

Análise de bases com métodos de inteligência artificial

*Alunos: Guilherme Eduardo
Matheus Bello*

4º período

Professora: Lívia Marcia Silva

MÉTODOS DOS TESTES

- ▶ Foram executados 12 testes com parâmetros diferentes para cada método.
- ▶ No geral foram utilizados os nodes File Reader, Category to Number, Missing Value, Normalizer e Partitioning, com algumas remoções de nodes em certas bases .
- ▶ Cada um dos testes foi realizado com um parâmetro pré-definido, e a partir do melhor dos 12 resultados criado um último resultado com os parâmetros aprimorados para cada base .

ABALONE

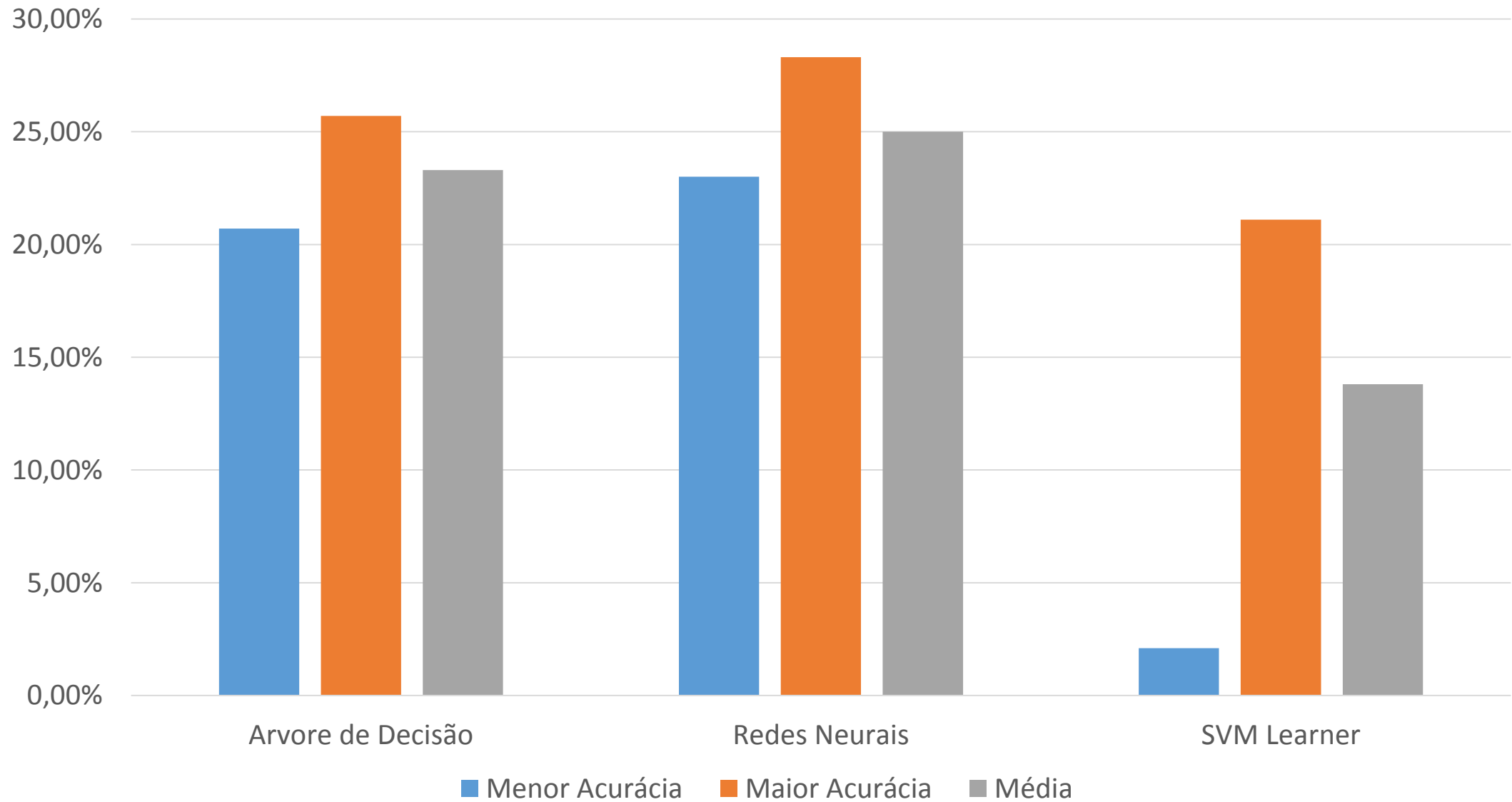
► A melhor inteligência para se usar nesta base são as redes neurais, tendo uma acurácia de 28.3%. Árvore de decisão e SVM ficaram com uma acurácia de 25.7% e 21.1%, respectivamente.

Arvore de Decisão								
Quality Measure	Pruning Method	Reduced Error Pruning	Min Number Records per Node	Number Records to Store for View	Avarage Split point	Number Threads	Skip Nominal Columns	Accuracy
Gain Ratio	No Pruning	V	2	10.000	V	4	V	20.7%
Gain Ratio	No Pruning	F	5	20.000	V	6	F	21.9%
Gain Ratio	No Pruning	V	8	30.000	F	8	V	24.2%
Gain Ratio	MDL	V	2	10.000	V	4	V	21.9%
Gain Ratio	MDL	F	5	20.000	V	6	F	23.4%
Gain Ratio	MDL	V	8	No Pruning 30.000	F	8	V	24.4%
Gini Index	No Pruning	V	2	10.000	V	4	V	23.0%
Gini Index	No Pruning	F	5	20.000	V	6	F	22.0%
Gini Index	No Pruning	V	8	30.000	F	8	V	22.0%
Gini Index	MDL	V	2	10.000	V	4	V	25.2%
Gini Index	MDL	F	5	20.000	V	6	F	25.5%
Gini Index	MDL	V	8	30.000	F	8	V	25.4%
Gini Index	MDL	V	5	20.000	F	6	V	25.7%

Redes Neurais			
Maximum Number of Iterations	Number of Hidden Layers	Number of hidden Neurons	Accuracy
50	1	5	25.4%
50	3	12	26.6%
75	3	20	28.7%
100	2	10	28.1%
100	4	5	26.6%
125	3	15	28.1%
150	5	10	23.0%
150	4	20	28.0%
175	3	5	28.2%
175	4	20	28.0%
200	5	15	28.1%
200	2	10	28.3%

