Machine Learning with WEKA

Bernhard Pfahringer (based on material by Eibe Frank, Mark Hall, and Peter Reutemann)

Department of Computer Science, University of Waikato, New Zealand

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

WEKA: the bird

The Weka or woodhen (Gallirallus australis) is an endemic bird of New Zealand. (Source: WikiPedia)



Copyright: Martin Kramer (mkramer@wxs.nl)

WEKA: the software

- Machine learning/data mining software written in Java (distributed under the GNU Public License)
- 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

History

Project funded by the NZ government since 1993

FRST App Number: 93-WKT-23-719 **7. PROGRAMME GOAL** (State the overall goal of the programme in a maximum of 5 lines).

The programme aims to build a state-of-the-art facility for developing techniques of machine learning and investigating their application in key areas of the New Zealand economy. Specifically we will create a workbench for machine learning, determine the factors that contribute towards its successful application in the agricultural industries, and develop new methods of machine learning and ways of assessing their effectiveness.

- Develop state-of-the art workbench of data mining tools
- Explore fielded applications
- Develop new fundamental methods

History (2)

- Late 1992 funding was applied for by Ian Witten
- 1993 development of the interface and infrastructure
 - WEKA acronym coined by Geoff Holmes
 - WEKA's file format "ARFF" was created by Andrew Donkin ARFF was rumored to stand for Andrew's Ridiculous File Format
- Sometime in 1994 first internal release of WEKA
 - TCL/TK user interface + learning algorithms written mostly in C
 - Very much beta software
 - Changes for the b1 release included (among others):
 "Ambiguous and Unsupported menu commands removed."
 "Crashing processes handled (in most cases :-)"
- October 1996 first public release: WEKA 2.1

History (3)

- July 1997 WEKA 2.2
 - Schemes: 1R, T2, K*, M5, M5Class, IB1-4, FOIL, PEBLS, support for C5
 - Included a facility (based on Unix makefiles) for configuring and running large scale experiments
- Early 1997 decision was made to rewrite WEKA in Java
 - Originated from code written by Eibe Frank for his PhD
 - Originally codenamed JAWS (JAva Weka System)
- May 1998 WEKA 2.3
 - Last release of the TCL/TK-based system
- Mid 1999 WEKA 3 (100% Java) released
 - Version to complement the Data Mining book
 - Development version (including GUI)

The GUI back then...

	WEKA 2.1 – Machine	ELearning Wor	rkbench		∇	
<u>W</u> EKA <u>T</u> rain	T <u>e</u> st <u>S</u> cheme <u>R</u> esults			<u>A</u> dvanced <u>I</u>	<u>H</u> elp	
Training File:	golf.arff	Scheme:	No scheme currently selected			
Testing File:	No Testing File	Description:				
Relation:	golf		No Scheme			
Attributes:	5 Instances: 14					
Attributes		Attribute li	nformation			
Includ	e All Exclude All Prune	Name:	outlook			
		Type:	Enumerated			
Number I	nfo Name	Missing:	0			
1	outlook	Value		Count		
2	✓ I temperature	sunny		5	$ \Delta $	
3	Image:	overcast		4		
4	V Windy	rain		5	- 11	
5	Class					
		$\overline{\mathbf{A}}$			\Box	
System Log						
12:49:34: (c) 1993-1996 The University of Waikato, Hamilton, New Zealand 12:49:34: WWW: http://www.cs.waikato.ac.nz/~ml/ 12:49:34: email: wekasupport@cs.waikato.ac.nz 12:49:34: using preval as the rule evaluator Image: Comparison of Compari						
13:24:46: L 13:25:57: L	oaded ARFF File - /home/ml/wekalite2.1/datasets. oaded ARFF File - /home/ml/wekalite2.1/datasets.	.lite/golf.ar: .lite/golf.ar:	tt ff			
Status				and the second second		
Okay			Schemes Running 075			

Ì

WEKA: versions

There are several versions of WEKA:

- WEKA 3.4: "book version" compatible with description in data mining book
- WEKA 3.5.5: "development version" with lots of improvements
- This talk is based on a nightly snapshot of WEKA 3.5.5 (12-Feb-2007)

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 – nominal attribute @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

. . .

java weka.gui.GUIChooser



5/14/2007

University of Waikato



5/14/2007

University of Waikato



5/14/2007

University of Waikato

java -jar weka.jar

Prepr	ocess Classify Clust	Explorer er Associate Select attributes Visualize }
Filter None	URL) Open DB	Generate) Undo Edit Save
Current relation Relation: None Instances: None	Attributes: None	Selected attribute Name: None Type: None Missing: None Distinct: None Unique: None
Attributes All None	Invert Patte	ern
		C Visualize
	Remove	

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, ...

$\Theta \Theta \Theta$		We	ka Explorer		
	Preprocess	Classify Cluster	Associate Select at	tributes Visu	alize }
Open file)	Open URL	Open DB	Generate) Unde	Edi	it Save
Filter Choose None					Apply
Current relation Relation: None Instances: None	Att	tributes: None	Selected attribute Name: None Missing: None	Distinct: No	Type: None one Unique: None
Attributes	None	Invert Pattern			Visualize All

Welcome to the Weka Explorer

Log 💉 x 0

) 🖯 🖯			Weka Explorer		
	Preprocess	Classify Cluste	er Associate Select attr	ibutes Visualize]
Open file)	Open URL	Open DB)	Generate Undo	Edit	Save
Filter Choose None					Apply
Current relation Relation: None Instances: None	At	tributes: None	Selected attribute Name: None Missing: None	Distinct: None	Type: None Unique: None
	None	Invert Patte			Visualize Al

Status Welcome to the Weka Explorer



🖯 🖯 🖯 Weka 🛛	Explorer	
Preprocess Classify Cluster A	ssociate Select attributes Visualize	
Open file Open URL Open DB Gene	erate Undo Edit Save)
Filter Choose None	A	pply
Current relation Relation: iris Instances: 150 Attributes: 5	Selected attribute Name: sepallength Missing: 0 (0%) Distinct: 35 Unique: 9 (6%	eric 6)
Attributes All None Invert Pattern No. Name 1 sepallength 2 sepalwidth 3 petallength 4 petalwidth 5 class	Statistic Value Minimum 4.3 Maximum 7.9 Mean 5.843 StdDev 0.828 Class: class (Nom) Visuali 30 34 28 25 16 10	ize All
(Remove)		1
	4.3 6.1	7.9

Status – OK

Log 💉 x 0



Status – OK

Log 🗸

A x 0

🖯 🖯 🖯 Weka I	xplorer	
Preprocess Classify Cluster A	ssociate Select attributes Visu	alize
Open file Open URL Open DB Gene	rate) Undo Ed	it Save
Filter Choose None		Apply
Current relation	Selected attribute	
Relation: iris Instances: 150 Attributes: 5	Name: class Missing: 0 (0%) Distinct: 3	Type: Nominal Unique: 0 (0%)
Attributes All None Invert Pattern No. Name 1 sepallength 2 sepalwidth	Label Iris-setosa 50 Iris-versicolor 50 Iris-virginica 50	Count
4 petalwidth 5 class	Class: class (Nom)	Visualize All
Remove	50 50	50

Status-	
ОК	



		We	ka Explorer			
	Preprocess	Classify Cluster	Associate Select	attributes	Visualize	
Open file)	Open URL	Open DB	enerate Ur	ndo C	Edit	Save
Filter Choose None						Apply
Current relation Relation: iris Instances: 150	А	ttributes: 5	Selected attribu Name: class Missing: 0 (0%	ite 6) Distir	nct: 3	Type: Nominal Unique: 0 (0%)
Attributes All No. 1 sepalle 2 sepalwi 3 petaller 4 petalwi	None of the second seco	Invert Pattern Name	Li Iris-setosa Iris-versicolor Iris-virginica Class: class (No	abel om)	50 50 50	Count
5 class	Remove		50	50		50

Status	-
ОК	



$\Theta \Theta \Theta$



$\Theta \Theta$	eka Explorer
Preprocess Classify Cluster	Associate Select attributes Visualize
Open file Open URL Open DB	Generate) Undo Edit Save
Filter Choose None	Apply
Current relation Relation: iris Instances: 150 Attributes: 5	Selected attribute Name: petallength Type: Numeric Missing: 0 (0%) Distinct: 43 Unique: 10 (7%)
Attributes All None Invert Patter No. Name 1 sepallength 2 sepalwidth	StatisticValueMinimum1Maximum6.9Mean3.759StdDev1.764
3 petallength 4 petalwidth 5 class	Class: class (Nom) Visualize A
Remove	2 1 3.95

