

ANÁLISE DE DADOS DE EXPRESSÃO GÊNICA

Estagiário PAE: Pablo Andretta Jaskowiak

Professor: Ricardo J. G. B. Campello

SCC0173

Mineração de Dados Biológicos

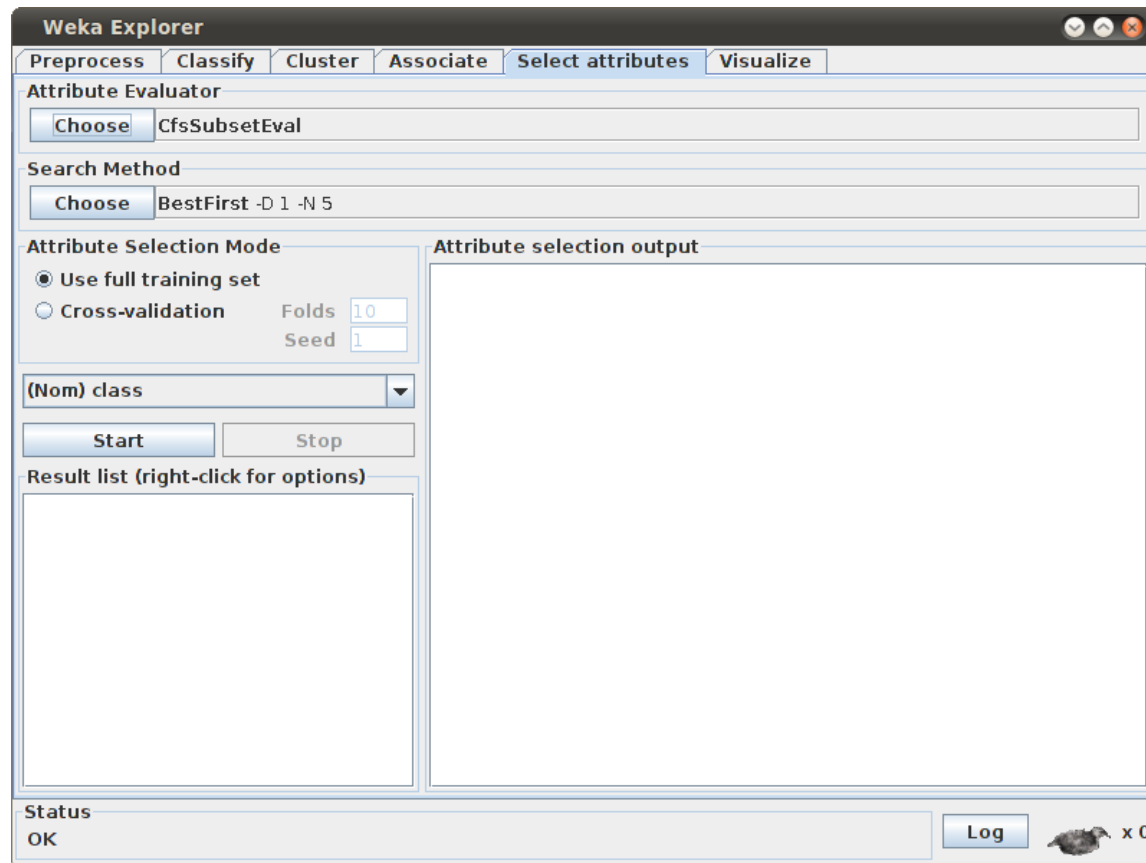
Sumário



- Seleção de Atributos
- Seleção + Classificação
- Prática!

Seleção de Atributos

- Explorer
 - ▣ *Aba Select Attributes*



Seleção de Atributos

- Como utilizar?
 - ▣ Escolher método de seleção
 - ▣ Escolher estratégia de busca
 - *GreedyStepwise*
- Só faz seleção de atributos
 - ▣ Não aplica um classificador

Seleção + Classificação

- Processo de Seleção
 - ▣ Somente na pasta de treinamento!
- Como fazer no WEKA?
- Aba *Classify*
 - ▣ Classificadores Meta
 - AttributeSelectedClassifier
 - ▣ Escolher
 - Método de seleção
 - Estratégia de busca
 - Classificador

Seleção + Classificação

- **AttributeSelectedClassifier**
 - ▣ Não mostra os atributos selecionados
 - ▣ Não permite visualizar atributos por pasta

- Para obter tais informações
 - ▣ Utilizar o ambiente *KnowledgeFlow*

Seleção + Classificação

Weka KnowledgeFlow Environment

DataSources DataSinks Filters Classifiers Clusterers Associations Evaluation Visualization

Visualization

Data Visualizer Scatter PlotMatrix Attribute Summarizer Model PerformanceChart CostBenefit Analysis Text Viewer Graph Viewer Strip Chart