SVM Learner										
Overlapping Penalty	Polynomial	Power	Bias	Gamma	HyperTangent	Kappa	Delta	RBF	Sigma	Accuracy
1	X	1.0	1	1.0						20.0%
1,5	X	1	1,5	1						20.0%
2	X	1,5	1	2						12.9%
2,5	X	2	0.5	0.5						19.4%
1					X	0,1	1.5			4.6%
1,5					X	0,3	1.2			4.1%
2					X	0,5	1			2.1%
2,5					X	1	0.5			2.2%
1								X	0.1	20.9%
1,5								X	0.3	19.5%
2								X	0.5	18.0%
2,5								X	0.7	15.8%
1								X	0.3	21.1%

Abalone



ADULT

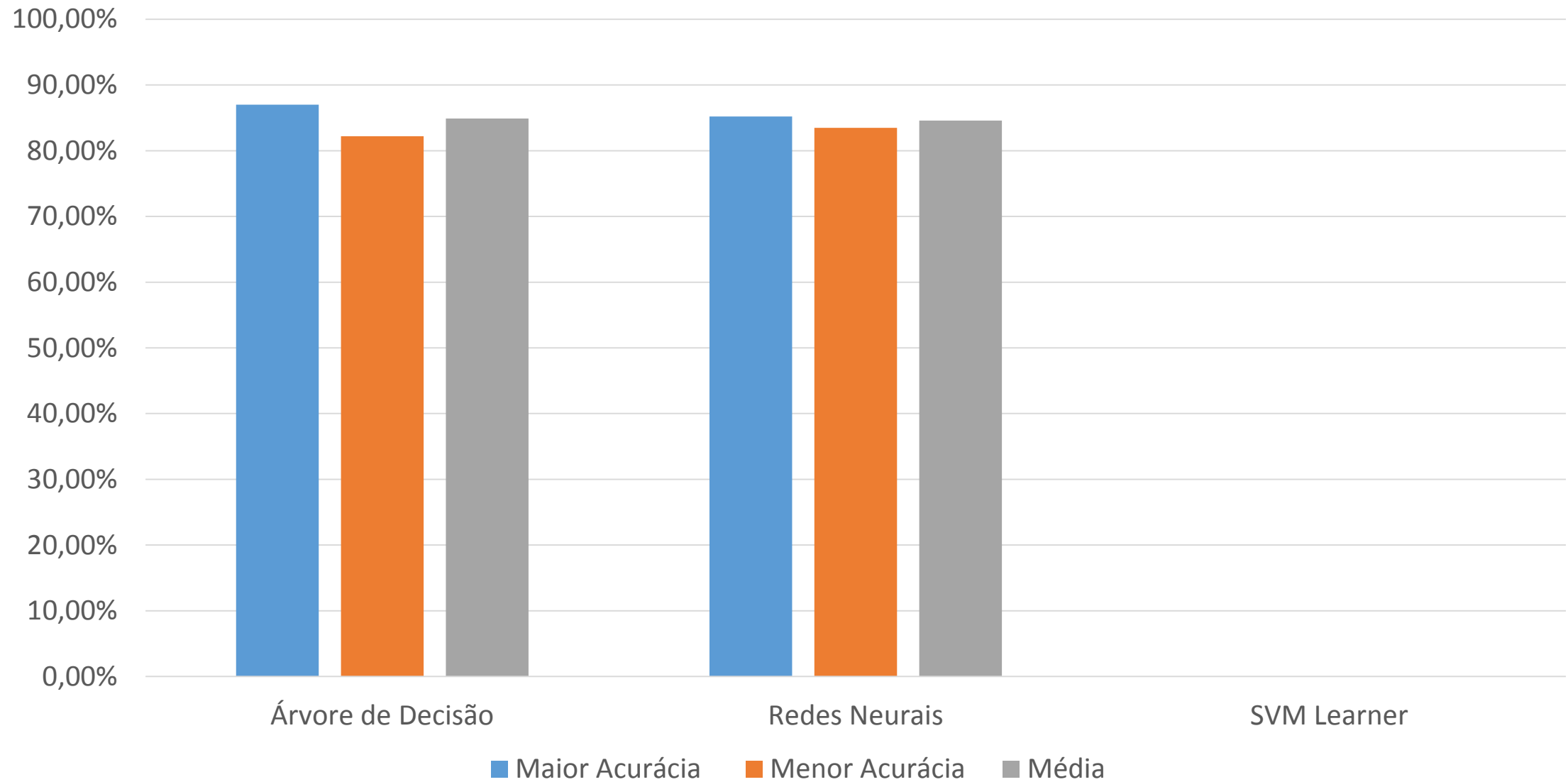
► A melhor inteligência para se usar nesta base é a árvore de decisão, tendo uma acurácia de 87%. Redes neurais ficaram com uma acurácia de 85,2%, já com a SVM não foi possível realizar os testes.

Arvore de Decisão								
Quality Measure	Pruning Method	Reduced Error Pruning	Min Number Records per Node	Number Records to Store for View	Avarage Split point	Number Threads	Skip Nominal Columns	Accu racy
Gain Ratio	No Pruning	V	2	10.000	V	4	V	83.0%
Gain Ratio	No Pruning	F	5	20.000	V	6	F	83.4%
Gain Ratio	No Pruning	V	8	30.000	F	8	V	84.7%
Gain Ratio	MDL	V	2	10.000	V	4	V	86.0%
Gain Ratio	MDL	F	5	20.000	V	6	F	86.4%
Gain Ratio	MDL	V	8	30.000	F	8	V	86.3%
Gini Index	No Pruning	V	2	10.000	V	4	V	82.2%
Gini Index	No Pruning	F	5	20.000	V	6	F	83.4%
Gini Index	No Pruning	V	8	30.000	F	8	V	84.1%
Gini Index	MDL	V	2	10.000	V	4	V	86.2%
Gini Index	MDL	F	5	20.000	V	6	F	86.1%
Gini Index	MDL	V	8	30.000	F	8	V	86.0%
Gain Ratio	MDL	F	9	20.000	V	3	F	87.0%

Redes Neurais			
Maximum Number of Iterations	Number of Hidden Layers	Number of hidden Neurons	Accuracy
50	1	5	83.5%
50	3	12	83.9%
75	3	20	84.6%
100	2	10	84.7%
100	4	5	84.1%
125	3	15	84.5%
150	5	10	85.1%
150	4	20	85.2%
175	3	5	84.9%
175	4	20	85.1%
200	5	15	85.0%
200	2	10	84.8%