St	atu	s-
0	к	





Status – OK

Log 🗸

P. × 0







⊖ ⊖ ⊖	Weka B	Explorer			
Preprocess C	lassify Cluster A	ssociate Select a	ttributes	Visualize]
Open file Open URL	Open DB Gene	erate Und	•) (Edit	Save
Choose Discretize -B 10 -M -1.0	–R first–last				Apply
Current relation		Selected attribute			
Relation: iris Instances: 150 Attrib	outes: 5	Name: petalle Missing: 0 (0%)	ngth Distin	ct: 43	Type: Numeric Unique: 10 (7%)
Attributes All None Inv No. Na 1 sepallength 2 sepalwidth	ert Pattern me	Statis Minimum Maximum Mean StdDev	tic	1 6.9 3.759 1.764	Value
3 petallength 4 petalwidth 5 class		Class: class (Non	1) 34	47	Visualize All
Remove		3			
L		1		3.95	6.9

Status	
ОК	





S	t	atu	IS-	
()	к		



P. × 0

7.54	-	-	-
1	1		()) ·
- 0	- 9-14		1.1

Weka Explorer

	General General General		Save
lter	weka.gui.Generic	ObjectEditor	Ĩ.
Choose Discretize	About	scretize	Apply
urrent relation Relation: iris Instances: 150 ttributes	An instance filter that discretizes a rang attributes in the dataset into nominal att	e of numeric tributes. Capabilities	Type: Numeric Unique: 10 (7%) Value
	attributeIndices	first-last	
No.	bins	10	1
1 - sepallength 2 - sepalwidth	desiredWeightOfInstancesPerInterval	-1.0	
3 petallength 4 petalwidth	findNumBins	(False	Visualize Al
5 Class	ignoreClass	(False	
	invertSelection	False	47
	makeBinary	(False	16
	useEqualFrequency	False 🛟	

2.5	-	-	~
- 1	100		6-3-
- 1	1	0	0
_			_

Weka Explorer

	General General		Jave
ilter	weka.gui.Generic	ObjectEditor	
Choose Discretize	About	scretize	Apply
Current relation Relation: iris Instances: 150	An instance filter that discretizes a rang attributes in the dataset into nominal att	tributes. More Capabilities	Type: Numeric Unique: 10 (7%) Value
		first last	Faroc
	attributeindices	Tirst-last)
No.	bins	10	1
1 🔄 sepallength 2 🗌 sepalwidth	desiredWeightOfInstancesPerInterval	-1.0	
3 petallength 4 petalwidth	findNumBins	(False	Visualize Al
5 Class	ignoreClass	False 🛟	
	invertSelection	False 🛟	47
	makeBinary	(False	16
	useEqualFrequency	False	

0 Weka Explorer Preprocess Classify Select attributes Cluster Associate Visualize Open file... Undo Open URL... Open DB... Edit... Save... Generate... $\Theta \Theta \Theta$ weka.gui.GenericObjectEditor Filter weka.filters.unsupervised.attribute.Discretize Choose Discretize Apply About Current relation An instance filter that discretizes a range of numeric More Relation: iris Type: Numeric attributes in the dataset into nominal attributes. Instances: 150 Unique: 10 (7%) Capabilities Value Attributes attributeIndices first-last No All bins 10 No. 1 sepallength desiredWeightOfInstancesPerInterval -1.0 sepalwidth 2 petallength 3 + False findNumBins \$ Visualize All 4 petalwidth 5 class + False ignoreClass 47 + False invertSelection + makeBinary False 16 + True useEqualFrequency 6.9 Save... OK Cancel Open...

Log

x 0

Status

OK

0 Weka Explorer Preprocess Classify Select attributes Cluster Associate Visualize Open file... Undo Open URL... Open DB... Edit... Save... Generate... $\Theta \Theta \Theta$ weka.gui.GenericObjectEditor Filter weka.filters.unsupervised.attribute.Discretize Choose Discretize Apply About Current relation An instance filter that discretizes a range of numeric More Relation: iris Type: Numeric attributes in the dataset into nominal attributes. Instances: 150 Unique: 10 (7%) Capabilities Value Attributes attributeIndices first-last No All 10 bins No. 1 sepallength desiredWeightOfInstancesPerInterval -1.0 sepalwidth 2 petallength 3 + False findNumBins \$ Visualize All 4 petalwidth 5 class + False ignoreClass 47 + False invertSelection + makeBinary False 16 + True useEqualFrequency 6.9 OK Cancel Open... Save... Status

OK

Log

x 0

🖯 🖯 💮 Weka B	Explorer	
Preprocess Classify Cluster A	ssociate Select attributes Visualize	
Open file Open URL Open DB Gene	rate) Undo Edit Save	
Filter Choose Discretize -F -B 10 -M -1.0 -R first-last	A	pply
Current relation	Selected attribute	
Relation: iris Instances: 150 Attributes: 5	Name: petallengthType: NumeMissing: 0 (0%)Distinct: 43Unique: 10 (7)	eric %)
Attributes	Statistic Value	
All None Invert Pattern No. Name 1 sepallength 2 sepalwidth	Minimum 1 Maximum 6.9 Mean 3.759 StdDev 1.764	
4 petalwidth 5 class	Class: class (Nom)	ze All
Remove	50 47 34 16	
	1 3.95	6.9

Status OK

Log



St	atu	s-
0	к	



x 0

$\Theta \Theta \Theta$			Wek	a Explorer					
(Preprocess	Classify	Cluster	Associate	Select attrib	outes	Visualize)	
Open file	Open URL	Open DI	3) (Ge	nerate)	Undo		Edit		Save
Filter Choose Disc	retize -F -B 10 -	M – 1.0 – R f	irst-last						Apply
Current relation Relation: iris- Instances: 150	weka.filters.unsu	pervised.att ttributes: 5	ribute.Dis	Selecte Nar Missir	d attribute ne: petallengtl ng: 0 (0%)	n Distine	ct: 10	Type: Unique:	Nominal 0 (0%)
Attributes All No. 1 sepa 2 sepa 3 peta 4 peta 5 class	None Illength Iwidth Ilength Iwidth	Invert Name	Pattern	'(-inf-1 '(1.35- '(1.45- '(1.55-) '(3.4-4) '(4.15- '(4.15- '(4.15- Class:	Label .35]' 1.45]' 1.55]' 3.4]' .15]' 4.55]' 4.55]' class (Nom)	16	11 12 14 16 16 18 17	Count	/isualize All
	Remove								
Status									

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, ...

 $\bigcirc \bigcirc \bigcirc \bigcirc$

Weka Explorer

		56773				
Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier Choose ZeroR						
Test options		lassifier ou	tput			
O 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 option	s)					



 $\bigcirc \bigcirc \bigcirc \bigcirc$

Weka Explorer

Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier Choose ZeroR						
Test options		lassifier ou	tput			
 Ose training set Supplied test set Cross-validation Folds Percentage split More options 						
(Nom) class Start Stop Result list (right-click for option	s)					

Status – OK



0	0	0	
0	0	0	

Weka Explorer

	15/1						
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier							
📁 weka							
🔻 📁 classifiers							
🕨 🕨 📁 bayes			sifier ou	tput			
functions							
🕨 📁 lazy							
🕨 📁 meta							
🕨 🕨 🧊 mi							
🕨 📁 misc							
🕨 📁 trees							
🔻 📁 rules							
🔄 Conju	nctiveRule						
Decisi	onTable						
JRip							
M5Rul	les						
NNge							
OneR							
PART							
Prism							
Ridor	1						

Status	
ОК	Log x 0



Weka Explorer

F	reprocess	Classify	Cluster	Associate	Select attrib	outes Visual	lize
Classifier							
weka	25.002						
haves			sifier ou	tout			
▶ inctions							
▶ 🗇 lazy							
► 📁 meta							
▶ 📁 mi							
▶ 📁 misc							
🔻 🧊 trees							
ADTree							
BFTree							
DecisionS	itump						
Id 3							
(
LMT							
M5P							
NBTree	3						
Random	orest						
Kandom I	ree		4				
KEPTree	-		* *				
SimpleCa	n		<u> </u>				
Eilter) Re	move filter	Close					
	-	111					

Status	
OK	Log x C







Status	
ОК	





\varTheta 🔿 🕘 🛛 w	eka.gui.GenericObjectEditor	-
weka.classifiers.trees.J	148	ze
About		-
Class for generating a	pruned or unpruned C4. More	
	Capabilities	
		E.
binarySplits	False	
confidenceFactor	0.25	
debug	False	
minNumObj	2	
numFolds	3	
reducedErrorPruning	False	
saveInstanceData	False	
seed	1	
subtreeRaising	True	
unpruned	False	
useLaplace	False	
Open)	Save OK Cancel	
	About Class for generating a binarySplits confidenceFactor debug minNumObj numFolds reducedErrorPruning saveInstanceData seed subtreeRaising unpruned useLaplace Open	 weka.gui.GenericObjectEditor weka.classifiers.trees.J48 About Class for generating a pruned or unpruned C4. More Capabilities binarySplits False confidenceFactor 0.25 debug False false minNumObj 2 numFolds 3 reducedErrorPruning False saveInstanceData False seed unpruned False unpruned False unpruned False Cancel

