

# Agentes e Inteligência Artificial Distribuída

up201604503 Tomás Nuno Fernandes Novo  
up201604828 João Pedro Viveiros Franco

**1ª Parte**

**Apresentação**





## Descrição do problema e experiências realizadas

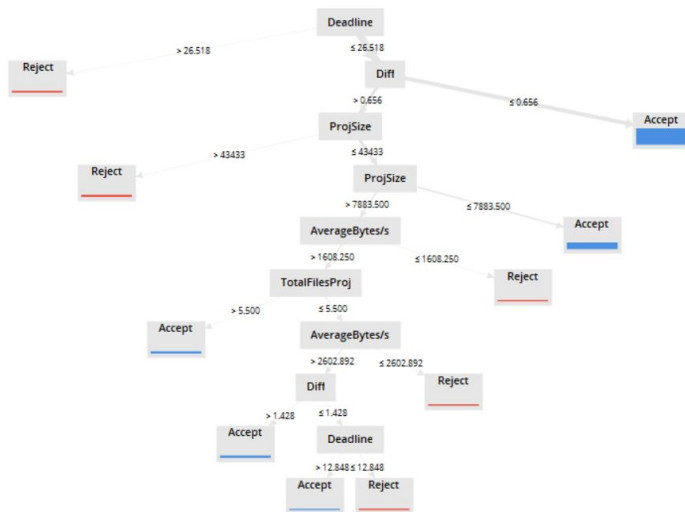
- ❑ No âmbito desta unidade curricular, foi-nos proposto, como 2º trabalho prático, a realização de um projeto cuja meta residia em analisar os dados de execução do 1º projeto. Descrevendo um pouco as idiossincrasias do projeto implementado, foi pedido analisar os dados gerados e analisar dois tipos de problemas: um problema de **classificação** e um problema de **regressão linear**.
- ❑ O problema de **classificação** consistia em tentar prever a aceitação, ou não aceitação, da re-negociação enviada pelo agente CPU, sendo esta **decisão** a variável **dependente** a ser analisada pelo agente Cliente usando uma árvore de decisão. As variáveis independentes são, então, o **tamanho**, em bytes, do projeto, o **número de ficheiros**, a **média** da velocidade de compilação dos últimos 30 projetos, o **deadline** do projeto e a **diferença** entre o tempo estimado pelo CPU para compilar e o deadline.
- ❑ Para a **regressão linear** a variável dependente é a diferença referida no problema de classificação enquanto que as variáveis independentes são as restantes (menos a decisão).



# Análise de Dados

- Para o problema de **classificação**, foram considerados os seguintes **resultados** e **estatísticas** gerados pelo **RapidMiner™**:

## Árvore de decisão:



## Estatísticas:

### PerformanceVector

```
PerformanceVector:
accuracy: 72.97%
ConfusionMatrix:
True:  Accept  Reject
Accept: 25      4
Reject: 6       2
classification_error: 27.03%
ConfusionMatrix:
True:  Accept  Reject
Accept: 25      4
Reject: 6       2
absolute_error: 0.264 +/- 0.413
correlation: 0.125
```



# Análise de Dados

- ❑ Para o problema de **regressão**, foram considerados os seguintes **resultados** e **estatísticas** gerados pelo **RapidMiner™**:

Tabela de Previsões:

Row No.	Diff	prediction(D...	ProjSize	TotalFilesProj	AverageByt...	Deadline	Compilation...
1	0	5.883	1514	1	1158.252	0.294	1.307
2	1.409	5.622	144	1	4340.318	0.598	0.362
3	8.052	4.713	2491	2	3924.108	0.897	2.051
4	0.201	4.419	11732	1	3563.787	4.758	1.480
5	0	-0.625	34963	4	3750.746	14.180	7.324
6	0.084	2.869	18697	3	3824.025	6.424	3.587
7	0	-0.112	48163	4	5500.054	23.913	2.563
8	0.189	2.165	20472	2	3705.346	6.772	4.773
9	6.976	4.218	5472	2	5523.788	1.971	1.926
10	0	-0.841	47550	5	4354.650	17.131	6.510
11	6.450	3.205	7343	3	5160.515	2.110	3.828
12	0.637	1.456	28756	3	5294.783	9.602	4.467
13	0	0.142	44433	3	8670.299	22.061	1.403
14	0	2.120	6760	5	7359.041	2.417	4.749
15	0.142	1.724	12691	6	2314.700	5.108	6.663
16	0	2.539	9388	3	6626.067	3.134	4.077
17	0.294	3.681	5977	3	2205.422	2.558	4.056

Estatísticas:

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value
AverageBytes/s	-0.000	0.000	-0.140	1.000	-1.221	0.226
Deadline	-0.181	0.145	-0.165	0.727	-1.253	0.214
CompilationTime	-0.564	0.331	-0.224	0.825	-1.705	0.093
(Intercept)	6.942	1.625	?	?	4.271	0.000

Função resultante:

$$Diff = -0.181 * Deadline - 0.564 * CompilationTime + 6.942$$



# Análise de Dados

- ❑ Como foi possível observar, a precisão dos métodos utilizados não é a melhor, o que é expectável dada a natureza estocástica do problema de distribuição dos ficheiros de compilação por cada CPU. A natureza referida é a resultante de várias nuances como a **velocidade de compilação** de cada CPU, a **versão do compilador** e o **projeto** do Cliente, sendo que estes fatores afetam direta e drasticamente as relações entre as variáveis analisadas no **RapidMiner™**.
- ❑ A árvore de decisão alcançou apenas 72% de precisão e a regressão linear apresenta valores altos de erro padrão e previsões pouco consistentes.



