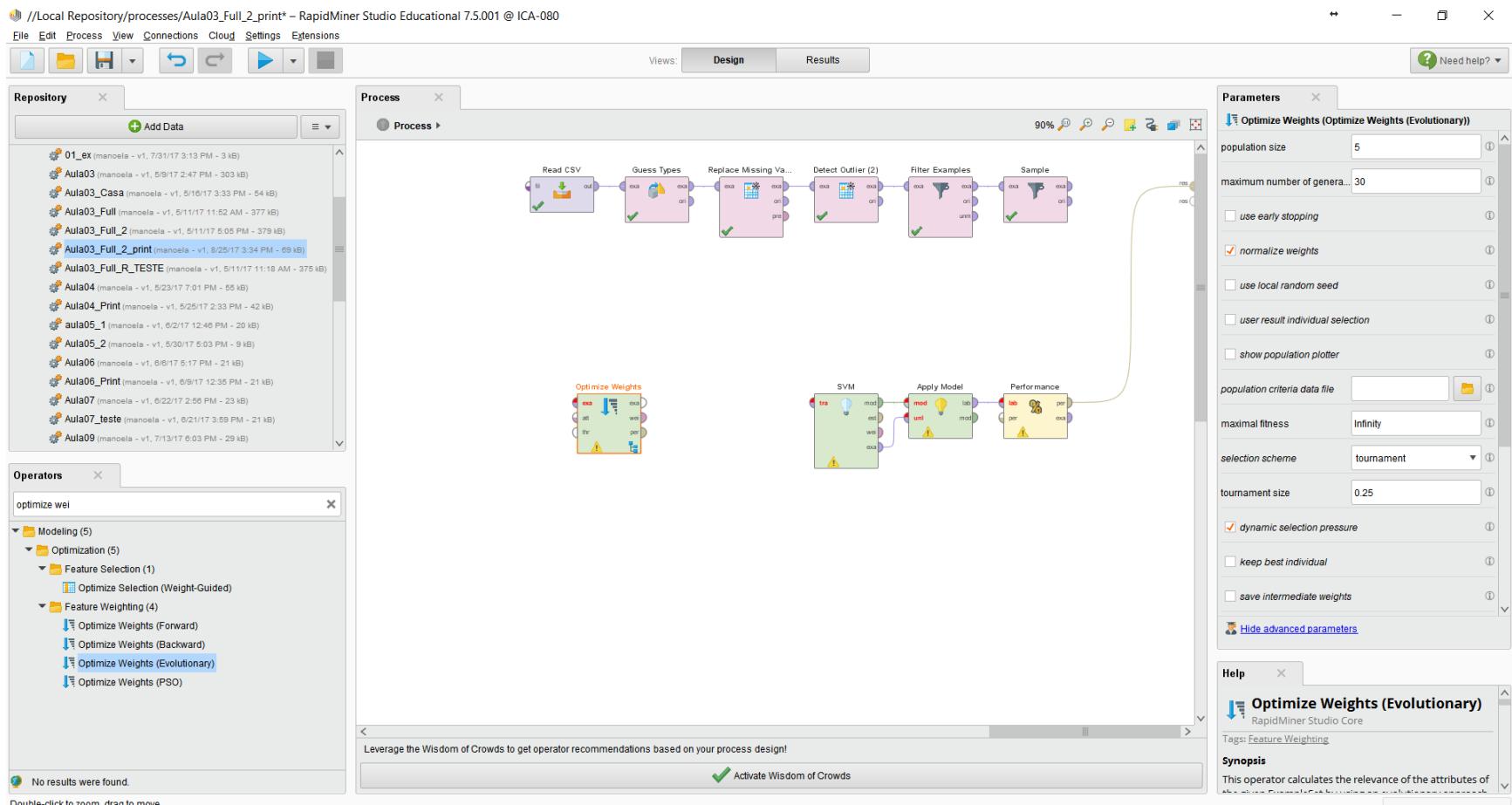
Row No.	att591	outlier	att1	att2	att3	att4	att5	att6	att7	att8	att9	att10	att11
1	-1	false	3030.930	2564	2187.733	1411.127	1.360	100	97.613	0.124	1.500	0.016	-0.003
2	-1	false	3095.780	2465.140	2230.422	1463.661	0.829	100	102.343	0.125	1.497	-0.000	-0.015
3	1	false	2932.610	2559.940	2186.411	1698.017	1.510	100	95.488	0.124	1.444	0.004	0.001
4	-1	false	2988.720	2479.900	2199.033	909.793	1.320	100	104.237	0.122	1.488	-0.012	-0.003
5	-1	false	3032.240	2502.870	2233.367	1326.520	1.533	100	100.397	0.123	1.503	-0.003	-0.007
6	-1	false	2946.250	2432.840	2233.367	1326.520	1.533	100	100.397	0.123	1.529	0.017	0.005
7	-1	false	3030.270	2430.120	2230.422	1463.661	0.829	100	102.343	0.125	1.582	-0.027	0.011
8	-1	false	3058.880	2690.150	2248.900	1004.469	0.788	100	106.240	0.118	1.515	0.016	0.001
9	-1	false	2967.680	2600.470	2248.900	1004.469	0.788	100	106.240	0.118	1.536	0.011	-0.007
10	-1	false	3016.110	2428.370	2248.900	1004.469	0.788	100	106.240	0.118	1.538	0.016	0.005
11	1	false	2994.050	2548.210	2195.122	1046.147	1.320	100	103.340	0.122	1.514	-0.019	0.001
12	1	false	2928.840	2479.400	2196.211	1605.758	0.996	100	97.916	0.126	1.469	0.017	-0.015
13	-1	false	2920.070	2507.400	2195.122	1046.147	1.320	100	103.340	0.122	1.531	-0.026	0.022
14	-1	false	3051.440	2529.270	2184.433	877.627	1.467	100	107.871	0.124	1.524	-0.021	-0.003
15	1	false	2963.970	2629.480	2224.622	947.774	1.292	100	104.849	0.120	1.447	0.014	-0.012
16	-1	false	2988.310	2546.260	2224.622	947.774	1.292	100	104.849	0.120	1.546	0.025	-0.002
17	-1	false	3028.020	2560.870	2270.256	1258.456	1.395	100	104.808	0.121	1.437	0.015	-0.004
18	-1	false	3032.730	2517.790	2270.256	1258.456	1.395	100	104.808	0.121	1.554	0.022	-0.003
19	-1	false	3040.340	2501.160	2207.389	962.532	1.204	100	104.031	0.121	1.548	-0.037	0.001
20	-1	false	2988.300	2519.050	2208.856	1157.722	1.551	100	107.802	0.123	1.536	-0.026	-0.018
21	-1	false	2987.320	2528.810	2200.547	1396.377	4.197	100	101.113	0.119	1.634	-0.026	0.012
22	-1	false	3014.453	2481.850	2207.389	962.532	1.204	100	104.031	0.121	1.556	0.000	-0.004
23	-1	false	3002.270	2497.450	2207.389	962.532	1.204	100	104.031	0.121	1.546	0.019	-0.011
24	1	false	2884.740	2514.540	2160.367	899.949	1.402	100	105.498	0.124	1.558	-0.032	-0.014
25	-1	false	3010.410	2632.800	2203.900	1116.413	1.264	100	102.273	0.120	1.423	0.019	0.007
26	-1	false	2979.740	2446.560	2257.167	1437.957	1.492	100	106.340	0.120	1.514	0.002	0.006