000	\varTheta 🔿 🔿 🛛 🛛 we	eka.gui.GenericObjectEditor	-
Pre	weka.classifiers.trees.J	48	ze
Classifier	About		
Choose J48 -C 0.2!	Class for generating a p	oruned or unpruned C4. More	-
est options		Capabilities	
🔿 Use training set			
O Supplied test set	binarySplits	False	
• Cross-validation F	confidenceFactor	0.25	
More option	debug	False	
	minNumObj	2	
(Nom) class	numFolds	3	
Start C	reducedErrorPruning	False	
lesult list (right-click f	saveInstanceData	False 🗘	
	seed	1	
	subtreeRaising	True	
	unpruned	False	
	useLaplace	False 🗘	
atus	(Open) (Save) (OK) (Cancel	
<			Log







Status	
ОК	











Status	
ОК	Log

x 0



St	ta	tu	s	-
C	ж			







OK

1	100	
1	LUG	1

x 0







Status
ОК



Weka Explorer

Preproce	ss Classify Cluster Associate Select attributes Visualize
Classifier	
Choose J48 -C 0.25 -M	2
Test options	Classifier output
O Use training set	=== Run information ===
O Supplied test set	Set Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2 Relation: iris
Cross-validation Folds	10 Instances: 150 Attributes: 5 66 sepallength
More options	sepalwidth petallength petalwidth class
(Nom) class	Test mode: split 66% train, remainder test === Classifier model (full training set) ===
Start S	top J48 pruned tree
Result list (right-click for opt	ions) petalwidth <= 0.6: Iris-setosa (50.0)
11:00:59 - trees.J48	petalwidth <= 1.7 petallength <= 4.9: Iris-versicolor (48.0/1.0) petallength > 4.9 petalwidth <= 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
	Size of the tree : 9



Weka Explorer

	Weka Explorer
Preproce	ess Classify Cluster Associate Select attributes Visualize
Classifier	2
148 -C 0.23 -M	۷
Test options	Classifier output
O Use training set	=== Run information ===
O Supplied test set	Set Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2 Relation: iris
 Cross-validation Folds Percentage split % 	10 Instances: 150 Attributes: 5 66 sepallength
More options	petallength petalwidth class
(Nom) class	=== Classifier model (full training set) ===
Start S	Stop J48 pruned tree
Result list (right-click for opt	tions) petalwidth <= 0.6: Iris-setosa (50.0)
11:00:59 - trees.J48	petalwidth > 0.0 petalwidth <= 1.7 petallength <= 4.9: Iris-versicolor (48.0/1.0) petallength > 4.9 petalwidth <= 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
	Size of the tree : 9



Weka Explorer

	frend Enpiorer	
Preprocess	Classify Cluster Associate Select a	ttributes Visualize
Classifier		
Choose J48 -C 0.25 -M 2		
est options	Classifier output	
🔾 Use training set		
O Supplied test set Set	Time taken to build model: 0.09	seconds
Cross-validation Folds 10	=== Summary ===	
Percentage split % 66 More options	Correctly Classified Instances Incorrectly Classified Instances Kappa statistic Mean absolute error Root mean squared error	49 96.0784 % S 2 3.9216 % 0.9408 0.0396 0.1579
Nom) class	Relative absolute error Root relative squared error Total Number of Instances	8.8979 % 33.4091 % 51
Start Stop	=== Detailed Accuracy By Class	
esult list (right-click for options)	TP Rate FP Rate Precision 1 0 1 1 0.063 0.905	Recall F-Measure ROC Area Class 1 1 1 Iris 1 0.95 0.969 Iris
1.00.33 - 4003,940		0.882 0.938 0.967 1115
	a b c < classified as 15 0 0 a = Iris-setosa 0 19 0 b = Iris-versicolo: 0 2 15 c = Iris-virginica	r
	IC.) 4 >



Weka Explorer

			1.11						
	Preprocess	Classify	Cluster	Associate	Select at	tributes	Visualize]	
Classifier	C 0.25 -M 2								
est options		C	lassifier ou	tput					
O Use training se	et								
O Supplied test s	set Set.		ime taken	to build mod	del: 0.09	seconds			
🔿 Cross-validati	on Folds 10		== Summary	y ===	spiit ===				
Percentage spl More	lit % 66 options		Correctly (Incorrectly Cappa station Content absolution Content absolution	Classified In Classified Stic The error Squared error	nstances Instances r		49 2 0.9408 0.0396 0.1579	96.0 3.9	0784 % 0216 %
(Nom) class		÷ F	Relative an Root relati Notal Numbe	solute error ve squared e er of Instan	r error ces		8.8979 % 33.4091 % 51		
Start	Stop		== Detaile	ed Accuracy I	By Class =				
tesult list (right-c	lick for options	5)	TP Rate I 1 1	P Rate Pro 0 0.063	ecision 1 0.905	Recall 1 1	F-Measure 1 0.95	ROC Area 1 0.969	Class Iris Iris
1.00.33 - 0283.34	°		0.882	0 Lon Matrix ==	1	0.882	0.938	0.967	Iris
			a b c 15 0 0 0 19 0 0 2 15	<pre>< classi: a = Iris-: b = Iris-: c = Iris-:</pre>	fied as setosa versicolor virginica				
							*****)4 +



 $\Theta \Theta \Theta$

Weka Explorer

				na Explorer				
Pre	process	Classify	Cluster	Associate	Select attribut	es Visualize]	
lassifier								
Choose J48 -C 0.25	5 -M 2							
est options			lassifier ou	tput				
🔘 Use training set								
O Supplied test set	Set.	Т	ime taken	to build mod	lel: 0.09 second	is		
			== Evaluat	ion on test	split ===			
Cross-validation	olds 10		Sululary	·		10.021		
Percentage split	% 66	CI	correctly (Classified In Classified	istances Instances	49 2	96.0)784 % 9216 %
More option	ns	K	appa stati	lstic te error		0.9408		
		R	toot mean s	squared error		0.1579		
(Nom) class		¢ R	cot relati	ve squared e	rror	33.4091 %		
		Т	otal Numbe	er of Instand	es	51		
Start	Ste	View in ma	in window	N Accuracy I	y Class ===			
Result list (right-click fo	or ontic	View in set	parate wir	dow	Recal	F-Measure	ROC Area	Class
1100-ED trace 148	or optic	Save result	t buffer		0.905 1	0.95	0.969	Iris
11:00:59 - trees.J48		Delete resu	ult buffer		0.88	0.938	0.967	Iris
		Load mode	el					
		Save mode	el a la la					
		Re-evaluat	te model o	on current te	est set r			
			0 2 15					
		Visualize c	lassifier e	rrors				
		visualize t	ree			*****) + +
		visualize n	nargin cui	rve				
atus		Visualize t	nreshold	curve		C	log	-
K		visualize c	ost curve		,	0	LUG	100



Caratura	
Status	
OK	(Log) X0
OK	

 $\Theta \Theta \Theta$

Weka Explorer

			Sector Sector					
Prep	rocess	Classify	Cluster	Associate	Select attribute	s Visualize]	
Choose 148 -C 0 25	-M 2						~	
(cildose) j +0 c 0.25	141 2							
Cest options Use training set		ſ	Classifier ou	itput				
O Supplied test set O Cross-validation Fo	Set Ids 10		Time taken === Evaluat === Summary	to build moo tion on test y ===	del: 0.09 second: split ===	3		
Percentage split % 66 More options			Correctly Classified Instances Incorrectly Classified Instances Kappa statistic Mean absolute error Root mean squared error			49 2 0.9408 0.0396 0.1579	96.0784 % 3.9216 %	
(Nom) class		•	Relative and Root relat: Total Number	ive squared e er of Instance	error ces	8.8979 % 33.4091 % 51		
Start Result list (right-click for 11:00:59 - trees.J48	Stop options	View in View in Save re Delete Load m Save m	n main wind separate v esult buffer result buff nodel	dow window er	Recall 1 0.882	F-Measure 1 0.95 0.938	ROC Area 1 0.969 0.967	Class Iris Iris Iris
		Re-eva Visuali	lluate mod	el on current r errors	test set			
atus		Visuali Visuali Visuali	ze tree ze margin ze thresho	curve Id curve	•		•)++
Ж		Visuali	ze cost cur	ve	•	C	Log	100



Status	
ОК	Log 💞

x 0

Weka Explorer

	frend Enpiorer	
Preprocess	Classify Cluster Associate Select a	ttributes Visualize
Classifier		
Choose J48 -C 0.25 -M 2		
est options	Classifier output	
🔾 Use training set		
O Supplied test set Set	Time taken to build model: 0.09	seconds
Cross-validation Folds 10	=== Summary ===	
Percentage split % 66 More options	Correctly Classified Instances Incorrectly Classified Instances Kappa statistic Mean absolute error Root mean squared error	49 96.0784 % S 2 3.9216 % 0.9408 0.0396 0.1579
Nom) class	Relative absolute error Root relative squared error Total Number of Instances	8.8979 % 33.4091 % 51
Start Stop	=== Detailed Accuracy By Class	
esult list (right-click for options)	TP Rate FP Rate Precision 1 0 1 1 0.063 0.905	Recall F-Measure ROC Area Class 1 1 1 Iris 1 0.95 0.969 Iris
1.00.33 - 4003,940		0.882 0.938 0.967 1115
	a b c < classified as 15 0 0 a = Iris-setosa 0 19 0 b = Iris-versicolo: 0 2 15 c = Iris-virginica	r
	IC.) 4 >