SVM Learner										
Overlapping Penalty	Polynomial	Power	Bias	Gamma	HyperTangent	Kappa	Delta	RBF	Sigma	Accuracy
1	X	1.0	1	1.0						Error
1,5	X	1	1,5	1						Error
2	X	1,5	1	2						Error
2,5	X	2	0.5	0.5						Error
1					X	0,1	1.5			Error
1,5					X	0,3	1.2			Error
2					X	0,5	1			Error
2,5					X	1	0.5			Error
1								X	0.1	Error
1,5								X	0.3	Error
2								X	0.5	Error
2,5								X	0.7	Error
1								X	0.3	Error

Adult



BREAST CANCER WISCONSIN

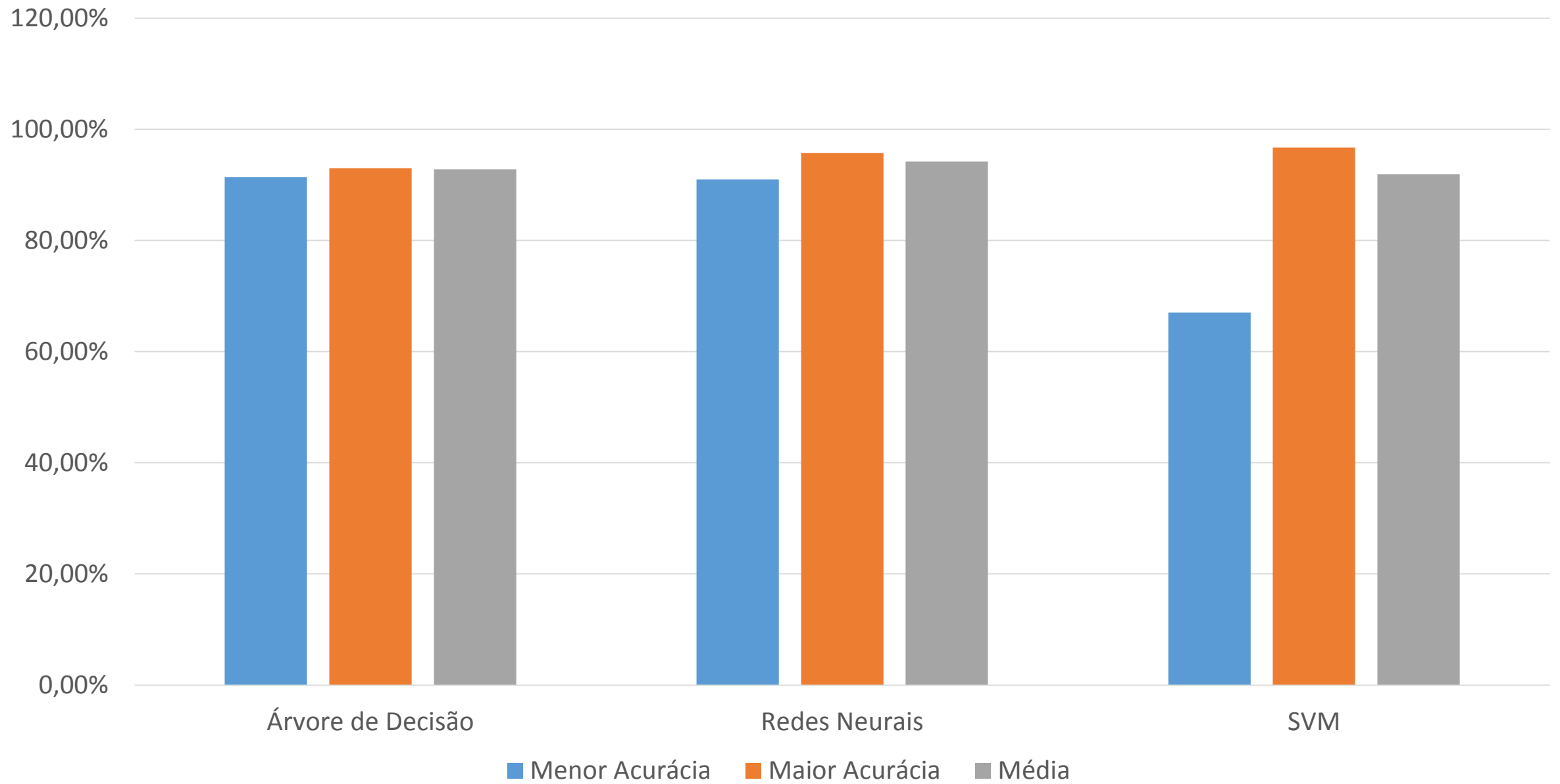
► A melhor inteligência para se usar nesta base são as redes neurais, com uma acurácia de 95,2%. SVM ficou com uma acurácia de 94,7%, já com a árvore de decisão ela foi de 93,8%.

Arvore de Decisão								
Quality Measure	Pruning Method	Reduced Error Pruning	Min Number Records per Node	Number Records to Store for View	Avarage Split point	Number Threads	Skip Nominal COLUMNS	Accu racy
Gain Ratio	No Pruning	V	2	10.000	V	4	V	91.4%
Gain Ratio	No Pruning	F	5	20.000	V	6	F	93.3%
Gain Ratio	No Pruning	V	8	30.000	F	8	V	93.8%
Gain Ratio	MDL	V	2	10.000	V	4	V	93.8%
Gain Ratio	MDL	F	5	20.000	V	6	F	93.8%
Gain Ratio	MDL	V	8	30.000	F	8	V	93.8%
Gini Index	No Pruning	V	2	10.000	V	4	V	93.8%
Gini Index	No Pruning	F	5	20.000	V	6	F	93.3%
Gini Index	No Pruning	V	8	30.000	F	8	V	91.9%
Gini Index	MDL	V	2	10.000	V	4	V	93.4%
Gini Index	MDL	F	5	20.000	V	6	F	92.3%
Gini Index	MDL	V	8	30.000	F	8	V	92.4%
Gain Ratio	MDL	F	5	20.000	V	4	V	93.8%

Redes Neurais			
Maximum Number of Iterations	Number of Hidden Layers	Number of hidden Neurons	Accuracy
50	1	5	94.7%
50	3	12	95.2%
75	3	20	94.7%
100	2	10	93.8%
100	4	5	95.2%
125	3	15	94.3%
150	5	10	96.2%
150	4	20	91.0%
175	3	5	95.2%
175	4	20	92.0%
200	5	15	93.3%
200	2	10	94.3%

