Machine Learning with WEKA

Eibe Frank

Department of Computer Science, University of Waikato, New Zealand

Revised by Pruet Boonma

- WEKA: A Machine Learning Toolkit
- The Explorer
 - Classification and Regression
 - Clustering
 - Association Rules
 - Attribute Selection
 - Data Visualization
- The Experimenter
- The Knowledge Flow GUI
- Conclusions

WEKA: the bird



Copyright: Martin Kramer (mkramer@wxs.nl)

University of Waikato

WEKA: the software

- Machine learning/data mining software written in Java (distributed under the GNU Public License)
 - Support MS Windows, Mac OS X and GNU/Linux
- Used for research, education, and applications
- Complements "Data Mining" by Witten & Frank
- Main features:
 - Comprehensive set of data pre-processing tools, learning algorithms and evaluation methods
 - Graphical user interfaces (incl. data visualization)
 - Environment for comparing learning algorithms

WEKA only deals with "flat" files

@relation heart-disease-simplified

@attribute age numeric
@attribute sex { female, male}
@attribute chest_pain_type { typ_angina, asympt, non_anginal, atyp_angina}
@attribute cholesterol numeric
@attribute exercise_induced_angina { no, yes}
@attribute class { present, not_present}

@data

. . .

63,male,typ_angina,233,no,not_present 67,male,asympt,286,yes,present 67,male,asympt,229,yes,present 38,female,non_anginal,?,no,not_present



WEKA only deals with "flat" files

numeric attribute

@attribute age numeric @attribute sex { female, male} @attribute chest_pain_type { typ_angina, asympt, non_anginal, atyp_angina} @attribute cholesterol numeric @attribute exercise_induced_angina { no, yes} @attribute class { present, not present}

@data 63,male,typ_angina,233,no,not_present 67,male,asympt,286,yes,present 67,male,asympt,229,yes,present 38,female,non_anginal,?,no,not_present

. . .

@relation heart-disease-simplified



University of Waikato



University of Waikato

Explorer: pre-processing the data

- Data can be imported from a file in various formats: ARFF, CSV, C4.5, binary
- Data can also be read from a URL or from an SQL database (using JDBC)
- Pre-processing tools in WEKA are called "filters"
- WEKA contains filters for:
 - Discretization, normalization, resampling, attribute selection, transforming and combining attributes, ...

000	Weka Knowledge Explorer						
	Preprocess Classify Cluster	Associate Select attributes Visualize					
Open file	Open URL O	Open DB Undo	Save				
Filter Choose None			Apply				
Current relation Relation: None Instances: None	Attributes: None	Selected attribute Name: None Missing: None Distinct: None	Type: None Unique: None				
Attributes							
			Visualize All				
Status							
Welcome to the Wek	a Knowledge Explorer	C	Log X 0				



Welcome to the Weka Knowledge Explorer





0) 🖯 🖯 Weka Knowledge Explorer							
		Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Ор	en file		en URL		Open DB	Undo		Save
Filter	None							Apply
Current rel Relatio	lation on: iris es: 150	At	ttributes: 5		Selected Nam Missin	lattribute ne:class ng:0(0%) Disti	nct: 3	Type: Nominal Unique: 0 (0%)
Attributes						Label	<u> </u>	Count
No. 1 2 3 4 5	sepallength sepalwidth petallength petalwidth class	Na h	ime		Iris-ver Iris-vir	sicolor ginica r: class (Nom)	50 50	Visualize All
					50	50		50
Status							6	100
OK								Log

$\Theta \Theta \Theta$	🖻 🖯 🖯 Weka Knowledge Explorer						
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Open file	Оре	n URL		Open DB	Undo		Save
Filter							
Choose None							Apply
Current relation				Selected	l attribute		
Relation: iris Instances: 150	At	tributes: 5		Nam Missin	ne: class ng: 0 (0%) Dis	tinct: 3	Type: Nominal Unique: 0 (0%)
Attributes					Label		Count
No. 1 sepalleng 2 sepalwidt 3 petalleng 4 petalwidt 5 class	Na th th h	me		Colour	osa sicolor ginica r: class (Nom)	50	Visualize All
-Status						C	Log x0
UN							ALC: NOT