 $\Theta \Theta \Theta$

Weka Explorer

0.0		inter (a						
Preprocess	Classify	Cluster	Associate	e Select	attributes	Visualize)	
Classifier								
Choose J48 -C 0.25 -M 2								
lest options		lassifier ou	tput					
O Use training set								
O Supplied test set	.) 1	rime taken	to build m	odel: 0.0	9 seconds			
O Cross-validation Folds 10		=== Evaluat === Summary	tion on tes	t split =				
Percentage split % 66 More options		Correctly (Incorrectly Kappa stat) Kappa stat)	Classified Classifie Istic Ite error	Instances d Instance	es	49 2 0.9408 0.0396 0.1579	96.0 3.9	784 % 216 %
(Nom) class	•	Relative al Root relati Total Numbe	osolute ern ive squared er of Insta	or l error inces		8.8979 % 33.4091 % 51		
Start Stop		=== Detaile	ed Accuracy	By Class	***			
Result list (right–click for option: 11:00:59 – trees.J48	5)	TP Rate I 1 1 0.882	FP Rate I 0 0.063 0	recision 1 0.905 1	Recall 1 0.882	F-Measure 1 0.95 0.938	ROC Area 1 0.969 0.967	Class Iris Iris Iris
		=== Confusi	ion Matrix					
		a b c 15 0 0 0 19 0 0 2 15	< class a = Iris b = Iris c = Iris	ified as -setosa -versicolo -virginic	or a			
					****	*****) + +

Log 💉 x 0



.... -

00		N N	eka Explo	irer				
Preproces	s Classify C	Cluster	Associ	iate Select	attributes :	Visualize]	
lassifier								
📁 weka 🗴 🗐 classifiers	non-4.10.2-0.	2 -N 5	600 -V 0 -	S 0 -E 20 -H	a			
Dayes		ifier o	utput					
V 🗊 functions								1
GaussianProcesses IsotonicRegression LeastMedSq		taker Evalua Summai	n to build ation on t	d model: 0.0 test split =	9 seconds			
LibSVM LinearRegression Logistic MultilayerPerceptro PaceRegression PLSClassifier	n	ectly prrectl a stat absol mean tive a relat	Classific Ly Classific Listic Lute error squared of absolute of Live squar per of Ins	ed Instances fied Instanc r error error red error stances	ces	49 2 0.9408 0.0396 0.1579 8.8979 % 33.4091 % 51	96.0 3.9)784 % }216 %
RBFNetwork SimpleLinearRegree	sion	Detail	led Accura	acy By Class	; ===			
SimpleLogistic SMO SMOreg SVMreg VotedPerceptron		ate 882 Confus b c	FP Rate 0 0.063 0 sion Matr: < cla	Precision 1 0.905 1 ix === assified as	Recall 1 1 0.882	F-Measure 1 0.95 0.938	ROC Area 1 0.969 0.967	Class Iris Iris Iris
<u>Filter</u> <u>Remove filter</u>	er <u>C</u> lose	19 0 2 15	b = 1 c = 1	ris-versicol ris-virginic	.or ta) 4 +



Minter F.

	weka Explorer			
Preprocess	Classify Cluster Associate Select attributes	Visualize]	
Classifier				
Choose MultilayerPerceptron	-L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a			
Test options	Classifier output			
O Use training set				2
O Supplied test set Set	Time taken to build model: 0.09 seconds			
O Cross-validation Folds 10	=== Evaluation on test split === === Summary ===			
Percentage split % 66 More options	Correctly Classified Instances Incorrectly Classified Instances Kappa statistic Mean absolute error Root mean squared error	49 2 0.9408 0.0396 0.1579	96.0 3.9)784 % 216 %
(Nom) class	Relative absolute error Root relative squared error Total Number of Instances	8.8979 % 33.4091 % 51		
Start Stop	=== Detailed Accuracy By Class ===			
Result list (right-click for options)-	TP Rate FP Rate Precision Recall	F-Measure	ROC Area	Class Iris
11:00:59 - trees.J48	1 0.063 0.905 1 0.882 0 1 0.882	0.95 0.938	0.969 0.967	Iris Iris
	=== Confusion Matrix ===			
	a b c < classified as 15 0 0 a = Iris-setosa 0 19 0 b = Iris-versicolor 0 2 15 c = Iris-virginica			
	C)++

Status-OK

) 🛷 × 0 Log

• 🖯 🖯	😝 🖯 😁 🛛 weka	a.gui.GenericObjectEditor		
Pre	weka.classifiers.functions	s.MultilayerPerceptron	s Visualize	.)
Classifier Choose MultilayerF	A Classifier that uses back instances.	propagation to classify More Capabilities		
Test options O Use training set	GUI	True		6
O Supplied test set (autoPunta	True	S	
Cross-validation	debug	False		
Percentage split	decay	False	49	96.0784 %
More option	hiddenLayers	a	0.9408 0.0396	
(Nom) class	learningRate	0.3	0.1579 8.8979 %	
(,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	momentum	0.2	51	-
Start C	nominalToBinaryFilter	True		
Result list (right-click fo	normalizeAttributes	True	F-Measure	ROC Area Class
11:00:59 - trees.J48	normalizeNumericClass	True	0.95	0.969 Iris 0.967 Iris
	randomSeed	0		
	reset	False		
	trainingTime	500		
	validationSetSize	0) + + (
atus	validationThreshold	20		
K			E	Log

Minter F.

	weka Explorer			
Preprocess CI	assify Cluster Associate Select attributes	s Visualize]	
Classifier				
Choose MultilayerPerceptron -	L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a			
Test options	Classifier output			
🔘 Use training set				6
O Supplied test set Set	Time taken to build model: 0.09 seconds	3		
O Cross-validation Folds 10	=== Evaluation on test split === === Summary ===			
Percentage split % 66 More options	Correctly Classified Instances Incorrectly Classified Instances Kappa statistic Mean absolute error Root mean squared error	49 2 0.9408 0.0396 0.1579	96.0 3.9	784 % 216 %
(Nom) class	Relative absolute error Root relative squared error Total Number of Instances	8.8979 % 33.4091 % 51		
Start Stop	=== Detailed Accuracy By Class ===			
Result list (right click for options)—	TP Rate FP Rate Precision Recall	F-Measure	ROC Area	Class Iris
11:00:59 - trees.J48	0.882 0 1 0.882	0.938	0.967	Iris
	=== Confusion Matrix ===			
	a b c < classified as 15 0 0 a = Iris-setosa 0 19 0 b = Iris-versicolor 0 2 15 c = Iris-virginica			
	C)++







Minter F.

	weka Explorer			
Preprocess Cla	assify Cluster Associate Select attributes	s Visualize]	
Classifier				
Choose MultilayerPerceptron -L	0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a -G -R			
Test options	Classifier output			
O Use training set				0
O Supplied test set Set	Time taken to build model: 5.91 seconds	\$		_
O Cross-validation Folds 10	=== Summary ===			- 1
Percentage split % 66 More options	Correctly Classified Instances Incorrectly Classified Instances Kappa statistic Mean absolute error Root mean squared error	50 1 0.9704 0.0239 0.1101	98.0 1.9	392 % 608 %
(Nom) class	Relative absolute error Root relative squared error Total Number of Instances	5.3594 % 23.2952 % 51		- 1
Start Stop	=== Detailed Accuracy By Class ===			- 1
Result list (right-click for options)	TP Rate FP Rate Precision Recall 1 0 1 1 1 0.031 0.95 1 0 941 0 1 0.941	F-Measure 1 0.974 0.97	ROC Area 1 0.998	Class Iris Iris
	=== Confusion Matrix ===			
	15 0 0 a = Iris-setosa 0 19 0 b = Iris-versicolor 0 1 16 c = Iris-virginica			
	(C)))4+