SVM Learner										
Overlapping Penalty	Polynomial	Power	Bias	Gamma	HyperTangent	Kappa	Delta	RBF	Sigma	Accuracy
1	X	1.0	1	1.0						95.7%
1,5	X	1	1,5	1						95.7%
2	X	1,5	1	2						93.8%
2,5	X	2	0.5	0.5						94.3%
1					X	0,1	1.5			93.3%
1,5					X	0,3	1.2			96.7%
2					X	0,5	1			89.0%
2,5					X	1	0.5			67.0%
1								X	0.1	89.5%
1,5								X	0.3	94.7%
2								X	0.5	94.7%
2,5								X	0.7	94.7%

Breast Cancer Winsconsin



CAR

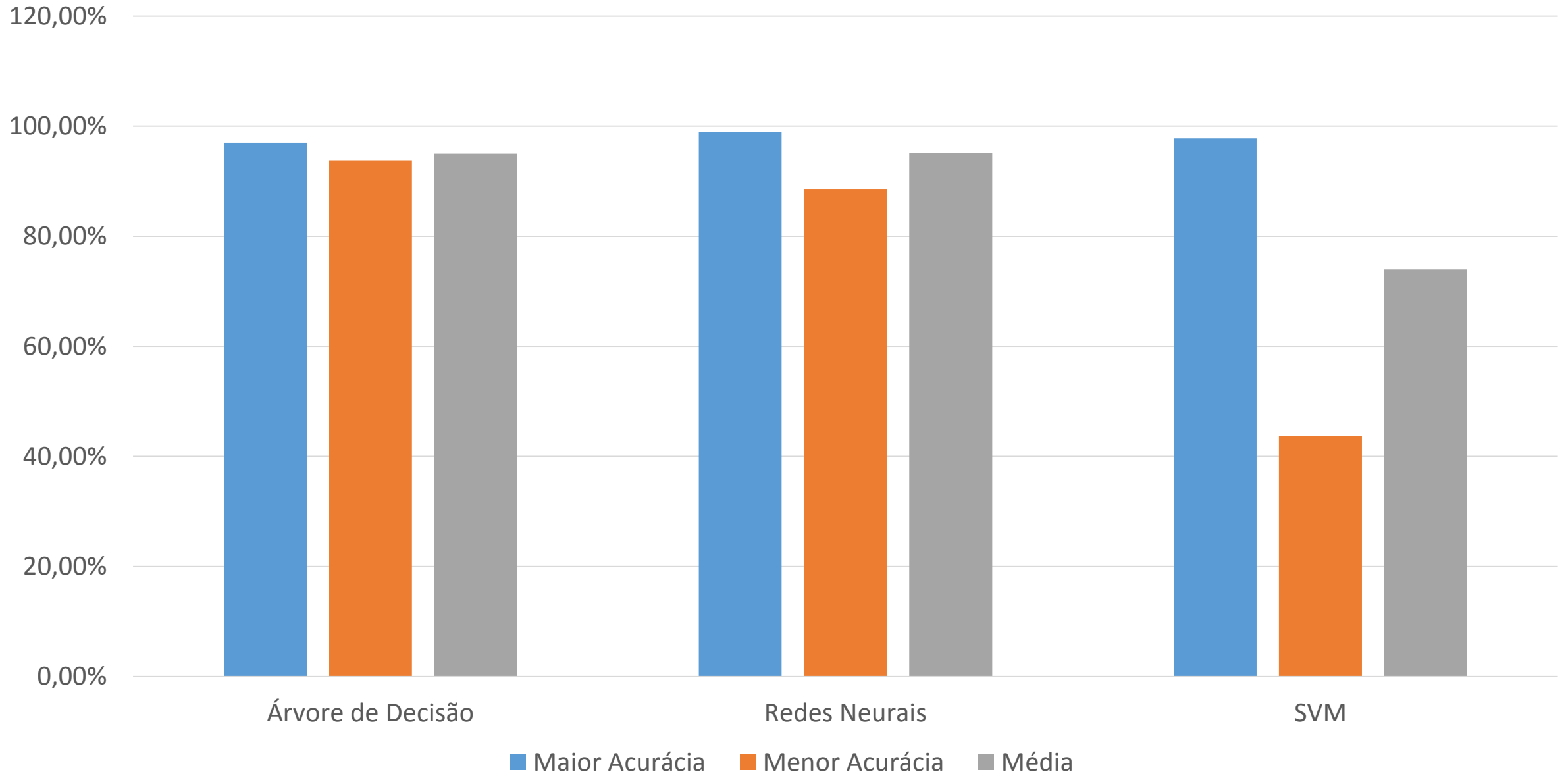
► A melhor inteligência para se usar nesta base são as redes neurais, tendo uma acurácia de 99%. SVM ficou com uma média de 97,8%, já com a árvore de decisão a acurácia foi de 97% .

Arvore de Decisão								
Quality Measure	Pruning Method	Reduced Error Pruning	Min Number Records per Node	Number Records to Store for View	Avarage Split point	Number Threads	Skip Nominal COLUMNS	Accu racy
Gain Ratio	No Pruning	V	2	10.000	V	4	V	96.3%
Gain Ratio	No Pruning	F	5	20.000	V	6	F	96.0%
Gain Ratio	No Pruning	V	8	30.000	F	8	V	93.8%
Gain Ratio	MDL	V	2	10.000	V	4	V	94.6%
Gain Ratio	MDL	F	5	20.000	V	6	F	94.6%
Gain Ratio	MDL	V	8	30.000	F	8	V	94.2%
Gini Index	No Pruning	V	2	10.000	V	4	V	95.5%
Gini Index	No Pruning	F	5	20.000	V	6	F	94.8%
Gini Index	No Pruning	V	8	30.000	F	8	V	93.8%
Gini Index	MDL	V	2	10.000	V	4	V	94.6%
Gini Index	MDL	F	5	20.000	V	6	F	94.6%
Gini Index	MDL	V	8	30.000	F	8	V	94.2%
Gain Ratio	MDL	V	2	10.000	V	4	V	97.0%

Redes Neurais			
Maximum Number of Iterations	Number of Hidden Layers	Number of hidden Neurons	Accuracy
50	1	5	88.6%
50	3	12	87.6%
75	3	20	96.7%
100	2	10	94.6%
100	4	5	92.6%
125	3	15	98.0%
150	5	10	96.5%
150	4	20	97.8%
175	3	5	93.2%
175	4	20	97.8%
200	5	15	96.3%
200	2	10	97.8%
175	3	19	99.0%