18 17 16 14

10

2







Status -

Log

P. × 0





$\Theta \Theta \Theta$	🖯 🖯 🖯 Weka Knowledge Explorer							
	Preprocess Classify	Cluster Associa	ate Select attributes Visualize	2				
Open file	Open URL	Open DB	. Undo (Save				
Filter								
Choose Discretiz	ze -B 10 -R first-last	\varTheta 🖯 🖯 🛛 wek	a.gui.GenericObjectEditor	Apply				
Current relation		weka.filters.unsuper	vised.attribute.Discretize					
Relation: iris Instances: 150	Attributes:	About An instance filter that of	discretizes a range of numeric	: Numeric : 10 (7%)				
Attributes		attributes in the datase	et into nominal attributes.	e				
No. 1 sepallengti	Name h	attributeIndices	first-last					
2 sepalwidth 3 petallength	1	bins	10					
4 petalwidth 5 class		findNumBins	False	•				
		invertSelection	False	Visualize All				
		makeBinary	False	•				
		useEqualFrequency	False	;				
		Open	Save OK Car	ncel				
		11	2 0 0 3 4	10				
Status		- 1	3.95	6.9				
OK			(Log x 0				





Explorer: building "classifiers"

- Classifiers in WEKA are models for predicting nominal or numeric quantities
- Implemented learning schemes include:
 - Decision trees and lists, instance-based classifiers, support vector machines, multi-layer perceptrons, logistic regression, Bayes' nets, ...
- "Meta"-classifiers include:
 - Bagging, boosting, stacking, error-correcting output codes, locally weighted learning, ...



cociato	Salact attributor	Viewalia
ssociate	Select attributes	VISUAIIZ



	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier							1
Choose ZeroR							
Test options		Clas	ssifier output	t			
🔘 Use training set							
O Supplied test se	t Set						
Cross-validation	n Folds 10						
Percentage split	% 66						
More op	ptions						
(Nom) class		•					
Start	Stop						
Result list (right-click for	r options) ———						
Status		1					1
ОК						C	Log 🔪 💉 🛛



-

Waka Knowledge Explorer

	weka kilowieuge Exp	lorer	
Preprocess Clas	ssify Cluster Associate	Select attributes Visualize	
Classifier			
Choose ZeroR			
Test options	Classifier output		
🔘 Use training set			
O Supplied test set Set			
• Cross-validation Folds 10			
O Percentage split % 66			
More options			
(Nom) class			
Start Stop			
Result list (right-click for options)			
	11		B



 $\mathbf{0}$

Weka Knowledge Explorer



100 æ

	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier weka Classifiers Classifier Classifi	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
V v j48 j48 N j48 N m5 Random Random REPTree UserCla	Forest Tree ssifier						
Status OK						C	Log 💦 🗙 X O



Status	
ОК	







OK

x 0

Log

x 0

Select attributes

Log

x 0

Preprocess Classify Cluster Associate Visualize Classifier 148 -C 0.25 -M 2 Choose Test options Classifier output Use training set === Run information === Supplied test set Set... ()weka.classifiers.trees.j48.J48 -C 0.25 -M 2 Scheme: Relation: iris Cross-validation Folds 10 Instances: 150 Attributes: 5 Percentage split % 66 (\bullet) sepallength sepalwidth More options... petallength petalwidth class + (Nom) class Test mode: split 66% train, remainder test === Classifier model (full training set) === Stop Start J48 pruned tree Result list (right-click for options) 11:49:05 - trees.j48.J48 petalwidth <= 0.6: Iris-setosa (50.0) petalwidth > 0.6 petalwidth <= 1.7 petallength <= 4.9: Iris-versicolor (48.0/1.0) petallength > 4.9petalwidth <= 1.5: Iris-virginica (3.0) petalwidth > 1.5: Iris-versicolor (3.0/1.0) petalwidth > 1.7: Iris-virginica (46.0/1.0) * Number of Leaves : 5