Weka Explorer

	Preprocess	Classify	Cluster	Associ	ate Select	t attributes	Visualize]	
Classifier									13
📁 weka 🔻 📁 classifiers	Daiye i								
🔻 🗊 bayes		sifier ou	tput						
AODE BayesNet ComplementNaiveBayes HNB NaiveBayes NaiveBayesSimple NaiveBayesUpdateable		es I	<pre>taken to build model: 5.91 seconds Evaluation on test split === Summary === rectly Classified Instances 50 98.0392 % prrectly Classified Instances 1 1.9608 % pa statistic 0.9704 absolute error 0.0239 mean squared error 0.1101 ative absolute error 5.3594 %</pre>						392 % 9608 %
► 📁 functions ► 📁 lazy)E		. relative squared error 23.2952 % .1 Number of Instances 51 Detailed Accuracy By Class ===						
 meta mi misc trees rules 			(ate) 941 Confus b c 0 0 19 0	FP Rate 0 0.031 0 ion Matr: < cla a = In b = In c = T	Precision 1 0.95 1 ix === assified as ris-setosa ris-versicol	Recall 1 0.941	F-Measure 1 0.974 0.97	ROC Area 1 0.998 0.998	Class Iris Iris Iris
<u>Filter</u>	<u>R</u> emove filter) <u>C</u> lose		c = 11	ris-virginic) 4 F

Status ОК





Weka Knowledge Explorer

- × 0

Log

Classifier	Cluster Associate Select attributes	visualize	
Choose NaiveBayes			
Test options	Classifier output		
🔘 Use training set	=== Evaluation on test split ===		0
O Supplied test set Set	=== Summary ===		
O Cross-validation Folds 10	Correctly Classified Instances Incorrectly Classified Instances	50 1	98.0392 % 1.9608 %
• Percentage split % 66	Kappa statistic Mean absolute error	0.9704 0.0239	
More options	Root mean squared error Relative absolute error Root relative squared error	0.1101 5.3594 % 23.2952 %	
(Nom) class	Total Number of Instances === Detailed Accuracy By Class ===	51	
Start Stop Result list (right-click for options)	TP Rate FP Rate Precision Recall	F-Measure	Class Iris-setosa
11:49:05 - trees.i48.i48	0.941 0 1 0.951	0.974	Iris-versicolor Iris-virginica
14:34:28 - functions.neural.NeuralNetwork	=== Confusion Matrix ===		
	a b c < classified as 15 0 0 a = Iris-setosa 0 19 0 b = Iris-versicolor 0 1 16 c = Iris-virginica		
	(C))4 + (

QuickTime?and a TIFF (LZW) decompressor are needed to see this picture.


Weka Knowledge Explorer

Manualt

Log x 0

Classifier	Cluster	Associa	ie Select a	tinbutes	Visualize	
Choose NaiveBayes						
Test options	Classifier o	utput				
O Use training set	=== Eval	uation on	test split =			6
O Supplied test set Set	Connechi				50	00.0202.6
O Cross-validation Folds 10	Incorrecti	y Classifi tly Classi	ed Instances fied Instance	es	1	1.9608 %
• Percentage split % 66	Kappa st Mean abs	atistic olute erro	r		0.9704 0.0239	
More options	Root mea Relative Root rel	n squared absolute ative squa	error error red error	0.1101 5.3594 % 23.2952 %		
(Nom) class	Total Nu	mber of In	stances		51	
Start Stop	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
Result list (right-click for options)	1	0 031	1	1	1	Iris-setosa
11:49:05 - trees.j48.J48	0.941	0	1	0.941	0.97	Iris-virginica
14:34:28 - functions.neural.leuralNetwork	=== Conf	usion Matr	ix ===			
	a b 15 0 0 19 0 1 1	c < cl 0 a = I 0 b = I 6 c = I	assified as ris-setosa ris-versicolo ris-virginico	or a		
	C					4 + (

QuickTime?and a TIFF (LZW) decompressor are needed to see this picture.

Weka Explorer

			1.11					
	Preprocess	Classify	Cluster	Associate	Select attribu	ites Visualiz	e	
Classifier Choose Naive	Bayes							
Test options		C	lassifier ou	tput				
O Use training se	et							1
O Supplied test s	et Set.	.) ¹	ime taken	to build mod	iel: 0.01 seco	nds		
Cross-validatio	on Folds 10	-	== Evaluat == Summary	; ===	split ===			
Percentage split More of	it % 66 options		orrectly o ncorrectly appa stati ean absolu oot mean s	Classified In Classified Stic te error Squared error	nstances Instances	48 3 0.9113 0.0447 0.1722	94.1 5.8	1176 % 8824 %
(Nom) class		R R T	elative ak oot relati otal Numbe	osolute error lve squared e er of Instand	error ces	10.0365 % 36.4196 % 51		
Start) Stop		== Detaile	ed Accuracy H	By Class ===			
Result list (right-cl 13:57:12 - functions	lick for options s.MultilayerPerce	s) T	P Rate I 1 0.947 0.882	FP Rate Pre 0 0.063 0.029	ecision Reca 1 1 0.9 0.9 0.938 0.8	11 F-Measure 1 47 0.923 82 0.909	ROC Area 1 0.988 0.988	Class Iris Iris Iris
14:01:20 - bayes.N	aiveBayes		== Confusi	ion Matrix ==	inted as			
			15 0 0 0 18 1 0 2 15	a = Iris-s b = Iris-v c = Iris-v	setosa versicolor virginica			
))4+

) 🛷 × 0

Log

Status -OK

. .

	Weka Explorer										
	Preprocess	Classify Cluster Associate Select attributes Visualize									
Classifier Choose NaiveB	ayes										
Test options		Classifier output									
 Use training set Supplied test set Cross-validation Percentage split More or 	t Set on Folds 10 t % 66 ptions	Time taken to build model: 0.01 seconds === Evaluation on test split === === Summary === Correctly Classified Instances 48 94.1176 % Incorrectly Classified Instances 3 5.8824 % Kappa statistic 0.9113 Mean absolute error 0.0447 Root mean squared error 0.1722									
(Nom) class Start Result list (right-cli	Stop ck for options	Relative absolute error 10.0365 % Root relative squared error 36.4196 % Total Number of Instances 51 === Detailed Accuracy By Class === TP Rate FP Rate Precision Recall F-Measure ROC Area Class									
13:57:12 - functions 14:01:20 - bayes.Na	.MultilayerPerce liveBayes	P View in separate window 0.947 0.923 0.988 Iris Save result buffer 0.882 0.909 0.988 Iris Delete result buffer 0.882 0.909 0.988 Iris									
		Save model or a A A A A A A A A A A A A A A A A A A									
tatus		Visualize margin curve Iris-setosa									
ок		Visualize threshold curve Iris-virginica									

Weka Classifier Visualize: ThresholdCurve. (Class value Iris-versicolor)



Weka Explorer

-				neku Explo	nei			_	
	Preprocess	Classify	Cluste	r Associ	iate Select	attributes	s Visualize)	
assifier									
🕨 📁 bayes	0.0.6		6						
functions			-						
🕨 📁 lazy			nifier	output					
🕨 📁 meta									
🕨 📁 mi			+ take	n to build	d model: 0.0	1 seconds			
🕨 📁 misc						223			
🔻 📁 trees			Summa	acion on '	cest spiit =				
ADTree			ect la	Classifi	ed Instances	r)	48	94 1	176 >
BFTree			prrect	ly Classi	fied Instances	es	3	5.8	3824 %
Decision!	Stump		ha sta	tistic lute erro	r		0.9113		
ld 3			mear	squared (error		0.1722		
J48			, rela	absolute (tive squa:	error red error		36.4196 %		
LMT			il Nur	aber of Ins	stances		51		
M5P			Detai	led Accura	acy By Class	===			
NBTree			late	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
📄 Random	Forest			0	1	1	1	1	Iri
📄 Random	Tree		947	0.063	0.938	0.947	0.923	0.988	Iri
REPTree			Confi	aion Matr	iv ===				
SimpleCa	art		Conre	ISTON MALL.	IX				
💾 UserClas	sifier		b c	c < cla	assified as ris-setosa				
🕨 📁 rules			v 18	b = I	ris-versicol	or			
	may film	Class	2 15	c = I	ris-virginic	a			
<u>Filter</u> <u>Re</u>	move filter	Liose							
) + +

Status OK

Log 💉 🗴 0

Weka Knowledge Explorer γ. .