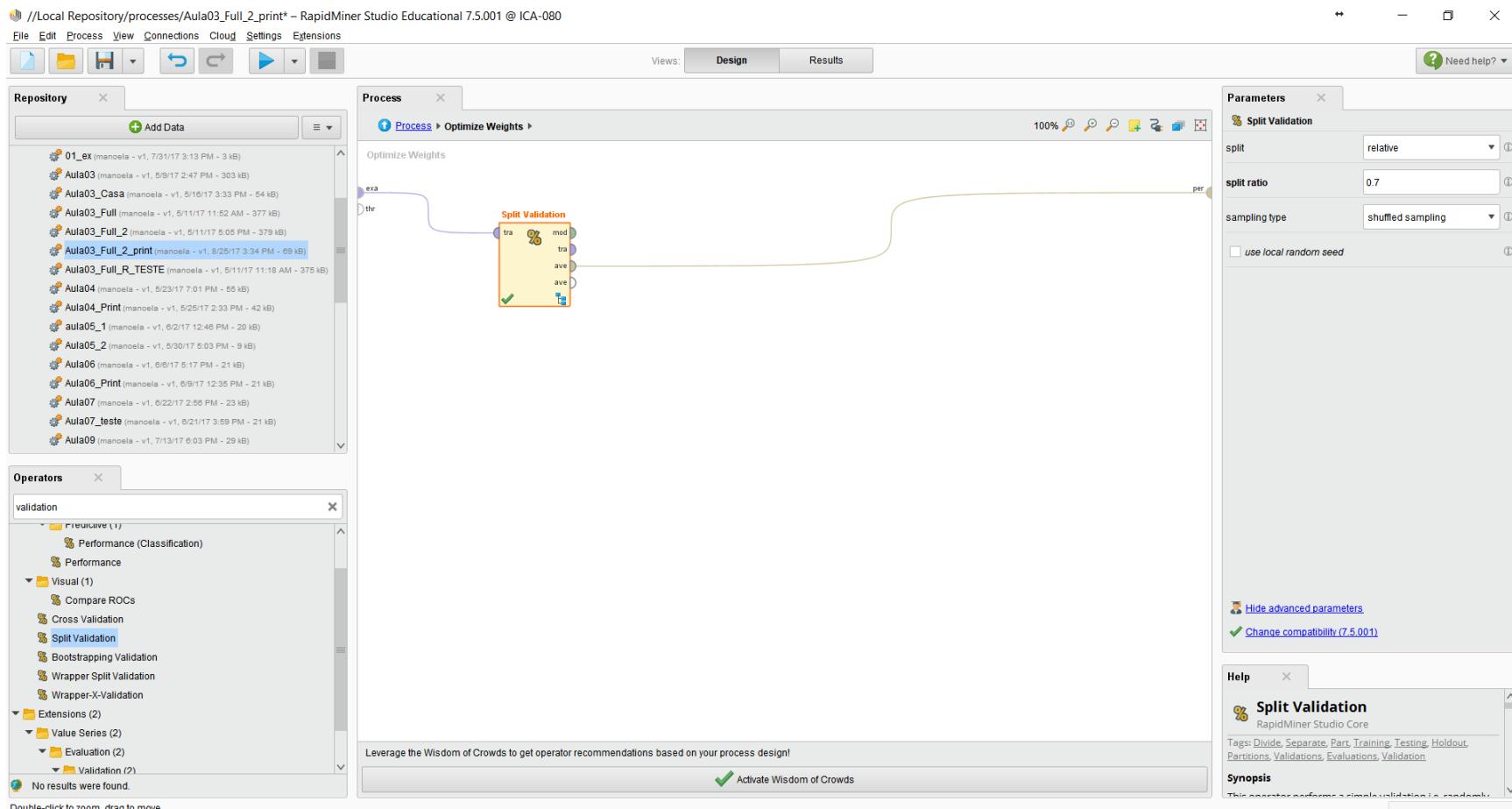
Repository Add Data

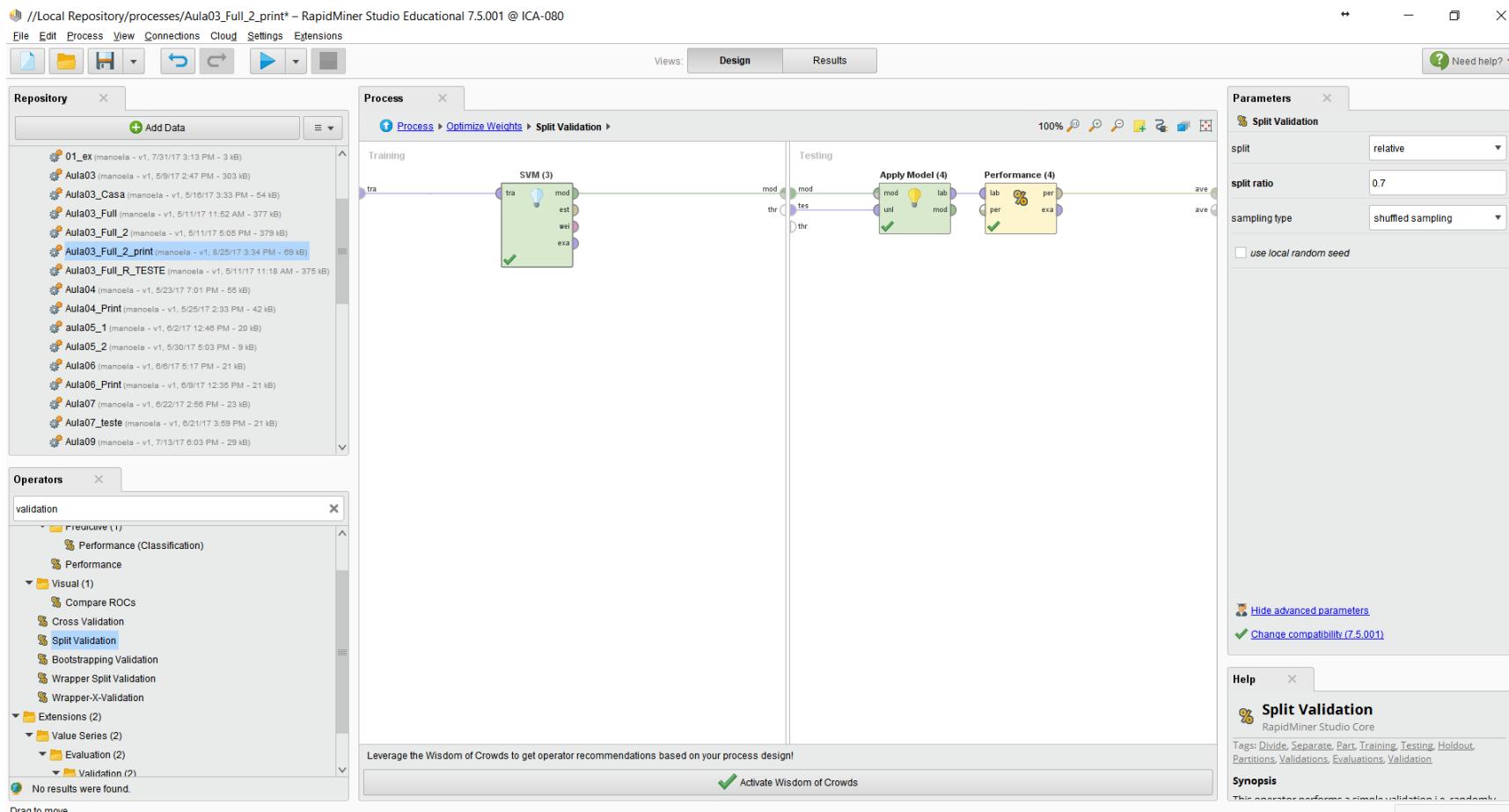
- 01_DecisionTree (mancela - v1, 5/2/17 2:04 PM - 2 kB)
- 01_ex (mancela - v1, 7/31/17 3:13 PM - 3 kB)
- Aula03 (mancela - v1, 5/16/17 2:47 PM - 303 kB)
- Aula03_Casa (mancela - v1, 5/16/17 3:33 PM - 54 kB)
- Aula03_Full (mancela - v1, 5/11/17 11:52 AM - 377 kB)
- Aula03_Full_2 (mancela - v1, 5/11/17 5:05 PM - 379 kB)
- Aula03_Full_2_print (mancela - v1, 8/26/17 3:34 PM - 89 kB)
- Aula03_Full_R_TESTE (mancela - v1, 5/11/17 11:18 AM - 1 kB)
- Aula04 (mancela - v1, 5/23/17 7:01 PM - 55 kB)
- Aula04_Print (mancela - v1, 5/25/17 2:33 PM - 42 kB)
- aula05_1 (mancela - v1, 6/2/17 12:46 PM - 20 kB)
- Aula05_2 (mancela - v1, 5/30/17 5:03 PM - 9 kB)
- Aula06 (mancela - v1, 6/6/17 5:17 PM - 21 kB)
- Aula06_Print (mancela - v1, 6/9/17 12:35 PM - 21 kB)
- Aula07 (mancela - v1, 6/22/17 2:25 PM - 23 kB)
- Aula07_Teste (mancela - v1, 6/21/17 3:59 PM - 21 kB)
- Aula09 (mancela - v1, 7/13/17 6:03 PM - 29 kB)
- Aula09_2 (mancela - v1, 7/6/17 9:49 AM - 48 kB)
- Aula09_3 (mancela - v1, 7/6/17 12:52 PM - 17 kB)
- aulaBi (mancela - v1, 7/13/17 6:22 PM - 29 kB)
- Copy of Aula07 (mancela - v1, 6/21/17 6:33 PM - 34 kB)
- Copy of Aula09 (mancela - v1, 7/12/17 5:32 PM - 33 kB)
- IBM_AttributeSelection (mancela - v1, 5/3/17 10:58 AM - 1 kB)
- IBM_AttributeSelection (mancela - v1, 5/4/17 12:12 PM - 1 kB)
- IBM_AttributeSelection_2 (mancela - v1, 5/4/17 12:12 PM - 1 kB)
- IBM_G_WeightSelection (mancela - v1, 5/4/17 12:18 PM - 1 kB)
- IBM_Limpeza (mancela - v1, 5/2/17 4:07 PM - 8 kB)
- IBM_Limpeza_NN_SVM (mancela - v1, 5/2/17 4:34 PM - 2 kB)
- IBM_MissingValues (mancela - v1, 5/4/17 12:12 PM - 13 kB)
- IBM_NeuralNet_DecisionTree (mancela - v1, 5/2/17 3:38 kB)
- New Text Document (mancela - v1, 5/30/17 5:48 PM - 6 kB)
- R (mancela - v1, 5/10/17 6:24 PM - 45 kB)
- teste (mancela - v1, 5/2/17 2:05 PM - 5 kB)
- teste_wrapper (mancela - v1, 5/5/17 2:28 PM - 3 kB)