 γ

×0

Log

nieprotess			sociate s	incer attim				
Classifier								
Choose J48 -C 0.25 -M 2								
Test options	Classifier	output						
O Use training set	Time ta	Time taken to build model: 0.24 seconds						
O Supplied test set Set	Fyral	Fueluation on test colit						
O Cross-validation Folds 10	=== Summary ===							
Percentage split % 66	Correct	y Classifi	ed Instances		49	96.0784 %		
More options	Incorrec	tly Classi	fied Instanc	es	2	3.9216 %		
More options	Mean abs	atistic	~		0.9408			
	Root mea	in squared	error		0.1579			
(Nom) class	Relative	absolute	error		8.8979 %			
	Root rel	ative squa	red error		33.4091 %			
Start Stop	Total Nu	mber of In	stances		51			
Result list (right-click for options)	=== Deta	niled Accur	acy By Class	===				
	TP Rate	FP Rate	Precision	Recall	F-Measure	Class		
11.49.05 - (1995.)46.)46	1	0	1	1	1	Iris-setosa		
	1	0.063	0.905	1	0.95	Iris-versicolor		
	0.882	0	1	0.882	0.938	Iris-virginica		
	=== Conf	usion Matr	ix ===					
	a b c < classified as							
	15 0 0 a = Iris-setosa							
	0 19	0 b = I	ris-versicol	or				
	0 2 1	.5 c = I	ris-virginic	a				

Select attributes

x 0

Log

Classify Cluster Preprocess Associate Visualize Classifier 148 -C 0.25 -M 2 Choose Test options Classifier output Use training set Time taken to build model: 0.24 seconds Supplied test set Set... === Evaluation on test split === === Summary === Cross-validation Folds 10 Percentage split % 66 96.0784 % Correctly Classified Instances 49 Incorrectly Classified Instances 2 3.9216 % More options... Kappa statistic 0.9408 Mean absolute error 0.0396 Root mean squared error 0.1579 + (Nom) class Relative absolute error 8.8979 % Root relative squared error 33.4091 % Total Number of Instances 51 Start Stop === Detailed Accuracy By Class === Result list (right-click for options) Precision Recall F-Measure TP Rate FP Rate Class 11:49:05 - trees.j48.J48 1 0 1 Iris-setosa 1 1 0.063 0.905 0.95 Iris-versicolor 1 1 0.882 0.882 0.938 0 1 Iris-virginica === Confusion Matrix === <-- classified as a b C 0 | a = Iris-setosa 15 0 19 0 | b = Iris-versicolor 0 2 15 | c = Iris-virginica 0

Preprocess	Classify Cluster Associate Sel	ect attributes Visu	alize
Classifier Choose J48 -C 0.25 -M 2			
Test options O Use training set O Supplied test set Set O Cross-validation Folds 10 • Percentage split % 66	Classifier output Time taken to build model: 0.24 === Evaluation on test split === === Summary === Correctly Classified Instances Incorrectly Classified Instances	seconds = 49	96.0784 %
More options (Nom) class	Kappa statistic Kappa statistic Mean absolute error Root mean squared error Root relative squared error Total Number of Instances === Detailed Accuracy By Class =	0.9408 0.0396 0.1579 8.8979 % 33.4091 % 51	3.9210 %
View 11:49:05 - trees.j48.J48 View Save Load Save Re-e Visua Visua View	in main window in separate window result buffer model waluate model on current test set alize classifer errors alize tree	Recall F-Measure 1 1 1 0.95 0.882 0.938	Class Iris-setosa Iris-versicolor Iris-virginica
OK Visua OK Visua	alize margin curve alize threshold curve alize cost curve		Log 💉 X O

()

()

()

 (\cdot)

Weka Knowledge Explorer Preprocess Classify Cluster Associate Select attributes Visualize Classifier 148 -C 0 25 -M 2 Choose Weka Classifier Tree Visualizer: 11:49:05 - trees.j48.J48 (iris) Test options Tree View Use training set Supplied test set Cross-validation petalwidth Percentage split 96.0784 % <= 0.6 > 0.63.9216 % More optic Iris-setosa (50.0) petalwidth >1.7 <= 1.7 (Nom) class Iris-virginica (46.0/1.0) petallength Start <= 4.9 Result list (right-click for Iris-versicolor (48.0/1.0) petalwidth 335 11:49:05 - trees.j48.J > 1.5 <= 1.5 is-setosa is-versicolor Iris-virginica (3.0) Iris-versicolor (3.0/1.0) is-virginica TTT9_960094 0 19 0 1 b = Iris-versicolor Å. 0 2 15 | c = Iris-virginica ¥