	Preprocess	Classify	Cluster	Associate	Select at	tributes	Visualize	
Classifier Choose	rpressor are needed to see this picu							
Test options		11	Classifier ou	tput				
 Use training set Supplied test set Cross-validation Percentage split 	t Set n Folds 10 % 66		=== Evalua === Summar Correctly Incorrect. Kappa stat Mean abso	ation on te ry === Classifie ly Classif: tistic lute error	ast split == i Instances ied Instance	-= ?S	48 3 0.9113 0.0447	94.1176 % 5.8824 %
(Nom) class	options	•	Root mean Relative a Root relat Total Numb	squared es absolute es tive square per of Ins led Accura	rror cror ad error tances cy By Class		0.1722 10.0365 % 36.4196 % 51	
Start Result list (right-click for 11:49:05 – trees.j48 14:34:28 – functions 14:48:05 – bayes.Na	Stop (aptions) .J48 s.neural: Jeura iveBayes	INetwork	TP Rate 1 0.947 0.882 === Confus a b c 15 0 0 0 18 1 0 2 15	FP Rate 0 0.063 0.029 sion Matri: < clas a = Ir: b = Ir: c = Ir:	Precision 1 0.9 0.938 < === asified as is-setosa is-versicolo is-virginica	Recall 1 0.947 0.882	F-Measure 1 0.923 0.909	Class Iris-setosa Iris-versicolor Iris-virginica
			<u> </u>)++
Status								
OK							C	Log X 0









	Weka Explorer										
Preprocess	Classify Cluster Associate Select attributes Visualize										
Classifier											
Choose UserClassifier											
Test options	Classifier output										
 Use training set 	Time taken to build model: 548 06 seconds	os									
O Supplied test set Set	=== Evaluation on test split ===										
O Cross-validation Folds 10	=== Summary ===										
Percentage split % 66 More options	Correctly Classified Instances4996.0784Incorrectly Classified Instances23.9216Kappa statistic0.9408Mean absolute error0.0319Root mean squared error0.1622	8									
(Nom) class	Relative absolute error 7.1634 % Root relative squared error 34.312 % Total Number of Instances 51	ſ									
Start Stop	=== Detailed Accuracy By Class ===										
Result list (right-click for options)	TP Rate FP Rate Precision Recall F-Measure ROC Area Classifier 1 0 1 1 1 1 Ir 1 0.063 0.905 1 0.95 0.969 Ir	ss is is									
13:57:12 - functions.MultilayerPercep 14:01:20 - bayes.NaiveBayes	tron 0.882 0 1 0.882 0.938 0.941 Ir.	is									
14.09.40 - trees.0serClassifier	a b c < classified as 15 0 0 a = Iris-setosa 0 19 0 b = Iris-versicolor 0 2 15 c = Iris-virginica										
		•									

Status-OK



	Weka Explorer										
Preprocess	Classify Cluster Associate Select attributes	Visualize									
Classifier											
Choose UserClassifier											
Test options	Classifier output										
O Use training set	Split on petallength AND petalwidth	(Not in Set): No iris-setos									
O Supplied test set Set	Time taken to build model: 548.06 second	is									
O Cross-validation Folds 10	=== Summary ===										
Percentage split % 66 More options	Correctly Classified Instances Incorrectly Classified Instances Kappa statistic Mean absolute error Root mean squared error	49 96.0784 % 2 3.9216 % 0.9408 0.0319 0.1622									
(Nom) class	Relative absolute error Root relative squared error Total Number of Instances	7.1634 % 34.312 % 51									
Start Stop	=== Detailed Accuracy By Class ===	-									
Result list (right-click for option.	1 0 1 1	r-Measure ROC Area Class									
13:57:12 - functions.MultilayerPercep	tron 1 0.063 0.905 1 0.882 0 1 0.882	0.95 0.969 Iris 0.938 0.941 Iris									
14:09:40 - trees.UserClassifier	=== Confusion Matrix ===										
	a b c < classified as 15 0 0 a = Iris-setosa 0 19 0 b = Iris-versicolor 0 2 15 c = Iris-virginica										
	C)4+									

Status-OK





Status

OK

Weka Explorer

Associate Select attributes Visualize

Preprocess Classify Cluster Classifier UserClassifier Choose Test options Classifier output Spilt on petallength AND petalwidth (Not in Set): No iris-setose Use training set Time taken to build model: 548.06 seconds Supplied test set Set... === Evaluation on test split === === Summary === Cross-validation Folds 10 Correctly Classified Instances 49 96.0784 % (Num) sepallength Incorrectly Classified Instances 3.9216 % 2 (Num) sepalwidth Kappa statistic 0.9408 Mean absolute error 0.0319 (Num) petallength Root mean squared error 0.1622 (Num) petalwidth Relative absolute error 7.1634 % ✓ (Nom) class Root relative squared error 34.312 % Total Number of Instances 51 === Detailed Accuracy By Class === Start Stop Precision Recall F-Measure TP Rate FP Rate ROC Area Class Result list (right-click for options)-Iris 1 0 1 1 1 1 0.905 0.95 0.969 Iris 1 0.063 1 13:57:12 - functions.MultilayerPerceptron 0.882 0.882 0.938 0.941 0 1 Iris 14:01:20 - bayes.NaiveBayes === Confusion Matrix === 14:09:40 - trees.UserClassifier <--- classified as bc a 15 0 0 a = Iris-setosa 0 19 0 b = Iris-versicolor 0 2 15 c = Iris-virginica 4 1

Log





Weka Explorer

	Preprocess	Classify	Cluster	Associ	ate Select	attributes	Visualize]	
Classifier	20								
 weka classifiers bayes functions lazy mi mi misc trees ADTree BFTree Decisi Id 3 	e e onStump		sifier o Split : taken Evalua Summan rectly prrect : absol . mean itive a . relat i Numb	on petal on petal n to build ation on r ry === Classific ly Classific ly Classific ly Classific squared of squared of absolute of tive squared	Tength AND p i model: 548 test split = ed Instances fied Instanc error error error stances	etalwidtn .06 secon == es	49 2 0.9408 0.0319 0.1622 7.1634 % 34.312 %	t): N6 ITIS 96.0 3.9	5-SetOS.
J48 LMT MSP NBTre Rando Rando REPTr Simple	e mForest mTree ee <u>Cart</u> <u>Remove filter</u>) <u>C</u> lose	Detai: (ate 882 Confus b c 0 0 19 0 2 15	led Accurate FP Rate 0 0.063 0 sion Matri < cla a = I b = I c = I	acy By Class Precision 1 0.905 1 ix === assified as ris-setosa ris-versicol ris-virginic	Recall 1 0.882	F-Measure 1 0.95 0.938	ROC Area 1 0.969 0.941	Class Iris Iris Iris

Status	
ок	Log x 0

1

Weka Explorer

		· · · · · · · · · · · · · · · · · · ·	Decision of the second s				
Pre	process	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier Choose M5P - M 4.0	0						
Test options			lassifier ou	tput			
🔘 Use training set		-	== Run inf	ormation ===	£		1
O Supplied test set	Set.	. S	cheme: elation: nstances:	weka.clas iris 150	sifiers.trees.M5P	-M 4.0	
Cross-validation F Percentage split More option	olds 10 % 66 ns		ttributes:	5 sepalleng sepalwidt petalleng petalwidt class	gth ch gth ch		
(Num) petallength		• • • • •	est mode: == Classif 5 pruned π	ier model (f	ull training set)	===	
Result list (right-click for 13:57:12 - functions.Mult	or option: ilayerPerce	s) p	etalwidth etalwidth class=1	<pre>> 0.8 : LM1 > 0.8 : LM1 cris-virginic</pre>	models) (50/9.298%) a <= 0.5 : LM2 (50	0/12.723%)	
14:01:20 - bayes.NaiveBayes 14:09:40 - trees.UserClassifier 14:25:51 - trees.M5P			class=1 M num: 1 etallength	ris-virginic	a > 0.5 : LM3 (50	0/15.631%)	
			0.1 - 0 + 0 + 0	0.1503 * sepail 0.1503 * sepa 0.715 * petal 0.9748	lwidth width		
		L p	etallength	=			

) 🛷 × 0

Log

Status -OK

Weka Explorer

			Station	-			
	Preprocess	Classify	Cluster	Associate	Select attribute	s Visualiz	ze
Classifier Choose M5P	-M 4.0						
Test options			lassifier ou	tput			
 Use training s Supplied test Cross-validat Percentage sp 	set Set ion Folds 10 lit % 66		LM num: 2 petallength 0.5 - (+ 1 + (- (1 = 075 * sepall 0.085 * sepal 0.1314 * peta 0.1083 * clas 0.0257	ength width lwidth s=Iris-virginica	a	
More (Num) petallengt Start	options h Stop		LM num: 3 betallength 0.7 - (+ (+ (+ (Number of F	1 = 278 * sepall 0.085 * sepal 0.2824 * peta 0.1083 * clas 0.3295 cules : 3	ength width width s=Iris-virginica	a	
Result list (right-o	click for option	s)	fime taken	to build mod	lel: 0.72 second	S	
14:01:20 - bayes.	NaiveBayes		== Evaluat == Summary	ion on test	split ===		
14:25:51 - trees.M	ISP	C M H H H J	Correlation Mean absolut Root mean s Relative ab Root relation Total Number	a coefficient ate error squared error solute error ve squared e er of Instanc	error	0.9889 0.1861 0.255 11.9578 % 14.9153 % 51	