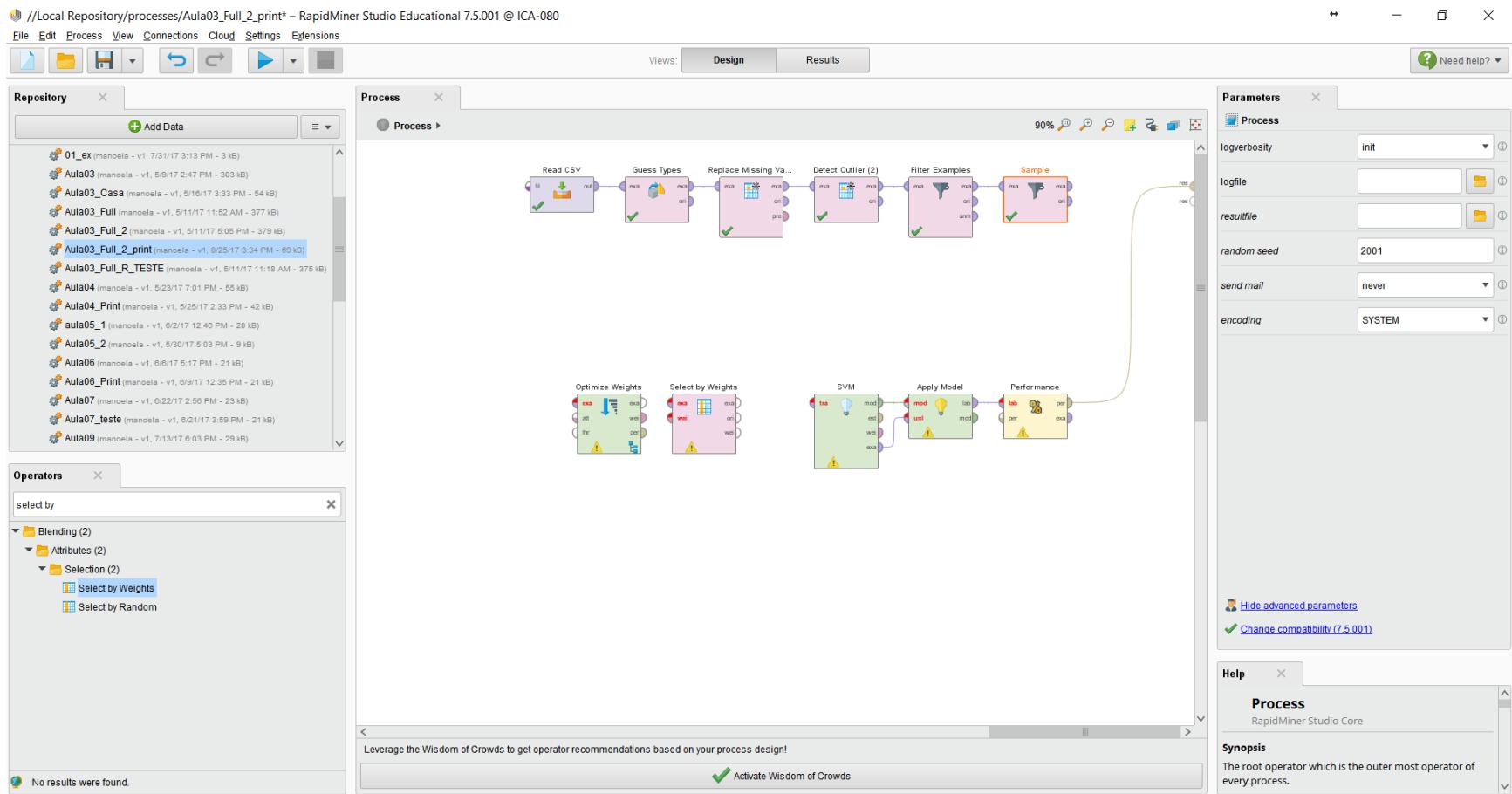
RESULTADOS

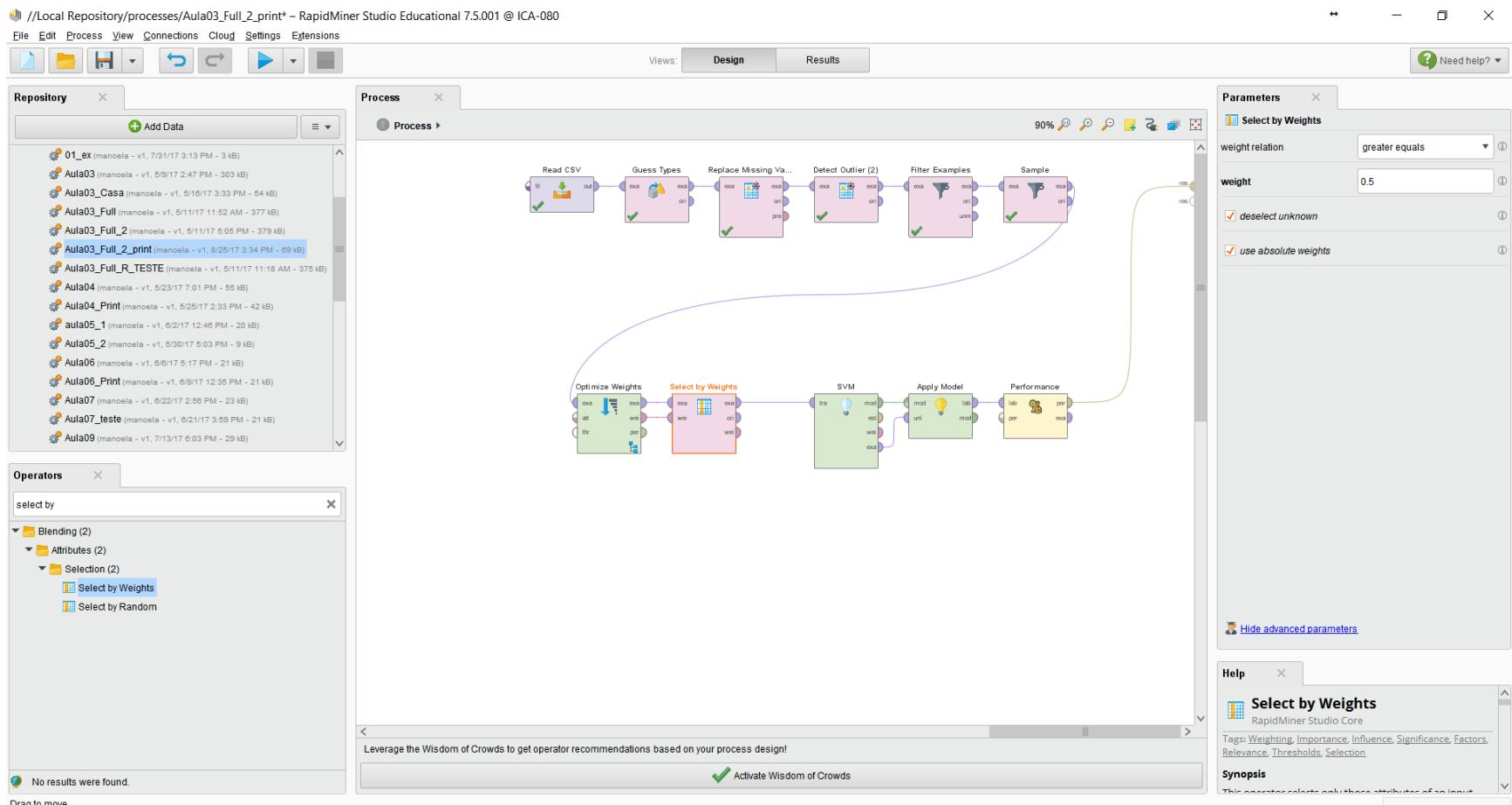
SVM	Acurácia	Kappa	Recall	Precisão
Missing	93.43%	0.018	50.48%	96.71%
+Balanceamento	80.29%	0.606	80.29%	80.31%
+Outliers	86.47%	0.729	86.44%	87.39%











//Local Repository/processes/Aula03_Full_2_print* – RapidMiner Studio Educational 7.5.001 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History AttributeWeights (Optimize Weights) ExampleSet (Optimize Weights)

Data attribute we... ↑

att19	0
att26	0
att155	0
att188	0
att231	0
att239	0
att313	0
att354	0
att371	0
att379	0
att422	0
att539	0
att554	0
att772	0.002
att159	0.002
att581	0.003
att147	0.003
att218	0.004
att497	0.005
att217	0.006
att410	0.011
att440	0.014
att511	0.015
att45	0.015

Repository Add Data

processes (mancela)

- 01_DecisionTree (mancela - v1, 5/2/17 2:04 PM - 2 kB)
- 01_ex (mancela - v1, 7/31/17 3:13 PM - 3 kB)
- Aula03 (mancela - v1, 5/9/17 2:47 PM - 303 kB)
- Aula03_Casa (mancela - v1, 5/16/17 3:33 PM - 54 kB)
- Aula03_Full (mancela - v1, 5/11/17 11:52 AM - 377 kB)
- Aula03_Full_2 (mancela - v1, 5/11/17 5:05 PM - 379 kB)
- Aula03_Full_2_print (mancela - v1, 8/25/17 3:34 PM - 89 kB)
- Aula03_Full_R_TESTE (mancela - v1, 5/11/17 11:18 AM - 1 kB)
- Aula04 (mancela - v1, 5/23/17 7:01 PM - 55 kB)
- Aula04_Print (mancela - v1, 5/25/17 2:33 PM - 42 kB)
- aula05_1 (mancela - v1, 6/2/17 12:46 PM - 20 kB)
- Aula05_2 (mancela - v1, 5/30/17 5:03 PM - 9 kB)
- Aula06 (mancela - v1, 6/6/17 5:17 PM - 21 kB)
- Aula06_Print (mancela - v1, 6/9/17 12:35 PM - 21 kB)
- Aula07 (mancela - v1, 6/22/17 2:56 PM - 23 kB)
- Aula07_teste (mancela - v1, 6/21/17 3:59 PM - 21 kB)
- Aula09 (mancela - v1, 7/13/17 6:03 PM - 29 kB)
- Aula09_2 (mancela - v1, 7/6/17 9:49 AM - 48 kB)
- Aula09_3 (mancela - v1, 7/6/17 12:52 PM - 17 kB)
- aulaBi (mancela - v1, 6/21/17 6:33 PM - 34 kB)
- Copy of Aula07 (mancela - v1, 6/21/17 6:33 PM - 34 kB)
- Copy of Aula09 (mancela - v1, 7/12/17 5:32 PM - 33 kB)
- IBM_AttributeSelection (mancela - v1, 5/3/17 10:58 AM - 1 kB)
- IBM_AttributeSelection (mancela - v1, 5/4/17 12:12 PM - 1 kB)
- IBM_AttributeSelection_2 (mancela - v1, 5/4/17 12:12 PM - 1 kB)
- IBM_JG_WeightSelection (mancela - v1, 5/4/17 12:18 PM - 1 kB)
- IBM_Limpeza (mancela - v1, 5/2/17 4:07 PM - 8 kB)
- IBM_Limpeza_NN_SVM (mancela - v1, 5/2/17 4:34 PM - 2 kB)
- IBM_MissingValues (mancela - v1, 5/4/17 12:12 PM - 13 kB)
- IBM_NeuralNet_DecisionTree (mancela - v1, 5/2/17 3:38 kB)
- New Text Document (mancela - v1, 5/30/17 5:48 PM - 6 kB)
- R (mancela - v1, 5/10/17 6:24 PM - 45 kB)
- teste (mancela - v1, 5/2/17 2:05 PM - 5 kB)

//Local Repository/processes/Aula03_Full_2_print* – RapidMiner Studio Educational 7.5.001 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History ExampleSet (Select by Weights)

ExampleSet (207 examples, 2 special attributes, 307 regular attributes)

Filter (207 / 207 examples): all

Row No.	att501	outlier	att3	att4	att7	att15	att16	att17	att24	att29	att30	att32	att33
1	1	false	1879.559	1087.347	52.216	6.395	213.952	4.657	-2402.914	64.800	1.317	1.849	72.203
2	1	false	1887.048	669.913	56.510	7.968	211.513	5.146	-1173.254	66.323	1.098	1.800	72.203
3	1	false	1887.984	1028.267	53.544	6.121	215.592	5.056	-2297.672	62.516	1.167	1.809	72.774
4	1	false	1912.408	606.919	57.335	6.556	216.858	4.891	-608.744	60.328	1.367	1.778	72.489
5	-1	false	1951.637	805.868	57.313	8.185	220.254	4.600	-2169.866	61.153	1.675	1.760	73.558
6	1	false	1857.170	576.294	57.690	7.811	223.238	4.420	-1228.473	60.328	1.167	1.785	72.489
7	-1	false	1874.373	774.033	54.836	4.618	212.640	4.331	-2252.250	63.352	1.493	1.666	72.924
8	1	false	1910.736	991.033	60.460	3.694	210.499	4.887	-2932.986	65.947	1.236	1.852	72.203
9	-1	false	1864.993	759.115	55.145	3.553	215.731	4.901	-3127.292	68.060	1.154	1.749	86.813
10	1	false	1909.791	982.864	60.887	4.327	214.538	4.777	-870.292	61.304	1.367	1.717	72.088
11	-1	false	1874.373	774.033	54.836	3.892	207.241	5.012	-2292.476	61.947	1.443	1.744	72.472
12	-1	false	1874.373	774.033	54.836	3.433	208.963	5.000	-3825.544	60.124	1.380	1.627	72.974
13	1	false	1862.462	580.857	57.156	3.430	209.871	5.046	-2560.110	66.666	1.449	1.773	71.718
14	1	false	1864.993	759.115	55.145	4.945	214.006	4.890	-2093.272	66.666	1.280	1.747	86.813
15	1	false	1869.635	665.483	56.652	3.219	208.675	5.135	-2369.367	67.138	1.499	1.771	71.864
16	1	false	1868.011	782.912	55.690	4.373	215.036	5.020	-2552.688	65.980	1.098	1.736	86.813
17	1	false	1860.647	772.664	54.967	1.512	203.817	4.976	-564.510	61.196	1.443	1.770	71.595
18	1	false	1911.701	962.545	60.185	2.832	207.738	5.510	-1791.795	61.389	1.424	1.799	72.512
19	1	false	1896.839	796.710	55.595	7.670	209.632	4.578	-3062.127	65.636	1.468	1.811	73.086
20	1	false	1880.515	812.402	56.030	9.516	218.867	4.557	-2510.829	65.754	1.336	1.779	71.864
21	-1	false	1893.572	647.051	57.039	10.719	220.635	4.548	-2351.703	62.902	1.462	1.753	72.483
22	-1	false	1893.572	647.051	57.039	11.503	221.087	4.559	-310.384	59.480	1.719	1.657	71.451
23	-1	false	1884.077	709.159	56.555	4.000	209.796	4.921	-2794.790	68.307	1.531	1.777	73.086
24	1	false	1925.026	763.792	54.748	3.191	207.783	4.824	-2324.390	60.649	1.531	1.692	71.451
25	1	false	1884.077	709.159	56.555	5.122	203.100	5.084	-1152.769	60.574	1.575	1.789	72.924
26	1	false	1919.256	1099.679	56.929	3.846	207.393	4.718	-1234.559	58.172	1.788	1.704	73.174

[1] Process 5:28

Repository Add Data

processes (mancela)