Status

OK



P	Preprocess	Classify Clust	er Associate	Select attribut	es Visua	lize
Classifier						
Choose J48 -C 0.2	25 -M 2					
Test options		Classifier output	t			
O Use training set		Time taken t	o build model:	0.24 seconds		6
O Supplied test set	Set	Evaluat	on on test enli	+ ===		
O Cross-validation	Folds 10	=== Summary	===	- L-		
Percentage split More option	% 66	Correctly Cl Incorrectly Kappa statis Mean absolut	assified Instar Classified Inst tic e error	nces 4 tances	9 2 0.9408 0.0396	96.0784 % 3.9216 %
(Nom) class	Stop	Root mean so Relative abs Root relative Total Number === Detailed	uared error colute error re squared error of Instances Accuracy By Cl	3 5 Lass ===	0.1579 8.8979 % 3.4091 % 1	
11:49:05 - trees.j48.J4	8 View i View i Save r	n main window n separate windo esult buffer	w	Recall F 1 1 0.882	-Measure 1 0.95 0.938	Class Iris-setosa Iris-versicolor Iris-virginica
	Load r Save r Re-ev	model nodel aluate model on	current test set			
	Visual Visual	ize classifer erro ize tree	rs	lor ca		•
OK	Visual Visual Visual	ize margin curve lize threshold cu lize cost curve	ve	• •		Log 💉 x 0





Log

x 0

Status

OK

Explorer: clustering data

- WEKA contains "clusterers" for finding groups of similar instances in a dataset
- Implemented schemes are:
 - *k*-Means, EM, Cobweb, *X*-means, FarthestFirst
- Clusters can be visualized and compared to "true" clusters (if given)
- Evaluation based on loglikelihood if clustering scheme produces a probability distribution



	Weka Knowledge Explorer						
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Clusterer							
Choose EM -I 1	00 -N -1 -S 1	00 -M 1.0E-	-6				
Cluster mode			Clust	erer output			
Use training set							
Supplied test set	s s	et					
O Percentage split		% 66					
O Classes to clust	ers evaluation						
(Nom) class		Ť					
Store clusters fo	r visualization						
Igno	ore attributes	*******	\supset				
Start	\supset	Stop					
Result list (right-click for	options)						
Status							

Log 💉 x 0



OK



Preprocess Classify Cluster Associate Select attributes Visualize

Treprocess	Classify	_ craster	rissociate	Serect attributes	Tradified	
Clusterer						
Choose Cobweb -A 1.0 -C 0	.002820947	91773878	15			
Cluster mode		Clust	terer output			
💿 Use training set						
O Supplied test set	Set					
Percentage split	% 66					
O Classes to clusters evaluation	ı					
(Nom) class	A V					
Store clusters for visualizatio	n					
Ignore attributes						
Start	Stop					
Result list (right-click for options)						
Status						
ОК					C	Log x 0



Ε

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

	c,		rissectate	server attributes	The damage	
Clusterer						
Choose Cobweb -A 1.0 -C 0.0	0028209479	917738781	15			
Cluster mode		Clust	erer output			
💿 Use training set						
O Supplied test set	et					
Percentage split	% 66					
Classes to clusters evaluation						
Nom) class	A V					
Store crysters for visualization						
		_				
Ignore attributes						
Start	Stop	\supset				
Result list (right-click for options)						
Status		110				
OK					C	Log x 0





Select attributes

Log



Status OK



Select attributes

Visualize



Status OK





Select attributes

Visualize

Log

x 0















OK

Log

x 0

Explorer: finding associations

- WEKA contains an implementation of the Apriori algorithm for learning association rules
 - Works only with discrete data
- Can identify statistical dependencies between groups of attributes:
 - ◆ milk, butter ⇒ bread, eggs (with confidence 0.9 and support 2000)
- Apriori can compute all rules that have a given minimum support and exceed a given confidence

🖯 🖯 🗎 Weka Knowledge Explorer							
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Associator							
Choose Apriori	-N 10 -T 0 -C	0.9 –D 0.0	5 -U 1.0 -	M 0.1 -S -1.0			
		Ass	ociator outpu	ut			
Start	Stop						
Result list (right-click for	r options)						
Status		176					
OK						C	Log x 0