SVM Learner										
Overlapping Penalty	Polynomial	Power	Bias	Gamma	HyperTangent	Kappa	Delta	RBF	Sigma	Accuracy
1	X	1.0	1	1.0						81.1%
1,5	X	1	1,5	1						81.5%
2	X	1,5	1	2						91.0%
2,5	X	2	0.5	0.5						90.5%
1					X	0,1	1.5			70.7%
1,5					X	0,3	1.2			70.1%
2					X	0,5	1			54.7%
2,5					X	1	0.5			43.7%
1								X	0.1	81.7%
1,5								X	0.3	95.3%
2								X	0.5	97.7%
2,5								X	0.7	96.3%
2,5								X	0.5	97.8%

Car



WDBC

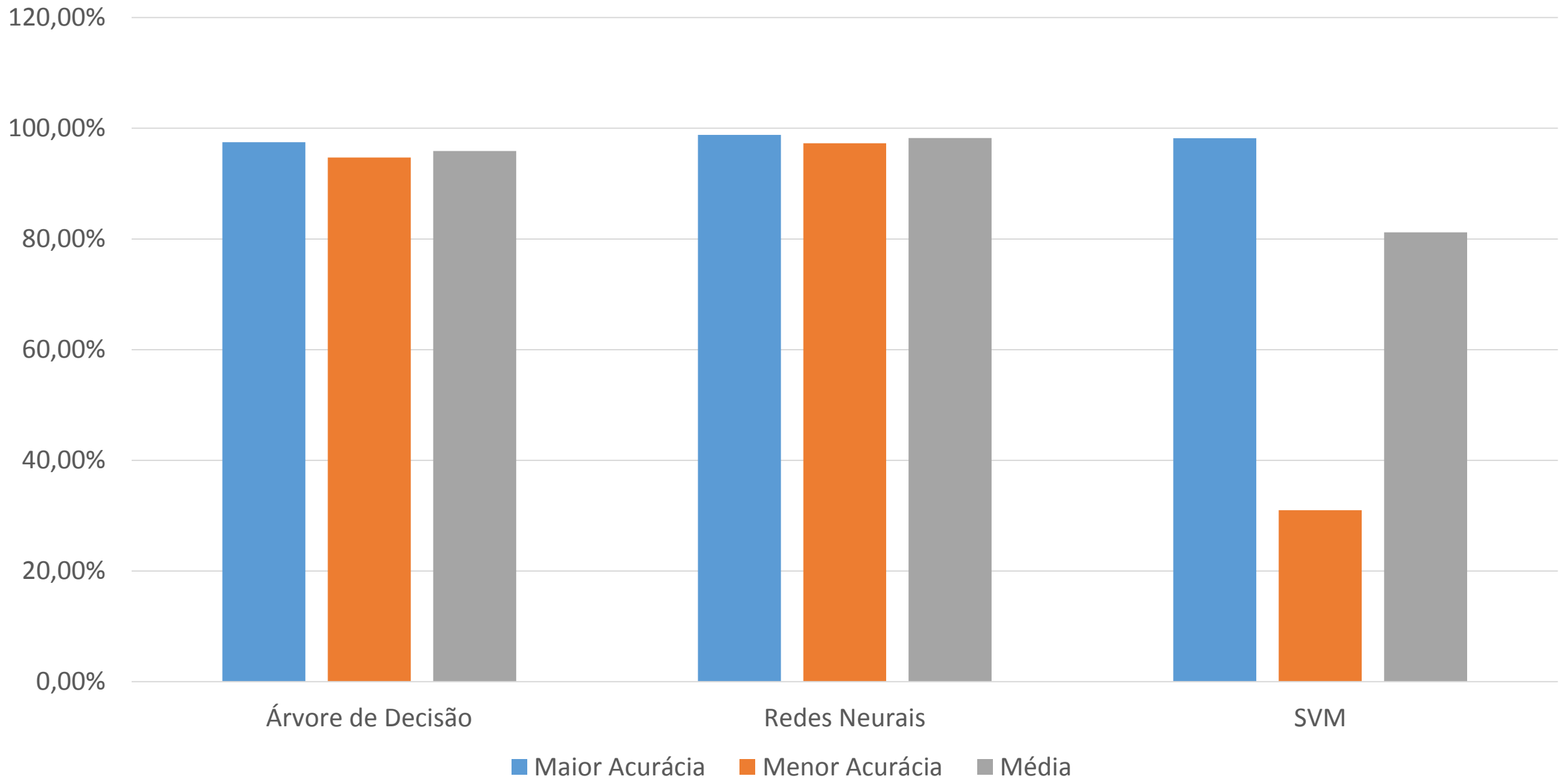
► A melhor inteligência para se usar nesta base são as redes neurais, tendo uma acurácia média de 98,8%. SVM ficou com uma acurácia de 98,2%, já com a árvore de decisão ela foi de 97,6%.

Arvore de Decisão								
Quality Measure	Pruning Method	Reduced Error Pruning	Min Number Records per Node	Number Records to Store for View	Avarage Split point	Number Threads	Skip Nominal COLUMNS	Accu racy
Gain Ratio	No Pruning	V	2	10.000	V	4	V	96.5%
Gain Ratio	No Pruning	F	5	20.000	V	6	F	95.9%
Gain Ratio	No Pruning	V	8	30.000	F	8	V	94.7%
Gain Ratio	MDL	V	2	10.000	V	4	V	95.3%
Gain Ratio	MDL	F	5	20.000	V	6	F	95.3%
Gain Ratio	MDL	V	8	30.000	F	8	V	95.3%
Gini Index	No Pruning	V	2	10.000	V	4	V	96.5%
Gini Index	No Pruning	F	5	20.000	V	6	F	96.5%
Gini Index	No Pruning	V	8	30.000	F	8	V	95.9%
Gini Index	MDL	V	2	10.000	V	4	V	95.9%
Gini Index	MDL	F	5	20.000	V	6	F	95.9%
Gini Index	MDL	V	8	30.000	F	8	V	95.9%
Gain Ratio	No Pruning	V	4	10.000	V	2	V	97.6%

Redes Neurais			
Maximum Number of Iterations	Number of Hidden Layers	Number of hidden Neurons	Accuracy
50	1	5	98.2%
50	3	12	97.6%
75	3	20	98.2%
100	2	10	98.2%
100	4	5	98.2%
125	3	15	97.6%
150	5	10	98.2%
150	4	20	98.8%
175	3	5	98.8%
175	4	20	98.2%
200	5	15	98.2%
200	2	10	98.8%
200	2	10	98.8%