- 01_DecisionTree (mancela - v1, 5/2/17 2:04 PM - 2 kB)
- 01_ex (mancela - v1, 7/5/17 3:13 PM - 3 kB)
- Aula03 (mancela - v1, 5/9/17 2:47 PM - 303 kB)
- Aula03_Casa (mancela - v1, 5/16/17 3:33 PM - 54 kB)
- Aula03_Full (mancela - v1, 5/11/17 11:52 AM - 377 kB)
- Aula03_Full_2 (mancela - v1, 5/11/17 5:05 PM - 379 kB)
- Aula03_Full_2_print (mancela - v1, 8/26/17 3:34 PM - 89 kB)
- Aula03_Full_R_TESTE (mancela - v1, 5/11/17 11:18 AM - 1 kB)
- Aula04 (mancela - v1, 5/23/17 7:01 PM - 55 kB)
- Aula04_Print (mancela - v1, 5/25/17 2:33 PM - 42 kB)
- aula05_1 (mancela - v1, 6/2/17 12:48 PM - 20 kB)
- Aula05_2 (mancela - v1, 5/30/17 5:03 PM - 9 kB)
- Aula06 (mancela - v1, 6/6/17 5:17 PM - 21 kB)
- Aula06_Print (mancela - v1, 6/9/17 12:35 PM - 21 kB)
- Aula07 (mancela - v1, 6/22/17 2:56 PM - 23 kB)
- Aula07_teste (mancela - v1, 6/2/17 3:59 PM - 21 kB)
- Aula09 (mancela - v1, 7/13/17 6:03 PM - 29 kB)
- Aula09_2 (mancela - v1, 7/6/17 9:49 AM - 48 kB)
- Aula09_3 (mancela - v1, 7/6/17 12:52 PM - 17 kB)
- aulaBi (mancela - v1, 7/13/17 6:22 PM - 29 kB)
- Copy of Aula07 (mancela - v1, 6/21/17 6:33 PM - 34 kB)
- Copy of Aula09 (mancela - v1, 7/12/17 5:32 PM - 33 kB)
- IBM_AttributeSelection (mancela - v1, 5/3/17 10:58 AM - 1 kB)
- IBM_AttributeSelection (mancela - v1, 5/4/17 12:12 PM - 1 kB)
- IBM_AttributeSelection_2 (mancela - v1, 5/4/17 12:12 PM - 1 kB)
- IBM_G_WeightingSelection (mancela - v1, 5/4/17 12:18 PM - 1 kB)
- IBM_Limpeza (mancela - v1, 5/2/17 4:07 PM - 8 kB)
- IBM_Limpeza_NN_SVM (mancela - v1, 5/2/17 4:34 PM - 2 kB)
- IBM_MissingValues (mancela - v1, 5/4/17 12:12 PM - 13 kB)
- IBM_NeuralNet_DecisionTree (mancela - v1, 5/2/17 3:38 PM - 6 kB)
- New Text Document (mancela - v1, 5/30/17 5:48 PM - 6 kB)
- R (mancela - v1, 5/10/17 6:24 PM - 45 kB)
- teste (mancela - v1, 5/2/17 2:05 PM - 5 kB)

//Local Repository/processes/Aula03_Full_2_print* – RapidMiner Studio Educational 7.5.001 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History PerformanceVector (Performance)

Criterion accuracy

accuracy: 84.54%

Table View Plot View

	true -1	true 1	class precision
pred. -1	100	28	78.12%
pred. 1	4	75	94.94%
class recall	96.15%	72.82%	