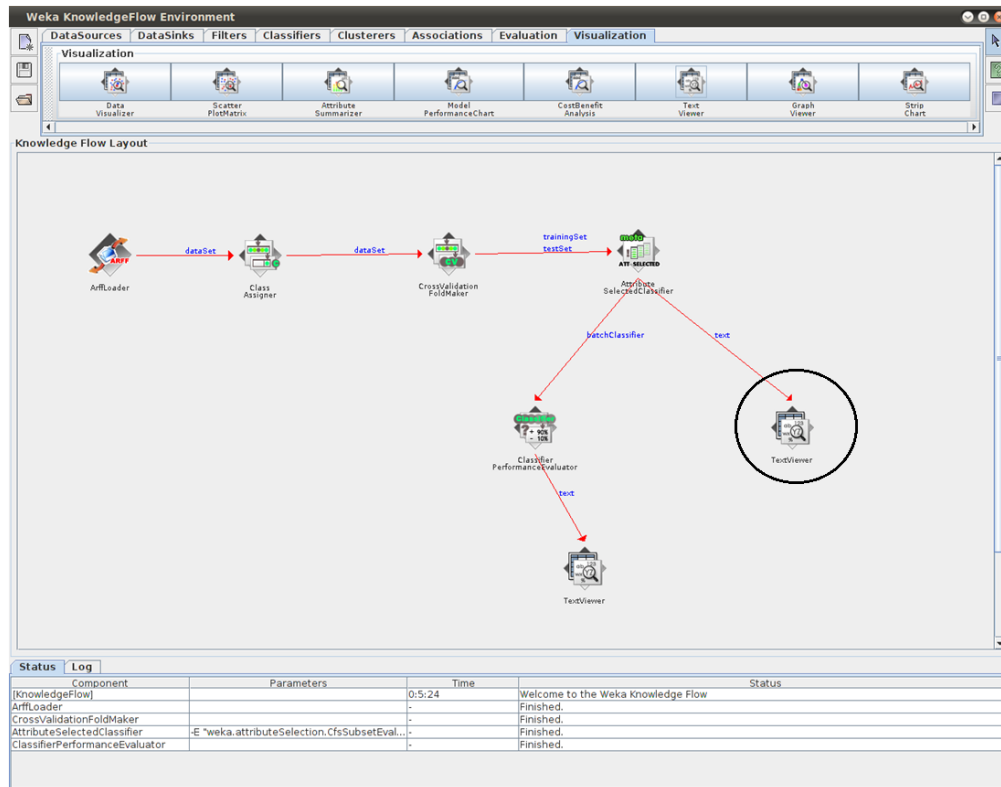
Knowledge Flow Layout

```
graph LR; ArffLoader[ArffLoader] -- dataSet --> ClassAssigner[Class Assigner]; ClassAssigner -- dataSet --> CrossValidationFoldMaker[CrossValidation FoldMaker]; CrossValidationFoldMaker -- trainingSet testSet --> AttributeSelectedClassifier[Attribute SelectedClassifier]; AttributeSelectedClassifier -- batchClassifier --> ClassifierPerformanceEvaluator[Classifier Performance Evaluator]; AttributeSelectedClassifier -- text --> TextViewer1[Text Viewer]; ClassifierPerformanceEvaluator -- text --> TextViewer2[Text Viewer];
```

Status Log

Component	Parameters	Time	Status
[KnowledgeFlow]		0:5:24	Welcome to the Weka Knowledge Flow
ArffLoader		-	Finished.
CrossValidationFoldMaker		-	Finished.
AttributeSelectedClassifier	-E "weka.attributeSelection.CfsSubsetEval..."	-	Finished.
ClassifierPerformanceEvaluator		-	Finished.

Seleção + Classificação



- Permite visualizar
 - Número de atributos selecionados em cada pasta
 - Atributos selecionados em cada pasta

Prática

- Para cada uma das bases
 - Na aba *Select Attributes*
 - Aplicar os métodos de seleção vistos em sala
 - Conjunto completo de dados
 - Utilizando validação cruzada de 10 pastas
 - Observar os resultados produzidos por
 - Filtros de subconjunto
 - Algoritmo de ordenação de atributos

Prática

- Para cada uma das bases
 - Na aba *Classify*
 - Aplicar os métodos de seleção vistos em sala
 - Utilizando validação cruzada de 10 pastas
 - Testar nos diferentes classificadores vistos em sala
 - Observar os resultados produzidos por
 - Filtros de subconjunto
 - Algoritmo de ordenação de atributos
 - Aplique os métodos de classificação sem seleção de atributos
 - Compare os resultados obtidos com e sem seleção