SVM Learner										
Overlapping Penalty	Polynomial	Power	Bias	Gamma	HyperTangent	Kappa	Delta	RBF	Sigma	Accuracy
1	X	1.0	1	1.0						98.2%
1.5	X	1	1.5	1						98.2%
2	X	1.5	1	2						98.2%
2.5	X	2	0.5	0.5						98.2%
1					X	0.1	1.5			83.0%
1.5					X	0.3	1.2			66.0%
2					X	0.5	1			44.4%
2.5					X	1	0.5			31.0%
1								X	0.1	66.0%
1.5								X	0.3	96.5%
2								X	0.5	97.6%
2.5								X	0.7	98.2%

Wdbc



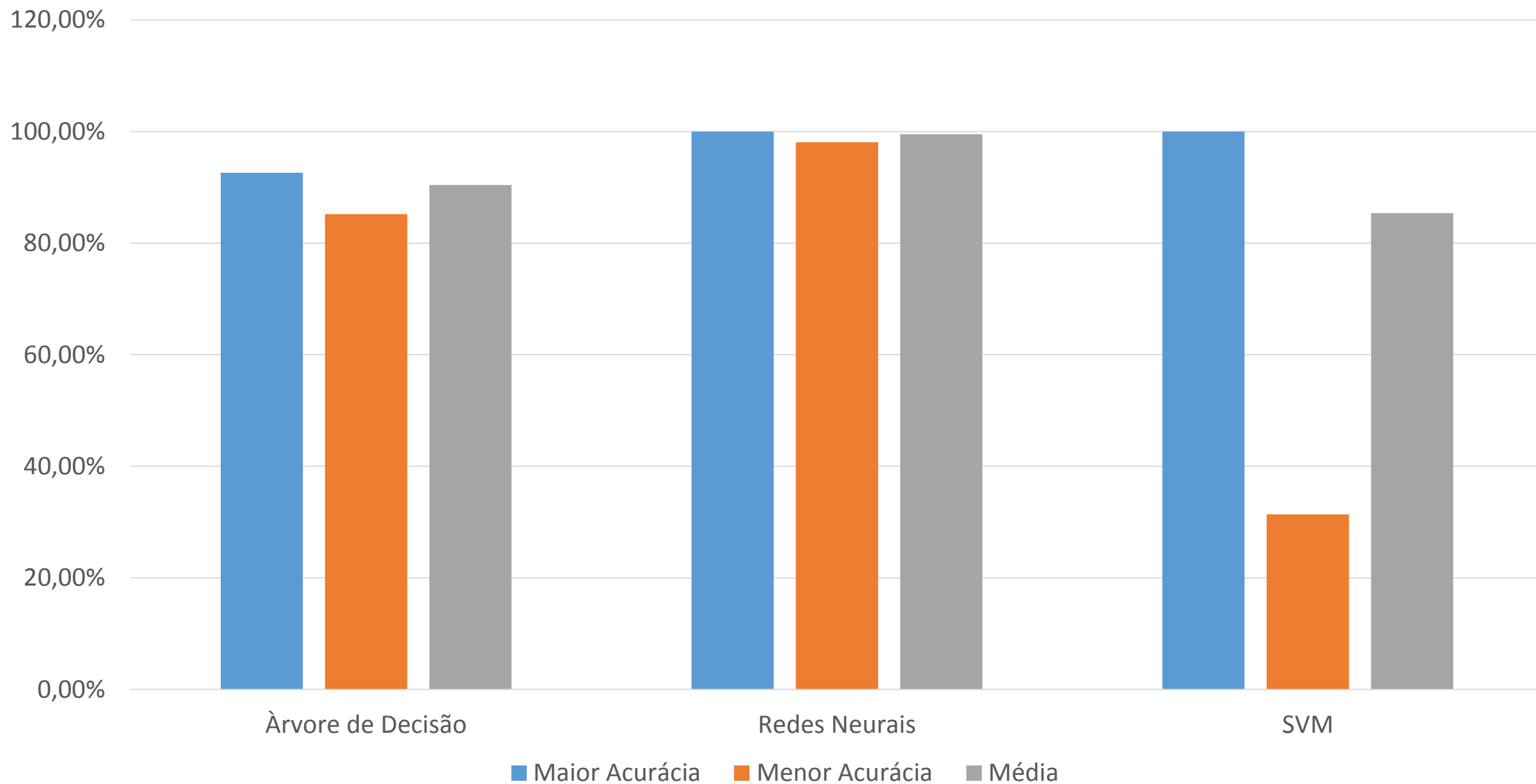
WINE

► A melhor inteligência para se usar nesta base são as redes neurais e SVM, tendo uma acurácia de 100%, já a árvore de decisão a acurácia foi de 92,6% .

Arvore de Decisão								
Quality Measure	Pruning Method	Reduced Error Pruning	Min Number Records per Node	Number Records to Store for View	Avarage Split point	Number Threads	Skip Nominal COLUMNS	Accu racy
Gain Ratio	No Pruning	V	2	10,000	V	4	V	88.9 %
Gain Ratio	No Pruning	F	5	20,000	V	6	F	88.9 %
Gain Ratio	No Pruning	V	8	30,000	F	8	V	85.2 %
Gain Ratio	MDL	V	2	10,000	V	4	V	88.9 %
Gain Ratio	MDL	F	5	20,000	V	6	F	88.9 %
Gain Ratio	MDL	V	8	30,000	F	8	V	88.9 %
Gini Index	No Pruning	V	2	10,000	V	4	V	92.6 %
Gini Index	No Pruning	F	5	20,000	V	6	F	92.6 %
Gini Index	No Pruning	V	8	30,000	F	8	V	92.6 %
Gini Index	MDL	V	2	10,000	V	4	V	92.6 %
Gini Index	MDL	F	5	20,000	V	6	F	92.6 %
Gini Index	MDL	V	8	30,000	F	8	V	92.6 %

Redes Neurais			
Maximum Number of Iterations	Number of Hidden Layers	Number of hidden Neurons	Accuracy
50	1	5	100%
50	3	12	100%
75	3	20	100%
100	2	10	100%
100	4	5	98.1%
125	3	15	98.1%
150	5	10	100%
150	4	20	100%
175	3	5	98.1%
175	4	20	100%
200	5	15	100%
200	2	10	100%