Description Annotations

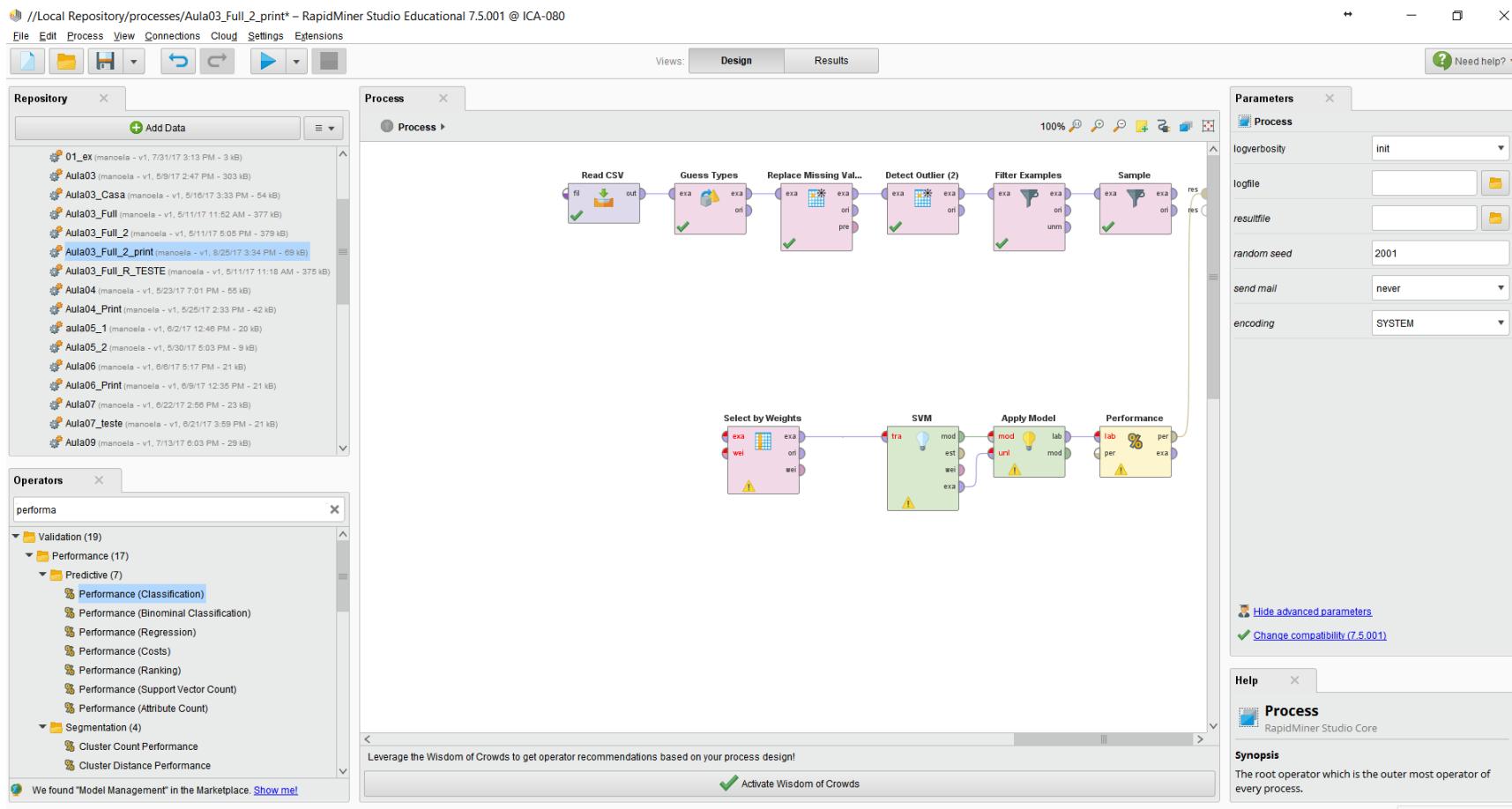
Repository

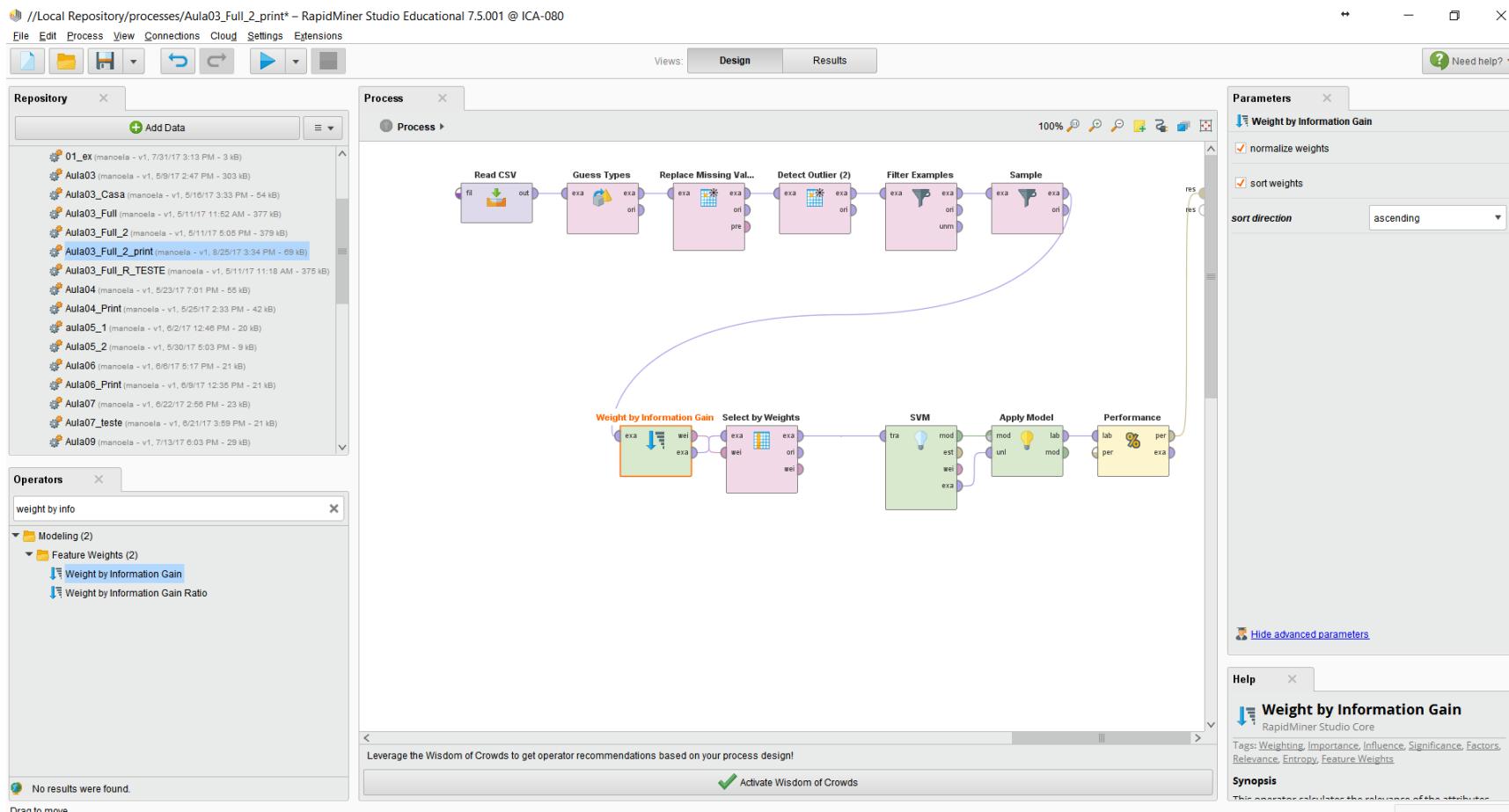
Add Data

- processes (mancela)
 - 01_DecisionTree (mancela - v1, 5/2/17 2:04 PM - 2 kB)
 - 01_ex (mancela - v1, 7/31/17 3:13 PM - 3 kB)
 - Aula03 (mancela - v1, 5/9/17 2:47 PM - 303 kB)
 - Aula03_Casa (mancela - v1, 5/16/17 3:33 PM - 54 kB)
 - Aula03_Full (mancela - v1, 5/11/17 11:52 AM - 377 kB)
 - Aula03_Full_2 (mancela - v1, 5/11/17 5:05 PM - 379 kB)
 - Aula03_Full_2_print (mancela - v1, 8/26/17 3:34 PM - 89 kB)
 - Aula03_Full_R_TESTE (mancela - v1, 5/11/17 11:18 AM - 1 kB)
 - Aula04 (mancela - v1, 5/23/17 7:01 PM - 55 kB)
 - Aula04_Print (mancela - v1, 5/25/17 2:33 PM - 42 kB)
 - aula05_1 (mancela - v1, 6/2/17 12:46 PM - 20 kB)
 - Aula05_2 (mancela - v1, 5/30/17 5:03 PM - 9 kB)
 - Aula06 (mancela - v1, 6/6/17 5:17 PM - 21 kB)
 - Aula06_Print (mancela - v1, 6/9/17 12:35 PM - 21 kB)
 - Aula07 (mancela - v1, 6/22/17 2:56 PM - 23 kB)
 - Aula07_teste (mancela - v1, 6/2/17 3:59 PM - 21 kB)
 - Aula09 (mancela - v1, 7/13/17 6:03 PM - 29 kB)
 - Aula09_2 (mancela - v1, 7/6/17 9:49 AM - 48 kB)
 - Aula09_3 (mancela - v1, 7/6/17 12:52 PM - 17 kB)
 - aulaBi (mancela - v1, 6/21/17 6:33 PM - 34 kB)
 - Copy of Aula07 (mancela - v1, 6/21/17 6:33 PM - 34 kB)
 - Copy of Aula09 (mancela - v1, 7/12/17 5:32 PM - 33 kB)
 - IBM_AttributeSelection (mancela - v1, 5/3/17 10:58 AM - 1 kB)
 - IBM_AttributeSelection (mancela - v1, 5/4/17 12:12 PM - 1 kB)
 - IBM_AttributeSelection_2 (mancela - v1, 5/4/17 12:12 PM - 1 kB)
 - IBM_JG_WeightSelection (mancela - v1, 5/4/17 12:18 PM - 1 kB)
 - IBM_Limpeza (mancela - v1, 5/2/17 4:07 PM - 8 kB)
 - IBM_Limpeza_NN_SVM (mancela - v1, 5/2/17 4:34 PM - 2 kB)
 - IBM_MissingValues (mancela - v1, 5/4/17 12:12 PM - 13 kB)
 - IBM_NeuralNet_DecisionTree (mancela - v1, 5/2/17 3:38 kB)
 - New Text Document (mancela - v1, 5/30/17 5:48 PM - 6 kB)
 - R (mancela - v1, 5/10/17 6:24 PM - 45 kB)
 - teste (mancela - v1, 5/2/17 2:05 PM - 5 kB)

RESULTADOS

SVM	Acurácia	Kappa	Recall	Precisão
Missing	93.43%	0.018	50.48%	96.71%
+Balanceamento	80.29%	0.606	80.29%	80.31%
+Outliers	86.47%	0.729	86.44%	87.39%
+Wrapper	84.54%	0.690	84.48%	86.53%





//Local Repository/processes/Aula03_Full_2_print* – RapidMiner Studio Educational 7.5.001 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History ExampleSet (Sample) AttributeWeights (Weight by Information Gain)

Data attribute weight

attribute	weight
att60	1
att421	0.968
att148	0.952
att283	0.938
att291	0.657
att65	0.606
att132	0.604
att66	0.596
att478	0.592
att12	0.588
att156	0.573
att429	0.573
att349	0.570
att84	0.558
att104	0.556
att154	0.556
att289	0.556
att427	0.556
att211	0.538
att206	0.515
att342	0.515
att131	0.502
att435	0.465
att432	0.450
...	...