$\Theta \Theta \Theta$			Weka Kn	owledge Expl	orer		
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Associator							
Choose Apriori	-N 10 -T 0 -C	0.9 -1 0.0	5 -U 1.0 -I	M 0.1 -S -1.0			
Start	Stop	Ass	osiator outpu	ıt —			
Result list (right-click for	r options)						
Status		1					
OK						C	Log x 0



Log

OK

< ____



Log

ок

🖯 🖯 🗎 Weka Knowledge Explorer							
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Associator							
Choose Apriori	-N 10 -T 0 -C	0.9 –D 0.0	5 -U 1.0 -	M 0.1 -S -1.0			
		Ass	ociator outpu	ut			
Start	Stop						
Result list (right-click for	r options)						
Status		176					
OK						C	Log x 0





Associator

Choose

Start

Stop

Result list (right-click for optic

16:29:37 - Apriori

Weka Knowledge Explorer

Cluster

Classify

Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0

Minimum metric <confidence>: 0.9

Generated sets of large itemsets:

Size of set of large itemsets L(1): 20

Size of set of large itemsets L(2): 17

Size of set of large itemsets L(3): 6

Size of set of large itemsets L(4): 1

Number of cycles performed: 11

Associator output-

Preprocess

Associate

Select attributes

Visualize

num metric <confidence>: 0.9 er of cycles performed: 11</confidence>	
rated sets of large itemsets:	
of set of large itemsets L(1): 20	
of set of large itemsets L(2): 17	
of set of large itemsets L(3): 6	
of set of large itemsets L(4): 1	
rules found:	
	the second se
adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 =>> Class adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicar physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> Class=democrat 201 physician-fee-freeze=n education-spending=n 202 ==> Class=democrat 201 physician-fee-freeze=n 247 ==> Class=democrat 245 conf:(0.99) al-salvador-aid=n Class=democrat 200 ==> aid-to-nicaraguan-contras=y 19 al-salvador-aid=n 208 ==> aid-to-nicaraguan-contras=y 204 conf:(0.98 adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y Class=democrat 19 al-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> Class=democrat 19 al-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> class=democrat 19 aid-to-nicaraguan-contras=y Class=democrat 218 ==> physician-fee-freeze	ss=democra raguan-con rat 210 conf:(97 conf 8) democrat 2 97 conf e=n 210

	<pre>1. adoption-of-the-budget-resolution=y 2. adoption-of-the-budget-resolution=y 3. physician-fee-freeze=n aid-to-nicara 4. physician-fee-freeze=n education-spe 5. physician-fee-freeze=n 247 ==> Class 6. el-salvador-aid=n Class=democrat 200 7. el-salvador-aid=n 208 ==> aid-to-nic 8. adoption-of-the-budget-resolution=y 9. el-salvador-aid=n aid-to-nicaraguan- 10. aid-to-nicaraguan-contras=y Class=de </pre>
Status	
OK	

Explorer: attribute selection

- Panel that can be used to investigate which (subsets of) attributes are the most predictive ones
- Attribute selection methods contain two parts:
 - A search method: best-first, forward selection, random, exhaustive, genetic algorithm, ranking
 - An evaluation method: correlation-based, wrapper, information gain, chi-squared, ...
- Very flexible: WEKA allows (almost) arbitrary combinations of these two







Status





University of Waikato

Performing experiments

- Experimenter makes it easy to compare the performance of different learning schemes
- For classification and regression problems
- Results can be written into file or database
- Evaluation options: cross-validation, learning curve, hold-out
- Can also iterate over different parameter settings
- Significance-testing built in!