Wine



WPBC

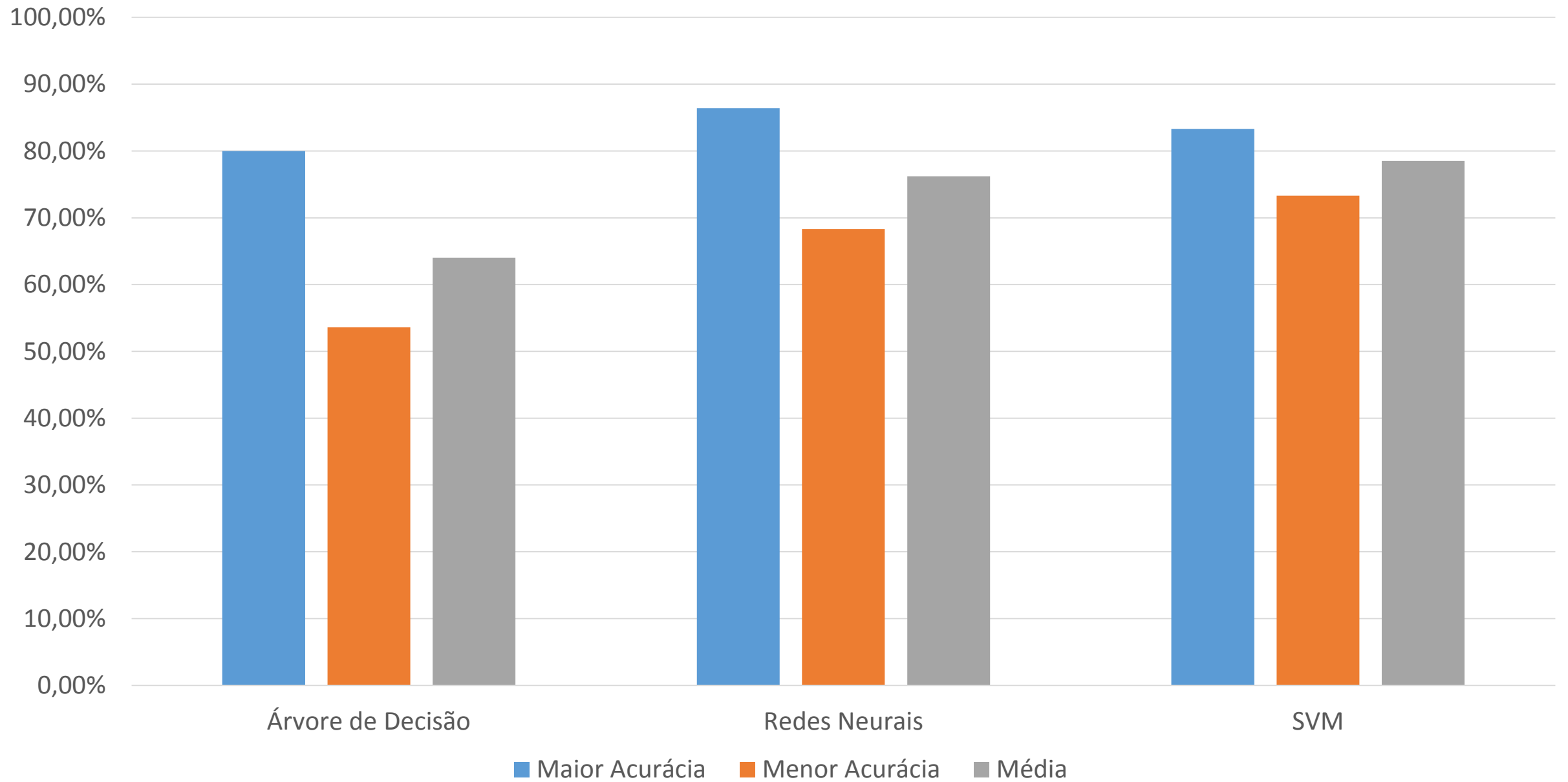
► A melhor inteligência para se usar nesta base são as redes neurais, tendo uma acurácia de 86,6%, já com a árvore de decisão a acurácia foi de 80%, e SVM 83,3% .

Arvore de Decisão								
Quality Measure	Pruning Method	Reduced Error Pruning	Min Number Records per Node	Number Records to Store for View	Avarage Split point	Number Threads	Skip Nominal COLUMNS	Accura cy
Gain Ratio	No Pruning	V	2	10,000	V	4	V	55.0%
Gain Ratio	No Pruning	F	5	20,000	V	6	F	66.6%
Gain Ratio	No Pruning	V	8	30,000	F	8	V	60.0%
Gain Ratio	MDL	V	2	10,000	V	4	V	78.3%
Gain Ratio	MDL	F	5	20,000	V	6	F	66.6%
Gain Ratio	MDL	V	8	30,000	F	8	V	58.3%
Gini Index	No Pruning	V	2	10,000	V	4	V	53.6%
Gini Index	No Pruning	F	5	20,000	V	6	F	56.6%
Gini Index	No Pruning	V	8	30,000	F	8	V	65.0%
Gini Index	MDL	V	2	10,000	V	4	V	65.0%
Gini Index	MDL	F	5	20,000	V	6	F	65.0%
Gini Index	MDL	V	8	30,000	F	8	V	65.0%
Gain Ratio	No Pruning	V	3	10,000	V	4	V	80.0%

Redes Neurais			
Maximum Number of Iterations	Number of Hidden Layers	Number of hidden Neurons	Accuracy
50	1	5	86.6%
50	3	12	68.3%
75	3	20	78.3%
100	2	10	76.6%
100	4	5	76.6%
125	3	15	70.0%
150	5	10	78.3%
150	4	20	76.6%
175	3	5	70.0%
175	4	20	76.6%
200	5	15	73.3%
200	2	10	73.3%

SVM Learner										
Overlapping Penalty	Polynomial	Power	Bias	Gamma	HyperTangent	Kappa	Delta	RBF	Sigma	Accuracy
1	X	1.0	1	1.0						83.3%
1.5	X	1	1.5	1						80.0%
2	X	1.5	1	2						78.3%
2.5	X	2	0.5	0.5						78.3%
1					X	0.1	1.5			73.3%
1.5					X	0.3	1.2			78.3%
2					X	0.5	1			78.3%
2.5					X	1	0.5			75.0%
1								X	0.1	78.3%
1.5								X	0.3	76.6%
2								X	0.5	80.0%
2.5								X	0.7	78.3%