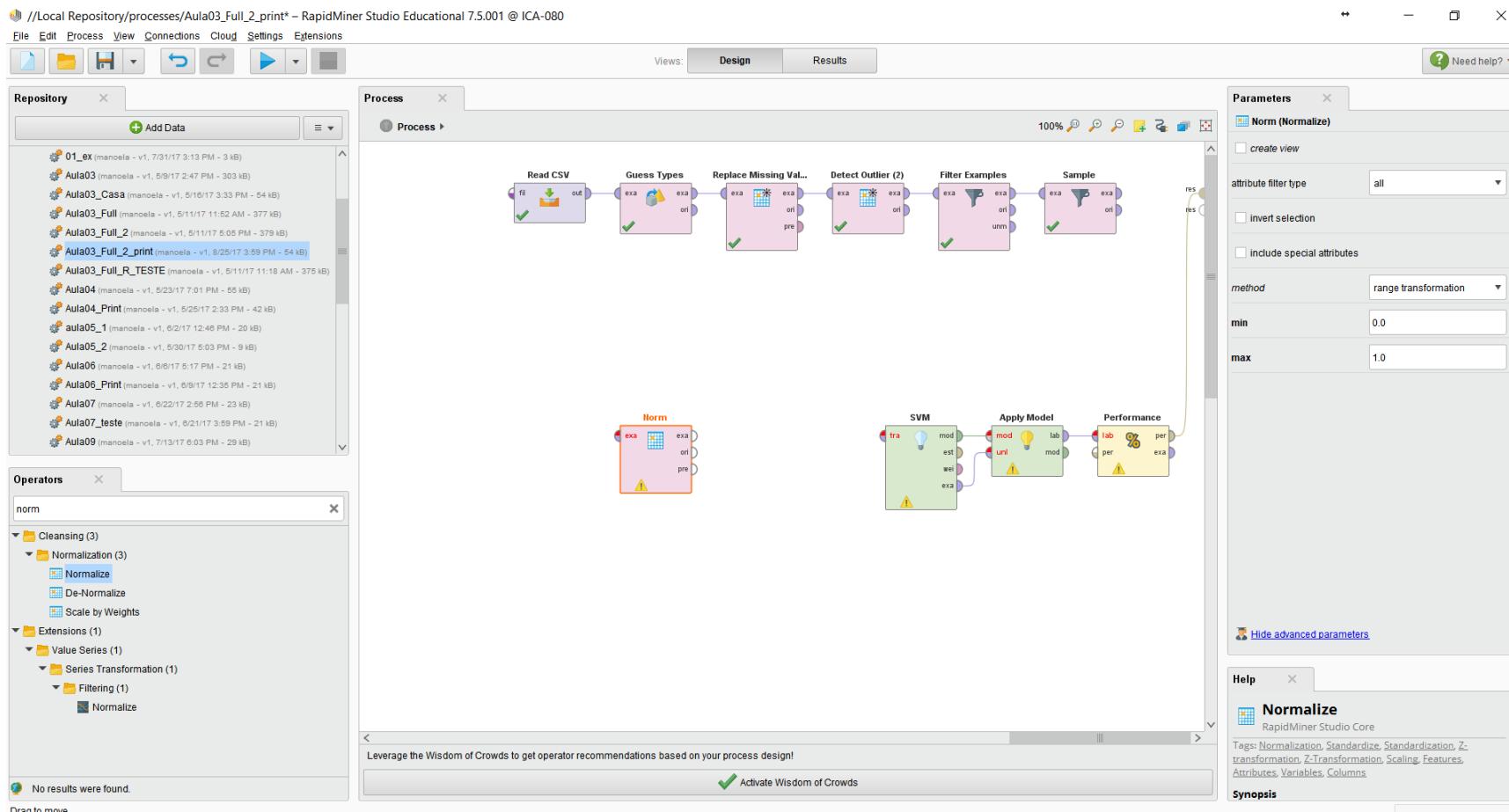
Repository

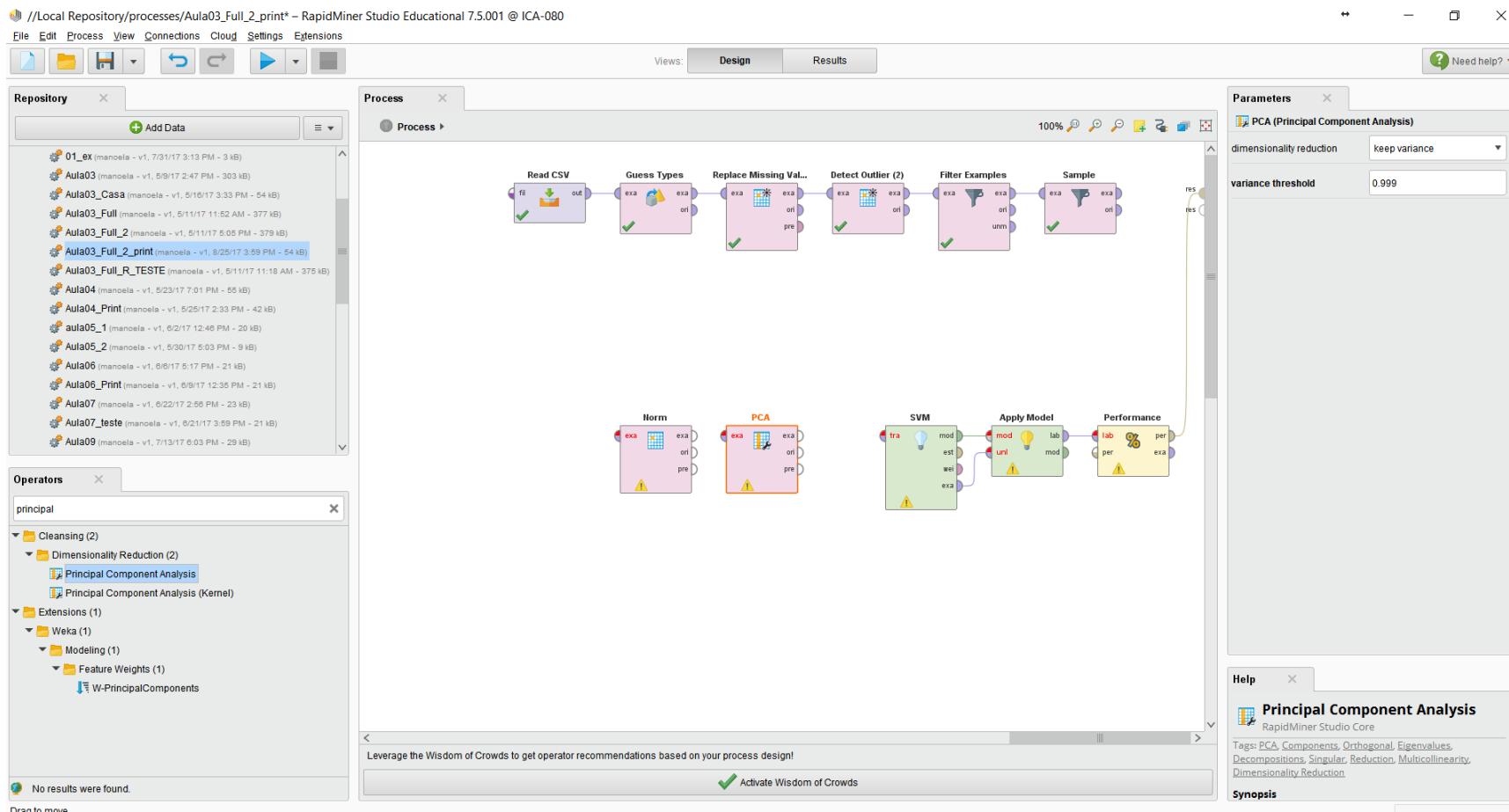
Add Data

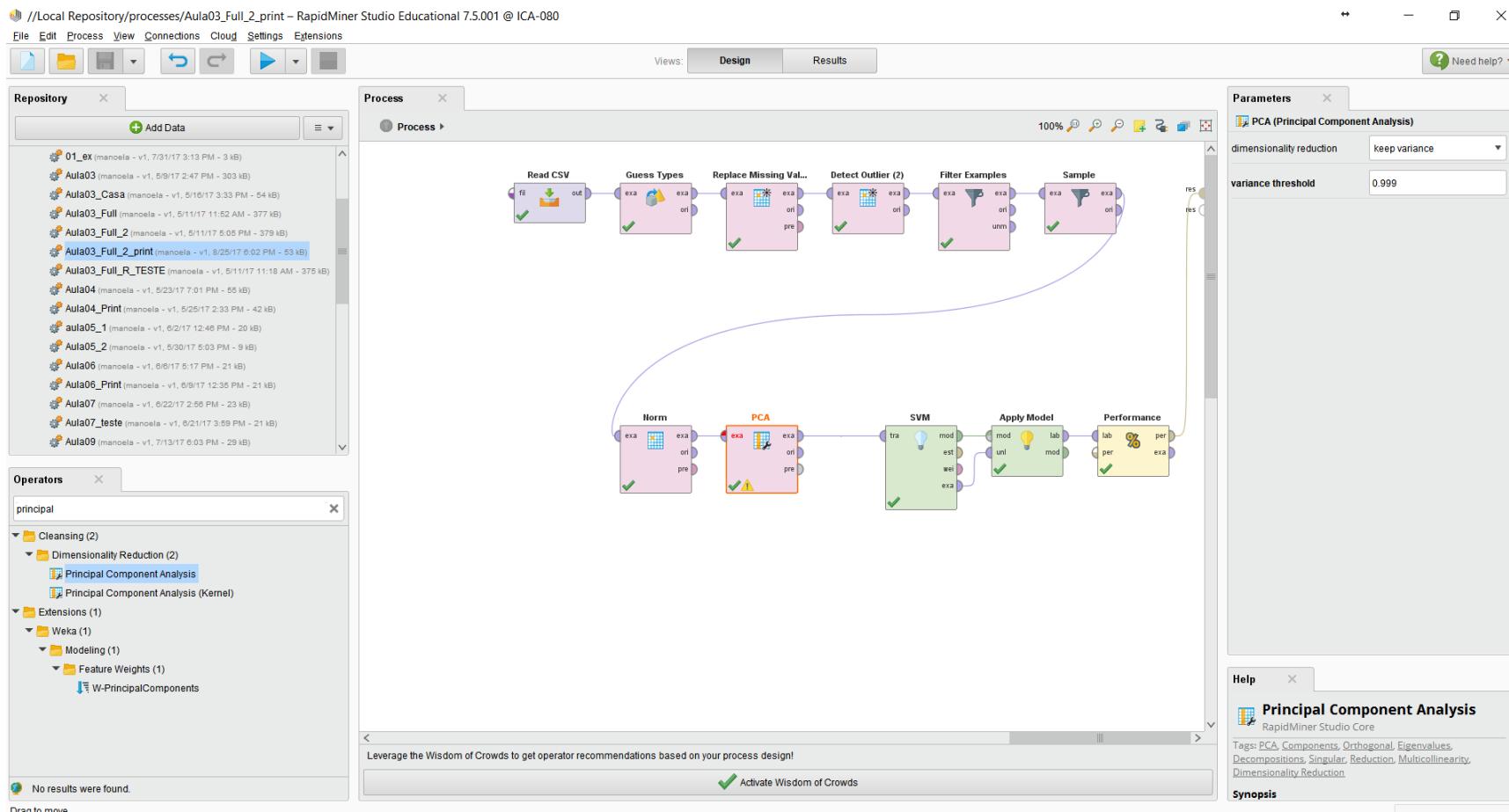
- processes (mancela)
 - 01_DecisionTree (mancela - v1, 5/2/17 2:04 PM - 2 kB)
 - 01_ex (mancela - v1, 7/31/17 3:13 PM - 3 kB)
 - Aula03 (mancela - v1, 5/9/17 2:47 PM - 303 kB)
 - Aula03_Casa (mancela - v1, 5/16/17 3:33 PM - 54 kB)
 - Aula03_Full (mancela - v1, 5/11/17 11:52 AM - 377 kB)
 - Aula03_Full_2 (mancela - v1, 5/11/17 5:05 PM - 379 kB)
 - Aula03_Full_2_print (mancela - v1, 8/26/17 3:34 PM - 89 kB)
 - Aula03_Full_R_TESTE (mancela - v1, 5/11/17 11:18 AM - 1 kB)
 - Aula04 (mancela - v1, 5/23/17 7:01 PM - 55 kB)
 - Aula04_Print (mancela - v1, 5/25/17 2:33 PM - 42 kB)
 - aula05_1 (mancela - v1, 6/2/17 12:46 PM - 20 kB)
 - Aula05_2 (mancela - v1, 5/30/17 5:03 PM - 9 kB)
 - Aula06 (mancela - v1, 6/6/17 5:17 PM - 21 kB)
 - Aula06_Print (mancela - v1, 6/9/17 12:35 PM - 21 kB)
 - Aula07 (mancela - v1, 6/22/17 2:56 PM - 23 kB)
 - Aula07_teste (mancela - v1, 6/2/17 3:59 PM - 21 kB)
 - Aula09 (mancela - v1, 7/13/17 6:03 PM - 29 kB)
 - Aula09_2 (mancela - v1, 7/6/17 9:49 AM - 48 kB)
 - Aula09_3 (mancela - v1, 7/6/17 12:52 PM - 17 kB)
 - aulaBi (mancela - v1, 6/21/17 6:33 PM - 34 kB)
 - Copy of Aula07 (mancela - v1, 7/12/17 6:33 PM - 34 kB)
 - Copy of Aula09 (mancela - v1, 7/12/17 5:32 PM - 33 kB)
 - IBM_AttributeSelection (mancela - v1, 5/3/17 10:58 AM - 1 kB)
 - IBM_AttributeSelection (mancela - v1, 5/4/17 12:12 PM - 1 kB)
 - IBM_AttributeSelection_2 (mancela - v1, 5/4/17 12:12 PM - 1 kB)
 - IBM_JG_WeightSelection (mancela - v1, 5/4/17 12:18 PM - 1 kB)
 - IBM_Limpeza (mancela - v1, 5/2/17 4:07 PM - 8 kB)
 - IBM_Limpeza_NN_SVM (mancela - v1, 5/2/17 4:34 PM - 2 kB)
 - IBM_MissingValues (mancela - v1, 5/4/17 12:12 PM - 13 kB)
 - IBM_NeuralNet_DecisionTree (mancela - v1, 5/2/17 3:38 kB)
 - New Text Document (mancela - v1, 5/30/17 5:48 PM - 6 kB)
 - R (mancela - v1, 5/10/17 6:24 PM - 45 kB)
 - teste (mancela - v1, 5/2/17 2:05 PM - 5 kB)

RESULTADOS

SVM	Acurácia	Kappa	Recall	Precisão
Missing	93.43%	0.018	50.48%	96.71%
+Balanceamento	80.29%	0.606	80.29%	80.31%
+Outliers	86.47%	0.729	86.44%	87.39%
Wrapper	84.54%	0.690	84.48%	86.53%
Ranker	71.98%	0.439	71.91%	74.30%







//Local Repository/processes/Aula03_Full_2_print – RapidMiner Studio Educational 7.5.001 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History PerformanceVector (Performance)

Criterion accuracy

accuracy: 98.55%

	true -1	true 1	class precision
pred. -1	102	1	99.03%
pred. 1	2	102	98.08%
class recall	98.08%	99.03%	