Status -OK



Weka Explorer

				eka Explorer			
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Classifier							
Choose M5P -M	14.0						
Test options		C	lassifier ou	itput			
○ Use training set		I	M num: 2				1
O Supplied test set	set.		etallengt	h = 5075 * sepall	length		
	Folds 10		-	0.085 * sepal	lwidth		
 Percentage split 	% 66		+	0.1083 * clas 0.0257	ss=Iris-virginica		
More op	otions		M num: 3 etallengt 0.	h = 7278 * sepall	length		
(Num) petallength		•	- + + +	0.085 * sepal 0.2824 * peta 0.1083 * clas 0.3295	lwidth alwidth ss=Iris-virginica		
Start	Stop		tumber of t	Bulos • 3			
Result list (right-clic	k for option	s) View	in main w	indow to window	onds		
13:57:12 - functions.	MultilayerPerce	eptro Save	result buf	fer			
14:01:20 - bayes.Nai 14:09:40 - trees User	veBayes Classifier	Delet	e result b	uffer			
14:25:51 - trees.M5P		Load	model	0.9889 0.1861			
		Save	model			0.255	
		Re-e	valuate m	odel on curre	ent test set	14.9153 % 51	
		Visua	lize classi	fier errors			-
		Visua	lize tree				
atus		Visua	lize marg	in curve			
к		Visua	lize thres	nold curve		Log	;_) "@^ >
		VISUa	IIIZE COSE (urve	P		

Weka Classifier Visualize: 14:25:51 - trees.M5P (iris)



Weka Classifier Visualize: 14:25:51 - trees.M5P (iris)



Weka Classifier Visualize: 14:25:51 - trees.M5P (iris)



Explorer: clustering data

- WEKA contains "clusterers" for finding groups of similar instances in a dataset
- Some 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 Explorer

00				Ra Explorer			
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Clusterer							
Choose EM -I 1	100 -N -1 -M	1.0E-6 -5	100				
Cluster mode			Clu	sterer output			
🕑 Use training set							
O Supplied test se	et	Set	\supset				
O Percentage split	t	% 66					
Classes to clust	ters evaluation	n					
(Nom) class		÷					
Store clusters for	or visualizatio	'n					
Ign	ore attributes	0	\supset				
Start	\supset	Stop	\supset				
Result list (right–cli	ck for option	5)					

Status -OK

Log

) 🛷 × 0



Status	
ОК	Log 🔊 × 0



Status	
ОК	





 $\bigcirc \bigcirc \bigcirc \bigcirc$

Weka Explorer

	weka Explorer
Preprocess Classify	Cluster Associate Select attributes Visualize
Clusterer	
Choose Cobweb -A 1.0 -C 0.002820947	79177387815 -S 42
Cluster mode	Clusterer output
🔘 Use training set	=== Run information ===
O Supplied test set Set	Scheme: weka.clusterers.Cobweb -A 1.0 -C 0.0028209479
O Percentage split % 66	Instances: 150 Attributes: 5
Classes to clusters evaluation	sepallength sepalwidth
(Nom) class	petallength petalwidth
Store clusters for visualization	Ignored: class
	Test mode: Classes to clusters evaluation on training date === Model and evaluation on training set ===
Ignore attributes	Number of merges: 0
Start Stop	Number of splits: 0 Number of clusters: 3
Result list (right-click for options)	node 0 [150]
16:11:09 - Cobweb	node 0 [150] leaf 2 [54]
	Clustered Instances
	1 100 (67%)
	2 50 (33%)
	Class attribute: class

Status-OK



) 🖯 🖯			We	ka Explorer			
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Clusterer Choose Cobwe	2b −A 1.0 −C 0	0.002820947	791773878	15 -S 42			
Cluster mode Use training set Supplied test set Percentage split Classes to clust (Nom) class Store clusters f Ign Start Result list (right-clit 16:11:09 - Cobweb	t et t ters evaluation for visualization nore attributes ick for options	Set % 66 n stop	Clus Num Num Num Num Num Num Clus Clus Clus Clus Clus Clus Clus Clus	sterer output Model and e ber of merge ber of split ber of clust leaf 1 [96] leaf 2 [54] stered Insta 100 (67 50 (33 ass attribute sses to Clus 2 < ass 0 Iris-s 0 Iris-s 0 Iris-s ster 1 < 1 ster 2 < 1 ster 2 < 1	evaluation on train es: 0 ers: 0 ers: 3 ances 7%) 3%) e: class sters: signed to cluster setosa versicolor virginica tris-versicolor tris-setosa astered instances	ning set ===	33.3333 %

Status-OK



			We	eka Explorer				
	Preprocess	Classify	Cluster	Associate	Select att	ributes	Visualize	
Clusterer								
Choose Cobw	eb -A 1.0 -C 0.	002820947	91773878	315 -S 42				
Cluster mode			Clu	sterer output				
O Use training se	et			= Modei and €	evaluation	on train	ing set ===	6
O Supplied test s	set S	et		nber of merge nber of split nber of clust	es: 0 ts: 0 ters: 3			
O Percentage spl	it	% 66	noo	ie 0 [150]				
Classes to cluster	sters evaluation			leaf 1 [96]	1			
(Nom) class		•	1	leaf 2 [54]	1			
Store clusters	for visualizatior	1	Clu	istered Insta	ances			
Ig	nore attributes			50 (33	75) 38)			
Start	\supset	Stop		ass attribute asses to Clus	e: class sters:			
Result list (right-c	lick for options))		1 2 < ass 0 50 Iris-s	signed to c setosa	luster		
16:11:09 - Cobweb		View in	n main wir	ndow	avai aal av			
		View in	n separate	window		lor		
		Save re	esult buffe	er				
		Delete	result but	ffer		ances :	50.0	33.3333 %
		Load m	nodel					v
		Save m	odel) 4 1
		Re-eva	luate mo	del on currer	nt test set			
Status		Visuali	ze cluster	assignment	s			00 x 0
ОК		Visuali	ze tree					



			We	eka Explorer			
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Clusterer Choose Coby	web -A 1.0 -C (0.002820947	791773878	15 -S 42			1
Cluster mode Use training s Supplied test Percentage sp Classes to clu (Nom) class	set set olit usters evaluatio	Set % 66 n	Clu Nur Nur Nur noc	sterer output Model and ender of merge mber of split mber of clust de 0 [150] leaf 1 [96] de 0 [150] leaf 2 [54]	evaluation on train es: 0 es: 0 ers: 3	ning set ===	
Store clusters	s for visualizatio	Stop		100 (67 50 (33 ass attribute asses to Clus	ances (%) 3%) e: class sters:		
Result list (right-	click for option	s)		1 2 < ass	signed to cluster		
16:11:09 - Cobwe	b View View Save Delet Load Save Re-e	in main wir in separate result buffe e result buf model model valuate mo	ndow window er ffer del on cur	rent test set	rsicolor rginica is-versicolor is-setosa tered instances :	: 50.0	33.3333 %
Status OK	Visua Visua	lize cluster	assignme	ents		Lo	g) 🛷 x





Explorer: finding associations

- WEKA contains the Apriori algorithm (among others) 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



S	t	a	t	u	s	-
()	ĸ				



V \varTheta 🕲 Weka Explorer							
	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Associator							
Choose Ap	riori -N 10 -T 0 -	C 0.9 -D 0.	05 -U 1.0 ·	-M 0.1 -S -1.	0 -c -1		
(choose) h		0.5 0 0.	05 0 1.0				
Start	Stop	A	ssociator o	utput			
Describelies (sinks	- dial factorian						
Result list (right	t-click for options	5)					

S	t	a	t	u	S	
(0	ĸ				





0	0	0
0	0	0

Weka Explorer

Prep	process Classify Cluster Associate Select attributes Visualize)
Associator Choose Apriori -N 1	10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1	
Start Stop Result list (right-click fo 16:23:55 - Apriori	Associator output Apriori Apriori Apriori Apriori Minimum metric <confidence>: 0.9 Number of cycles performed: 11 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(2): 17 Size of set of large itemsets L(3): 6 Size of set of large itemsets L(4): 1 Best rules found: 1. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid- 3. physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> Class=democrat 4. physician-fee-freeze=n education-spending=n 202 ==>> Class=democrat 5. physician-fee-freeze=n 247 ==> Class=democrat 245 conf:(0.99 6. el-salvador-aid=n Class=democrat 200 ==> aid-to-nicaraguan-contras=y 204 co 8. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y 204 co 9. el-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> class=democrat 218 ==> physician-fee</confidence>	==> Class=democr to-nicaraguan-cc s=democrat 210 crat 201 conf:)) cras=y 197 cor onf:(0.98) y Class=democrat mocrat 197 cor e==freeze=n 210