Wpbc



YEAST

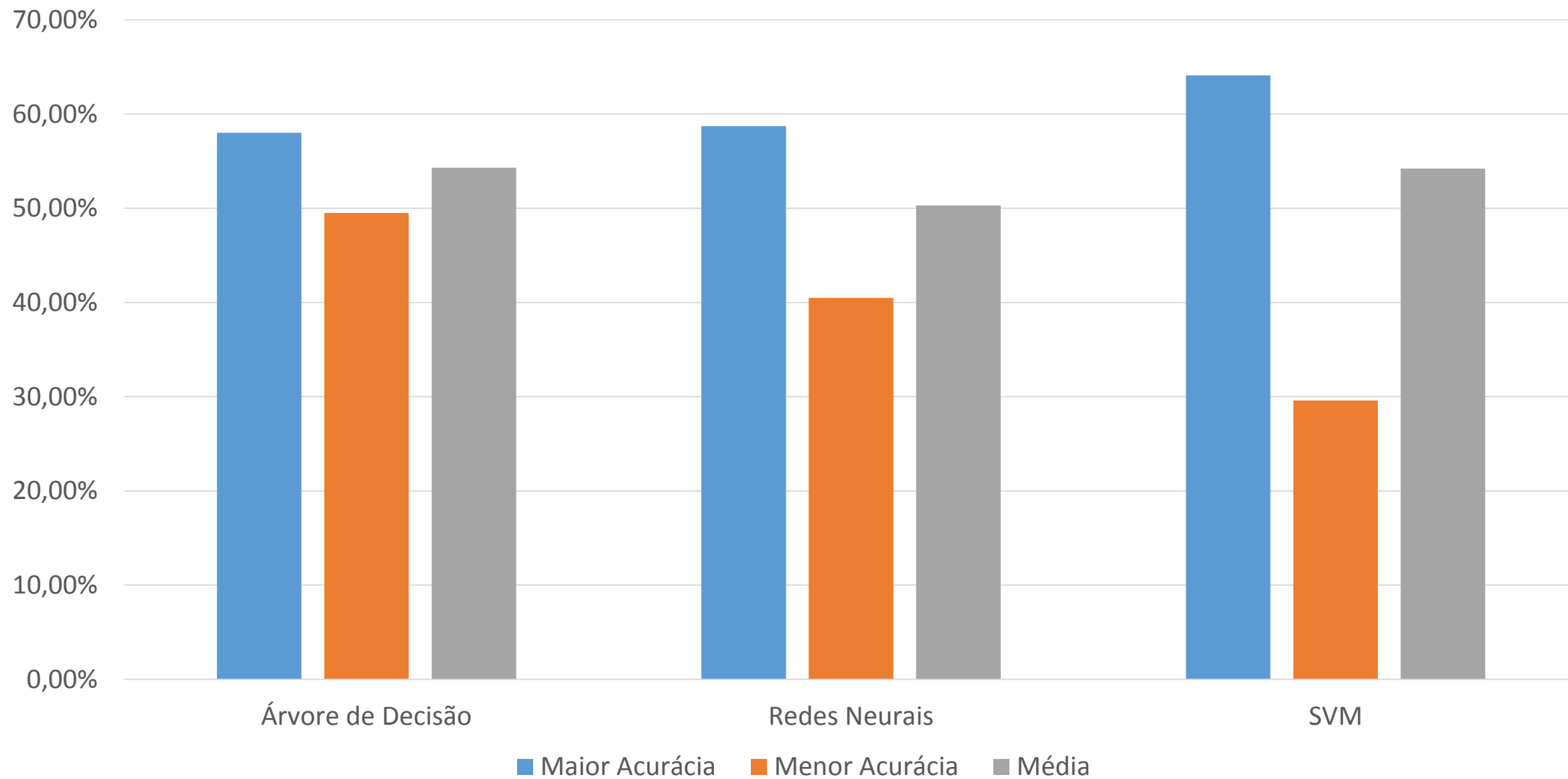
► A melhor inteligência para se usar nesta base foi a SVM, tendo uma acurácia de 64,1%, já com a árvore de decisão a acurácia foi de 58%, e redes neurais 58,7% .

Arvore de Decisão								
Quality Measure	Pruning Method	Reduced Error Pruning	Min Number Records per Node	Number Records to Store for View	Avarage Split point	Number Threads	Skip Nominal COLUMNS	Accuracy
Gain Ratio	No Pruning	V	2	10,000	V	4	V	52.0%
Gain Ratio	No Pruning	F	5	20,000	V	6	F	49.5%
Gain Ratio	No Pruning	V	8	30,000	F	8	V	50.2%
Gain Ratio	MDL	V	2	10,000	V	4	V	55.6%
Gain Ratio	MDL	F	5	20,000	V	6	F	54.0%
Gain Ratio	MDL	V	8	30,000	F	8	V	53.4%
Gini Index	No Pruning	V	2	10,000	V	4	V	54.0%
Gini Index	No Pruning	F	5	20,000	V	6	F	54.2%
Gini Index	No Pruning	V	8	30,000	F	8	V	51.8%
Gini Index	MDL	V	2	10,000	V	4	V	58.0%
Gini Index	MDL	F	5	20,000	V	6	F	57.8%
Gini Index	MDL	V	8	30,000	F	8	V	57.6%
Gini Index	MDL	V	8	30,000	F	8	V	57.6%

Redes Neurais			
Maximum Number of Iterations	Number of Hidden Layers	Number of hidden Neurons	Accuracy
50	1	5	51.3%
50	3	12	43.2%
75	3	20	56.5%
100	2	10	57.4%
100	4	5	40.5%
125	3	15	58.7%
150	5	10	57.1%
150	4	20	57.6%
175	3	5	56.7%
175	4	20	58.5%
200	5	15	56.7%
200	2	10	58.3%

SVM Learner										
Overlapping Penalty	Polynomial	Power	Bias	Gamma	HyperTangent	Kappa	Delta	RF	Sigma	Accuracy
1	X	1.0	1	1.0						52.9%
1.5	X	1	1.5	1						54.2%
2	X	1.5	1	2						58.5%
2.5	X	2	0.5	0.5						57.6%
1					X	0.1	1.5			51.3%
1.5					X	0.3	1.2			51.1%
2					X	0.5	1			43.9%
2.5					X	1	0.5			29.6%
1								X	0.1	57.4%
1.5								X	0.3	63.6%
2								X	0.5	61.4%
2.5								X	0.7	59.6%
2								X	0.3	64.1%

Yeast



Conclusão

A partir dos resultados finais pode-se concluir que o método mais efetivo nas bases trabalhadas são as redes neurais, com uma margem de resultados positivos considerável em relação à árvore de decisão e SVM.