Description Annotations

Repository

Add Data

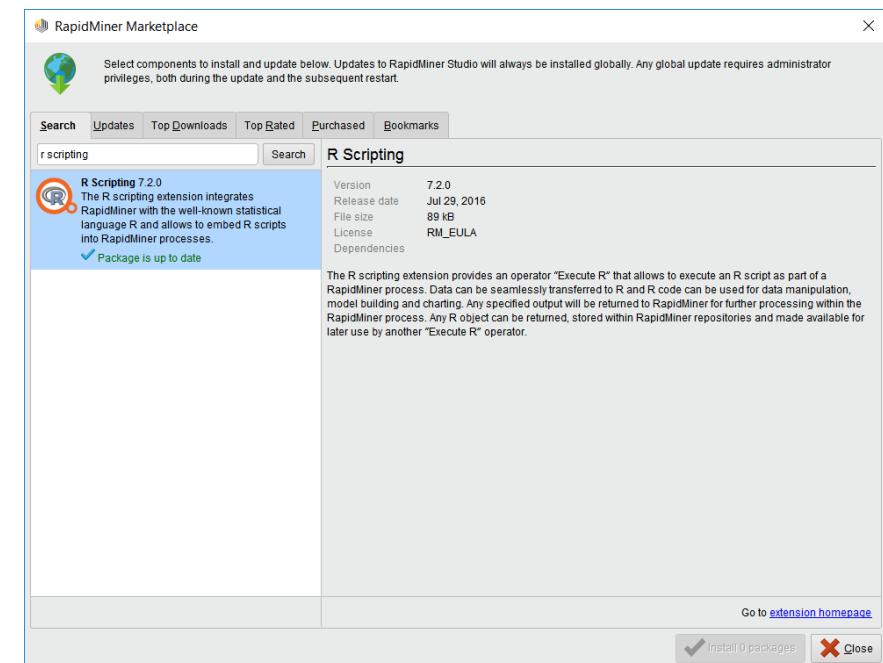
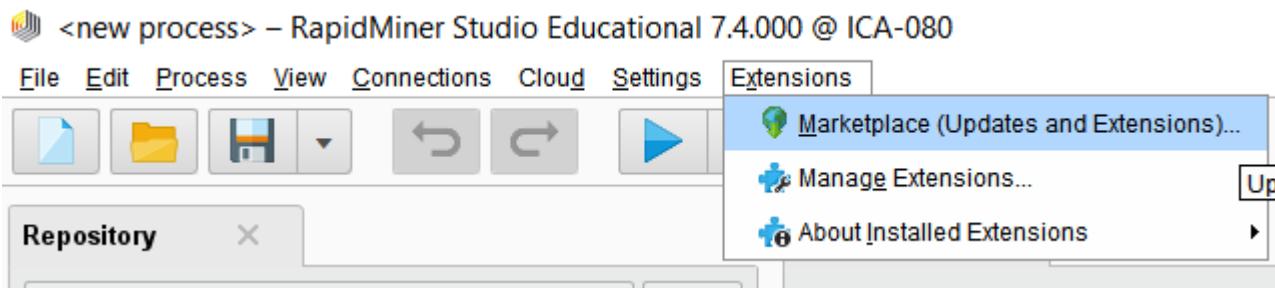
- processes (mancela)
 - 01_DecisionTree (mancela - v1, 5/2/17 2:04 PM - 2 kB)
 - 01_ex (mancela - v1, 7/31/17 3:13 PM - 3 kB)
 - Aula03 (mancela - v1, 5/9/17 2:47 PM - 303 kB)
 - Aula03_Casa (mancela - v1, 5/16/17 3:33 PM - 54 kB)
 - Aula03_Full (mancela - v1, 5/11/17 11:52 AM - 377 kB)
 - Aula03_Full_2 (mancela - v1, 5/11/17 5:05 PM - 379 kB)
 - Aula03_Full_2_print (mancela - v1, 8/25/17 6:02 PM - 53 kB)**
 - Aula03_Full_R_TESTE (mancela - v1, 5/11/17 11:18 AM - 1 kB)
 - Aula04 (mancela - v1, 5/23/17 7:01 PM - 55 kB)
 - Aula04_Print (mancela - v1, 5/25/17 2:33 PM - 42 kB)
 - aula05_1 (mancela - v1, 6/2/17 12:46 PM - 20 kB)
 - Aula05_2 (mancela - v1, 5/30/17 5:03 PM - 9 kB)
 - Aula06 (mancela - v1, 6/6/17 5:17 PM - 21 kB)
 - Aula06_Print (mancela - v1, 6/9/17 12:35 PM - 21 kB)
 - Aula07 (mancela - v1, 6/22/17 2:56 PM - 23 kB)
 - Aula07_teste (mancela - v1, 6/2/17 3:59 PM - 21 kB)
 - Aula09 (mancela - v1, 7/13/17 6:03 PM - 29 kB)
 - Aula09_2 (mancela - v1, 7/6/17 9:49 AM - 48 kB)
 - Aula09_3 (mancela - v1, 7/6/17 12:52 PM - 17 kB)
 - aulaBi (mancela - v1, 6/21/17 6:33 PM - 34 kB)
 - Copy of Aula07 (mancela - v1, 6/21/17 6:33 PM - 34 kB)
 - Copy of Aula09 (mancela - v1, 7/12/17 5:32 PM - 33 kB)
 - IBM_AttributeSelection (mancela - v1, 5/3/17 10:58 AM - 1 kB)
 - IBM_AttributeSelection (mancela - v1, 5/4/17 12:12 PM - 1 kB)
 - IBM_AttributeSelection_2 (mancela - v1, 5/4/17 12:12 PM - 1 kB)
 - IBM_JG_WeightSelection (mancela - v1, 5/4/17 12:18 PM - 1 kB)
 - IBM_Limpeza (mancela - v1, 5/2/17 4:07 PM - 8 kB)
 - IBM_Limpeza_NN_SVM (mancela - v1, 5/2/17 4:34 PM - 2 kB)
 - IBM_MissingValues (mancela - v1, 5/4/17 12:12 PM - 13 kB)
 - IBM_NeuralNet_DecisionTree (mancela - v1, 5/2/17 3:38 kB)
 - New Text Document (mancela - v1, 5/30/17 5:48 PM - 6 kB)
 - R (mancela - v1, 5/10/17 6:24 PM - 45 kB)
 - teste (mancela - v1, 5/2/17 2:05 PM - 5 kB)

RESULTADOS

SVM	Acurácia	Kappa	Recall	Precisão
Missing	93.43%	0.018	50.48%	96.71%
+Balanceamento	80.29%	0.606	80.29%	80.31%
+Outliers	86.47%	0.729	86.44%	87.39%
Wrapper	84.54%	0.690	84.48%	86.53%
Ranker	71.98%	0.439	71.91%	74.30%
PCA	98.55%	0.971	98.55%	98.55%

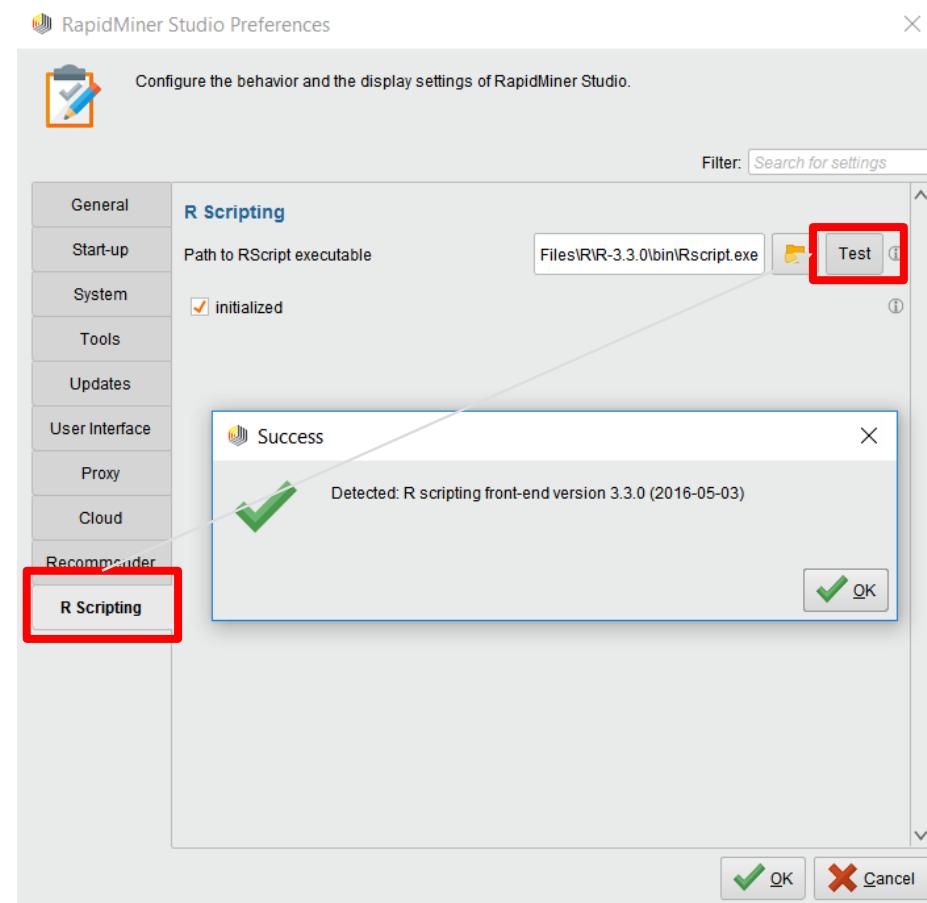
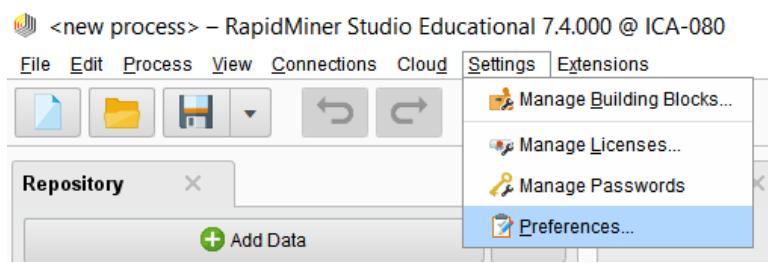
RAPIDMINER

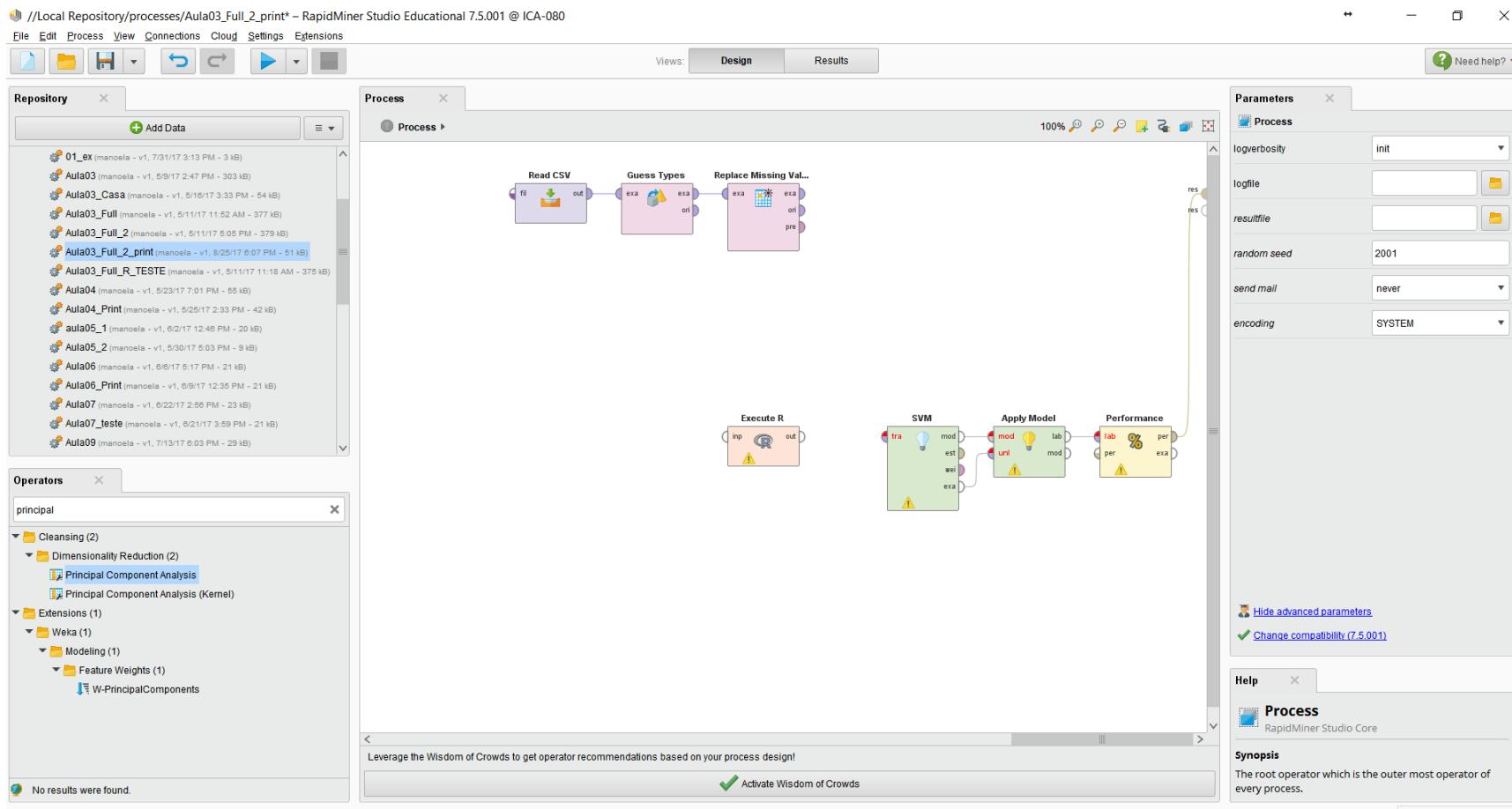
Extensão para integrar o R ao Rapidminer