# Conclusões

- ❑ Em suma, a realização deste projeto foi benéfico para os elementos constituintes do grupo devido à aquisição de conhecimentos acerca de *machine learning*, além de que a manipulação do **RapidMiner™** foi bastante interessante.
- ❑ A implementação do projeto encontra-se em concordância com o discutido com o professor e foi efetuada com sucesso, pelo que achamos que a sua concepção foi bastante positiva.

## 2ª Parte

# Informação Adicional







## Outras experiências

- ❑ Como experiências extra, foi realizada a modelação dos dados no **RapidMiner™** com o método **auto-model**, o que possibilitou uma maior compreensão dos dados em relação a variáveis como o valor de correlação, o *id-ness*, etc, o que nos possibilitou, por exemplo, a deteção do baixo grau de correlação no modelo referido por parte da variável **CompilationTime**.
- ❑ O método usado executa a modelação dos dados em diferentes algoritmos de classificação, tal como **random forest**, **support vector machines**, entre outras e selecciona o melhor tendo em conta parâmetros como a **precisão**, a **velocidade de cálculo**, etc.
- ❑ O algoritmo proposto pelo programa foi o classificador de **Naive-Bayes**, com uma precisão de 89%, que é significativamente superior à da árvore de decisão de apenas 72%.



# Outras experiências

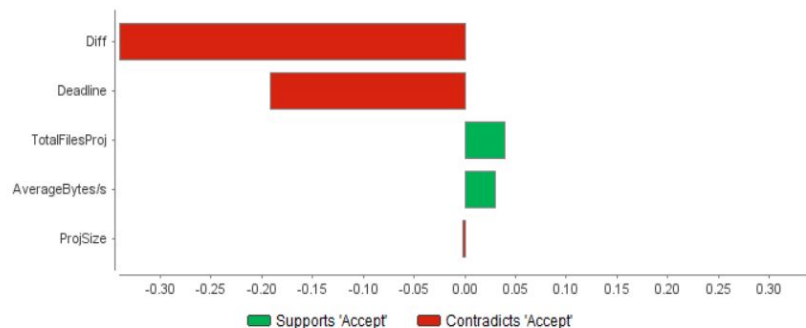
## Naive Bayes - Weights

Attribute	Weight
Deadline	0.094
Diff	0.035
ProjSize	0.028
TotalFilesProj	0.013
AverageBytes/s	0.011

## Performances

Criterion	Value	Standard Deviation
Accuracy	88.6%	± 6.4%
Classification Error	11.4%	± 6.4%
AUC	66.0%	± 38.8%
Precision	84.8%	± 9.4%
Recall	100.0%	± 0.0%
F Measure	91.6%	± 5.3%
Sensitivity	100.0%	± 0.0%
Specificity	?	?

## Important Factors for **Accept**





# Outras experiências

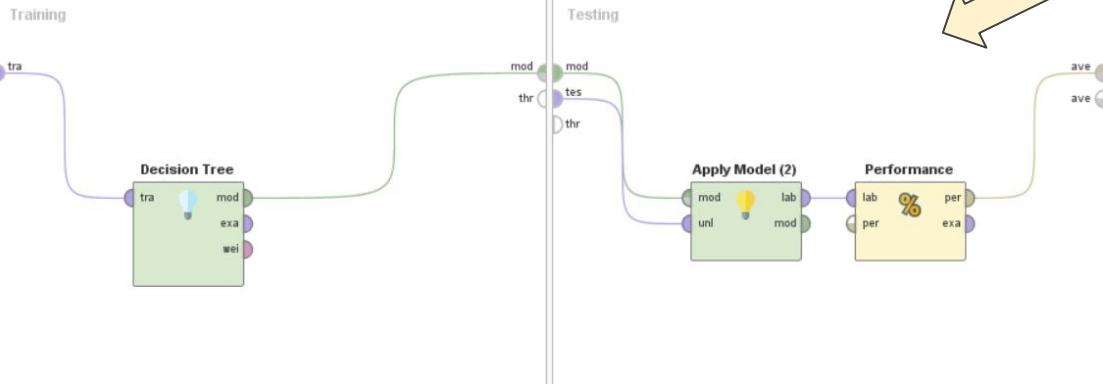
## Naive Bayes - Predictions

Row No.	Decision	prediction(D...	confidence(...	confidence(...	cost	ProjSize	TotalFilesProj	AverageByt...	Deadline	Diff ↓
16	Accept	Accept	0.324	0.676	0.352	3265	2	2954.700	1.090	0
18	Accept	Accept	0.324	0.676	0.352	1417	1	2637.959	0.574	0
19	Accept	Accept	0.324	0.676	0.352	2138	1	1356.048	3.496	0
20	Accept	Accept	0.157	0.843	0.686	6628	1	3505.844	10.837	0
21	Accept	Accept	0.324	0.676	0.352	2242	2	849.289	3.666	0
22	Reject	Reject	0.657	0.343	0.314	35686	7	1500.073	25.337	0
23	Accept	Accept	0.157	0.843	0.686	44433	3	16085.454	22.061	0
25	Accept	Accept	0.324	0.676	0.352	160	1	1190.164	0.665	0
29	Reject	Accept	0.324	0.676	0.352	32689	4	3589.102	11.448	0
30	Accept	Accept	0.157	0.843	0.686	31214	3	5103.039	11.245	0
32	Reject	Accept	0.324	0.676	0.352	39437	3	4205.424	10.134	0
34	Accept	Accept	0.324	0.676	0.352	6939	6	3645.735	2.481	0
35	Accept	Accept	0.324	0.676	0.352	23030	3	3592.115	6.619	0
37	Accept	Accept	0.324	0.676	0.352	8518	4	1746.345	4.229	0



# Processes: Classification

Process



# Processes: Regression