😑 🖯 💮 🛛 Weka Experin	nent Environment	
Setup	Run Analyse	
Experiment Configuration Mode:	 Simple 	Advanced
Open S	ave	New
Results Destination		
JDBC database 🗘 Filename:		Browse)
Experiment Type	Iteration Control	
Cross-validation	Number of repetit	ions:
Number of folds:	 Data sets first 	t
Classification Regression	 Algorithms fir 	st
Datasets	Algorithms	
Add new Delete selected Use relative paths Notes	Add new.	Delete selected

🖯 🖯 🖯 Weka	Experime	nt Environment			
Se	tup Ru	n Analyse			
Experiment Configuration Mode:		🖲 Simple	Advanced		
Open	Sav	/e	New		
Results Destination					
JDBC database 🗘 Filename:			Browse)		
Experiment Type		Iteration Control			
Cross-validation	×	Number of repetit	ions:		
Number of folds:		 Data sets first 			
Classification Classification		 Algorithms fir 	st		
Datasets		Algorithms			
Add new Delete selecte	d)	Add new	. Delete selected		
Use relative paths					
_ Notes	1	11	J		
1			1		

😇 🖯 💭 Weka Experime	ent Environment
Setup Ru	un Analyse
Experiment Configuration Mode:	 Simple Advanced
Open Sa	we New
Results Destination	
JDBC database 🗘 URL: jdbc:idb=experiments.prp	Browse)
Experiment Type	Iteration Control
Cross-validation	Number of repetitions: 10
Number of folds: 10	• Data sets first
Classification Regression	Algorithms first
Datasets	Algorithms
Add new Delete selected Use relative paths //Users/eibe/Documents/datasets/UCI/iris.arff //Users/eibe/Documents/datasets/UCI/vote.arff //Users/eibe/Documents/datasets/UCI/glass.arff	Add new Delete selected J48 -C 0.25 -M 2 NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a NaiveBayes
Notes	

🖯 🖯 💛 Weka Experime	Weka Experiment Environment			
Setup Ru	n Analyse			
Experiment Configuration Mode:	Shaple Advanced			
Open Sa	ve New			
Results Destination				
JDBC database 🗘 URL: jdbc:idb=experiments.prp	Browse)			
Experiment Type	Iteration Control			
Cross-validation	Number of repetitions: 10 Data sets first Algorithms Add new Delete selected J48 -C 0.25 -M 2 NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a NaiveBayes			
Notes				

🖯 🖯 Weka Experiment Environment			
	Setup Run Analyse		
(Start	Stop	
Log			
Status Not running			



😑 😑 💮 Weka Experiment Environment			
	Setup Run Analyse		
Start		Stop	
Log			
10:33:04: Started 13:41:15: Finished 13:41:15: There were 0 errors			
Status			
Not running			


😁 🖯 💭 Weka Experiment Environment							
			Setup	Run	Analyse		
Source							
No source						File Database Experiment	
Configure test			Test out	put			
Row key fields	Select keys	\supset					
Run field		× v					
Column key fields	Select keys	\supset					
Comparison field		× v					
Significance	0.05						
Test base	Select base	\bigcirc					
Show std. deviations							
Perform test	Save output	\supset					
Result list		_					

$\Theta \Theta \Theta$	it					
			Setup	Run	Analyse	
Source						
No source						File Database Experiment
Configure test			Test out	put		
Row key fields	Select keys	\bigcirc				
Run field		A Y				
Column key fields	Select keys	\supset				
Comparison field		× v				
Significance	0.05					
Test base	Select base	\bigcirc				
Show std. deviations						
Perform test	Save output	\supset				
Result list						
1			- 1			





University of Waikato

The Knowledge Flow GUI

- New graphical user interface for WEKA
- Java-Beans-based interface for setting up and running machine learning experiments
- Data sources, classifiers, etc. are beans and can be connected graphically
- Data "flows" through components: e.g.,
 "data source" -> "filter" -> "classifier" -> "evaluator"
- Layouts can be saved and loaded again later





































Conclusion: try it yourself!

WEKA is available at

http://www.cs.waikato.ac.nz/ml/weka

Also has a list of projects based on WEKA

WEKA contributors:

Abdelaziz Mahoui, Alexander K. Seewald, Ashraf M. Kibriya, Bernhard Pfahringer , Brent Martin, Peter Flach, Eibe Frank ,Gabi Schmidberger ,Ian H. Witten , J. Lindgren, Janice Boughton, Jason Wells, Len Trigg, Lucio de Souza Coelho, Malcolm Ware, Mark Hall ,Remco Bouckaert , Richard Kirkby, Shane Butler, Shane Legg, Stuart Inglis, Sylvain Roy, Tony Voyle, Xin Xu, Yong Wang, Zhihai Wang