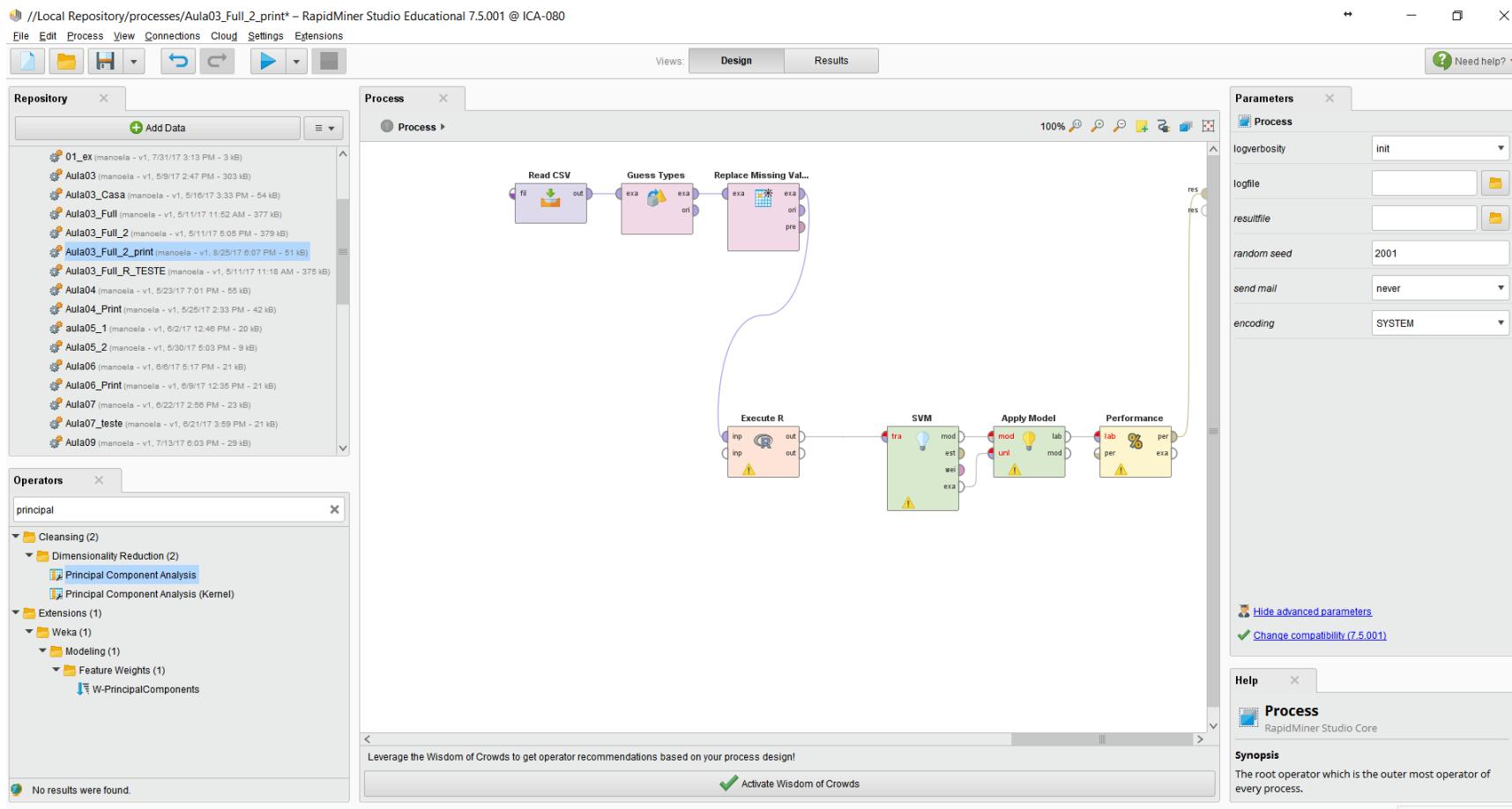


RAPIDMINER

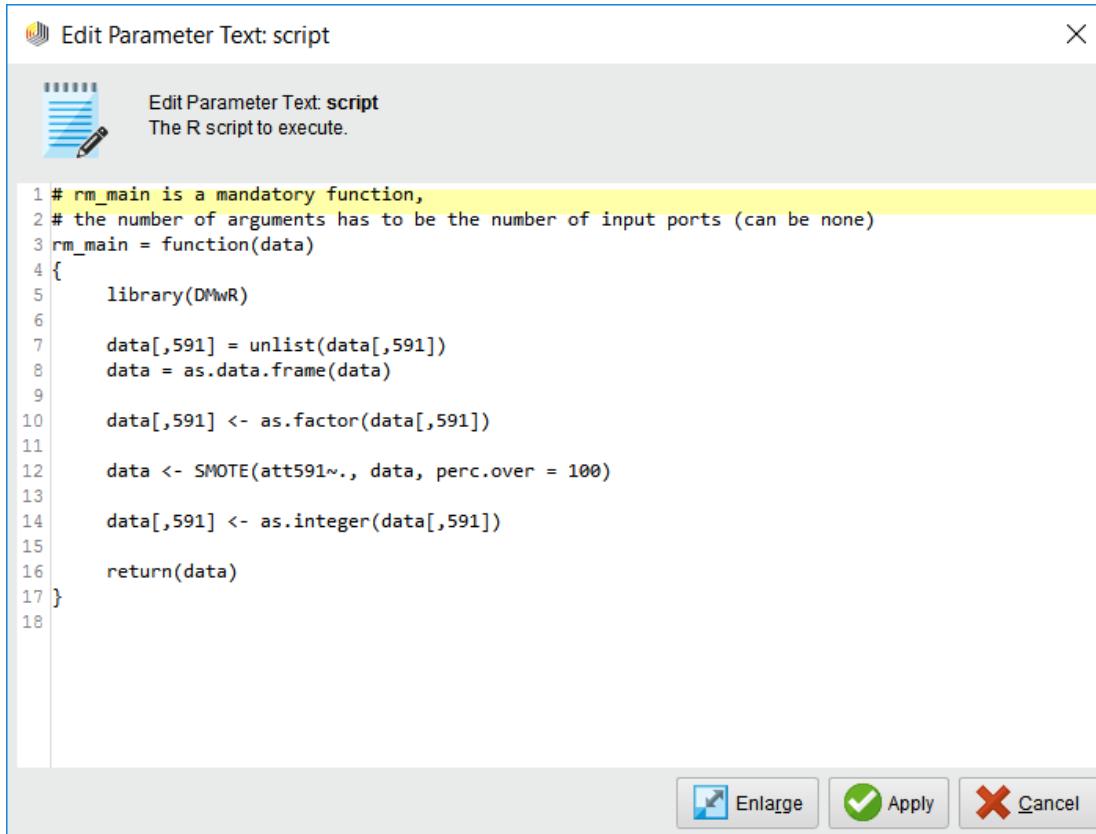
Testar a integração







RAPIDMINER



```
rm_main = function(data)
{
  library(DMwR)

  data[,591] = unlist(data[,591])
  data = as.data.frame(data)

  data[,591] <- as.factor(data[,591])

  data <- SMOTE(att591~, data, perc.over = 100)

  data[,591] <- as.integer(data[,591])

  return(data)
}
```

//Local Repository/processes/Aula03_Full_2_print – RapidMiner Studio Educational 7.5.001 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History PerformanceVector (Performance)

Criterion accuracy

accuracy: 90.87%

Table View Plot View

	true 1	true 2	class precision
pred. 1	195	25	88.64%
pred. 2	13	183	93.37%
class recall	93.75%	87.98%	

Description Annotations

Repository

Add Data

- processes (mancela)
 - 01_DecisionTree (mancela - v1, 5/2/17 2:04 PM - 2 kB)
 - 01_ex (mancela - v1, 7/31/17 3:13 PM - 3 kB)
 - Aula03 (mancela - v1, 5/9/17 2:47 PM - 303 kB)
 - Aula03_Casa (mancela - v1, 5/16/17 3:33 PM - 54 kB)
 - Aula03_Full (mancela - v1, 5/11/17 11:52 AM - 377 kB)
 - Aula03_Full_2 (mancela - v1, 5/11/17 5:05 PM - 379 kB)
 - Aula03_Full_2_print (mancela - v1, 8/25/17 6:09 PM - 49 kB)
 - Aula03_Full_R_TESTE (mancela - v1, 5/11/17 11:18 AM - 1 kB)
 - Aula04 (mancela - v1, 5/23/17 7:01 PM - 55 kB)
 - Aula04_Print (mancela - v1, 5/25/17 2:33 PM - 42 kB)
 - aula05_1 (mancela - v1, 6/2/17 12:46 PM - 20 kB)
 - Aula05_2 (mancela - v1, 5/30/17 5:03 PM - 9 kB)
 - Aula06 (mancela - v1, 6/6/17 5:17 PM - 21 kB)
 - Aula06_Print (mancela - v1, 6/9/17 12:35 PM - 21 kB)
 - Aula07 (mancela - v1, 6/22/17 2:56 PM - 23 kB)
 - Aula07_teste (mancela - v1, 6/2/17 3:59 PM - 21 kB)
 - Aula09 (mancela - v1, 7/13/17 6:03 PM - 29 kB)
 - Aula09_2 (mancela - v1, 7/6/17 9:49 AM - 48 kB)
 - Aula09_3 (mancela - v1, 7/6/17 12:52 PM - 17 kB)
 - aulaBi (mancela - v1, 6/21/17 6:33 PM - 34 kB)
 - Copy of Aula07 (mancela - v1, 7/12/17 5:32 PM - 33 kB)
 - Copy of Aula09 (mancela - v1, 7/12/17 10:58 AM - 12 kB)
 - IBM_AttributeSelection (mancela - v1, 5/3/17 10:58 AM - 1 kB)
 - IBM_AttributeSelection (mancela - v1, 5/4/17 12:12 PM - 1 kB)
 - IBM_AttributeSelection_2 (mancela - v1, 5/4/17 12:12 PM - 1 kB)
 - IBM_JG_WeightSelection (mancela - v1, 5/4/17 12:18 PM - 1 kB)
 - IBM_Limpeza (mancela - v1, 5/2/17 4:07 PM - 8 kB)
 - IBM_Limpeza_NN_SVM (mancela - v1, 5/2/17 4:34 PM - 2 kB)
 - IBM_MissingValues (mancela - v1, 5/4/17 12:12 PM - 13 kB)
 - IBM_NeuralNet_DecisionTree (mancela - v1, 5/2/17 3:38 kB)
 - New Text Document (mancela - v1, 5/30/17 5:48 PM - 6 kB)
 - R (mancela - v1, 5/10/17 6:24 PM - 45 kB)
 - teste (mancela - v1, 5/2/17 2:05 PM - 5 kB)

RESULTADOS

SVM	Acurácia	Kappa	Recall	Precisão
Missing	93.43%	0.018	50.48%	96.71%
+Balanceamento	80.29%	0.606	80.29%	80.31%
+Outliers	86.47%	0.729	86.44%	87.39%
Wrapper	84.54%	0.690	84.48%	86.53%
Ranker	71.98%	0.439	71.91%	74.30%
PCA	98.55%	0.971	98.55%	98.55%
SMOTE	90.87%	0.817	90.87%	91.00%

R Script