Log

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	
ОК	Log

A x 0
000

Weka Explorer

Attribu	ute select ation mo	Associate tion output de: eva	Select attributes	Visualize
Attribu Evalu === A Searc	ute select ation mo .ttribute	tion output class de: eva	luate on all trair	ning data
Attribu Evalu === A Searc	ute select ation mo .ttribute	tion output de: eva	luate on all trair	ning data
Attribu Evalu === A Searc	ute select ation mo .ttribute	tion output de: eva	luate on all trair	ning data
Evalu === A Searc	ation mo	de: eva	luate on all trair	ning data
Searc	ttribute	Selection	an all danut date	
•	n Method Best	l: first.	on all input data	7 ===
	Start Searc Stale Total Merit	h direction search af number of of best s	n: forward ter 5 node expansi subsets evaluated ubset found: 0.	ions 1: 85 .729
Attri	bute Sub CFS S Inclu ted attr	set Evalua ubset Evalu iding local ibutes: 3, ad ph	tor (supervised, 0 uator ly predictive attr 4,10,11 : 4 loption-of-the-budg ysician-fee-freeze	Class (nominal): 17 Class): ributes get-resolution
	Attri	Attribute Sub CFS S Inclu	Attribute Subset Evalua CFS Subset Evalua Including local Selected attributes: 3, ad ph im	Attribute Subset Evaluator (supervised, of CFS Subset Evaluator Including locally predictive attributes: Selected attributes: 3,4,10,11 : 4 adoption-of-the-budg physician-fee-freeze immigration synfuels-corporation

Status-OK



000

Weka Explorer

Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Attribute Evaluator						
Choose CfsSubsetEval						
Search Method						
Choose BestFirst -D 1 -N 5						
Attribute Selection Mode	At	tribute sel	ection output			
 Use full training set Cross-validation Folds Seed 	10 1 =	valuation == Attribu	mode: eva	luate on all train on all input data	ing data	(
(Nom) Class Start Stop Result list (right-click for options		earch Meth Bes Sta Sea Sta Tot Mer	od: it first. irt set: no a irch directio le search af al number of it of best s	ttributes m: forward ter 5 node expansi subsets evaluated subset found: 0.	ons : 85 729	Π
16:36:05 – BestFirst + CfsSubsetEva	A S	ttribute S CFS Inc elected at	Subset Evalua Subset Eval Subset Eval Subset Eval Subset Eval Subset Sub	tor (supervised, C uator ly predictive attr 4,10,11 : 4 option-of-the-budg ysician-fee-freeze migration nfuels-corporation	lass (nominal): 17 Cl ibutes et-resolution -cutback	.ass):

Status – OK



0	0	0	
0	0	0	

Weka Explorer

	Preprocess	Classify	Cluster	Associate	Select attributes Visualize
Attribute Evaluator					
weka weka weka cfsSubsete ChiSquaree ClassifierSi Consistence GainRatioA InfoGainAt OneRAttrib PrincipalCo ReliefFAttri SVMAttribu Symmetric Symmetric Symmetric WrapperSu	tion Eval dAttributeEval ubsetEval sySubsetEval attributeEval outeEval outeEval alUncertAttribut alUncertAttribut ubsetEval	teEval teSetEval	E308 -N pute sele uation Attribu ch Meth Bes Sta Sea Sta Tot Mer ibute S CFS Inc cted at	te Selection od: tfirst. rt set: no a rch directio le search af al number of it of best s ubset Evalua Subset Evalua luding local tributes: 3, ad ph im	attributes on on all input data ==== attributes on: forward ifter 5 node expansions of subsets evaluated: 85 subset found: 0.729 mator (supervised, Class (nominal): 17 Class): iluator ally predictive attributes 5,4,10,11 : 4 idoption-of-the-budget-resolution obysician-fee-freeze mmigration synfuels-corporation-cutback

Status-	-
ОК	







Status-OK



x 0

Explorer: data visualization

- Visualization very useful in practice: e.g. helps to determine difficulty of the learning problem
- WEKA can visualize single attributes (1-d) and pairs of attributes (2-d)

To do: rotating 3-d visualizations (Xgobi-style)

- Color-coded class values
- "Jitter" option to deal with nominal attributes (and to detect "hidden" data points)
- "Zoom-in" function

🖯 🖯 🖯 Weka I	Explorer
Preprocess Classify Cluster A	ssociate Select attributes Visualize
Open file Open URL Open DB Gene	erate Undo Edit Save
Filter Choose None	Apply
Current relation Relation: Glass Instances: 214 Attributes: 10	Selected attribute Name: RI Type: Numeric Missing: 0 (0%) Distinct: 178 Unique: 145 (68%)
Attributes All None Invert Pattern No. Name 1 RI 2 Na	StatisticValueMinimum1.511Maximum1.534Mean1.518StdDev0.003
3 Mg 4 Al 5 Si 6 K 7 Ca 8 Ba 9 Fe 10 Type	Class: Type (Nom) Visualize All
(Remove)	3 4 3 3 0 1 1.51 1.52 1.53

Status -OK

Log 💉 x 0

000

Weka Explorer



Log

x 0

OK

000

Weka Explorer



2	t	d	ι	u	S
C)	K			



x 0

000

Weka Knowledge Explorer

Log



OK

 $\Theta \Theta \Theta$

Weka Explorer



Log

OK



build wind	float	build wind non-floa	at	vehic wind float
vehic wind	non-float		tableware	headlamps



Class colour

build wind float build wind non-float vehic wind float vehic wind non-float containers tableware headlamps



build wind float vehic wind non-float

build wind non-float tableware vehic wind float headlam



5/14/2007

University of Waikato



5/14/2007

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!

🖶 \ominus 😁 🛛 🛛 🕅	eka Experiment Environment	
(Setup Run Analyse	
Experiment Configuration Mode:	Simple	O <u>A</u> dvanced
<u>Open</u>	<u>S</u> ave) (New
Results Destination ARFF file Filename:		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 Edit select Dele	te sele) (Add new) (Edi	it selected) Delete selected
Up Down	Load optio)	Save optio) Up Down
\subset	Notes	

🖯 🖯 Weka Experime	ent Environment			
Setup Ru	In Analyse			
Experiment Configuration Mode:	Simple ○ Advanced			
<u>Open</u> <u>S</u> a	ve <u>N</u> ew			
Results Destination				
ARFF file Filename: results.arff	Browse)			
Experiment Type	Iteration Control			
Cross-validation	Number of repetitions: 10			
Number of folds: 10	Data sets first			
Olassification ○ Regression	O Algorithms first			
Datasets	Algorithms			
Add new) (Edit select) (Delete sele)	(Add new) (Edit selected) (Delete selected)			
Use relative p	148 - C 0 25 - M 2			
/Users/bernhard/snapshot/glass.arff /Users/bernhard/snapshot/iris.arff /Users/bernhard/snapshot/vote.arff	NaiveBayes AdaBoostM1 -S 1 -I 10 -W weka.classifiers.trees.J48C 0.29			
Up Down	Load optio) Save optio) Up Down			
(Ne	otes			

00		Weka Experiment Environment		
		Setup Run Analyse		
\square	<u>S</u> tart		Stop	
Log				
Status				
Not running				



	Wel	eka Experiment Environment				
		Setup Run Analyse				
Source						
No source		<u>File</u> <u>Database</u> <u>Experiment</u>				
Configure test		Test output				
Testing <u>w</u> ith	Paired T-Tester 🛊					
<u>R</u> ow	Select					
<u>C</u> olumn	Select					
Co <u>m</u> parison field						
Significance	0.05					
<u>S</u> orting (asc.) by		~				
Test <u>b</u> ase	(Select)					
Displayed Columns	Select					
Show std. devi <u>a</u> tions	Θ					
<u>O</u> utput Format	Select					
Perform test	Save output					
Result list						









5/14/2007

University of Waikato



5/14/2007

University of Waikato

The Knowledge Flow GUI

- 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
- cf. Clementine ™







Knowledge Flow Layout

ArffLoader	
¢	A ¥ •
Status Welcome to the Weka Knowledge Flow	Log



Knowledge Flow Layout ArffLoader Data Visualizer 4.1 Status

Welcome to the Weka Knowledge Flow

Log



Knowledge Flow Layout





Knowledge Flow Layout

ARTE	
dataSet	
Data Visualizer	
***********)4
ne to the Weka Knowledge Flow	(-L





Knowledge Flow Layout







Welcome to the Weka Knowledge Flow

5/14/2007






Welcome to the Weka Knowledge Flow

5/14/2007





00	Text View	wer			
Result list	Text				
18:01:14 - SMO	=== Evaluation result ===				
	Scheme: SMO Relation: iris-weka.filters.supervised	i.attribute.A	ttributeSel	ection-Eweka.attrik	outeSele
	Correctly Classified Instances Incorrectly Classified Instances Kappa statistic Mean absolute error Root mean squared error Relative absolute error Root relative squared error Total Number of Instances	144 6 0.94 0.2311 0.288 52 % 60.8201 %	96 4	16 B	
	TP Rate FP Rate Precision Recall 1 0 1 1 0.96 0.04 0.923 0.96 0.92 0.02 0.958 0.92 === Confusion Matrix === a b c $<$ classified as 50 0 0 a = Iris-setosa 0 48 2 b = Iris-versicolor 0 4 46 c = Iris-virginica	L F-Measure 1 0.941 0.939	ROC Area 1 0.96 0.971	Class Iris-setosa Iris-versicolor Iris-virginica	
	C)41





5/14/2007



5/14/2007

Sourceforge.net – Downloads



Sourceforge.net – Web Traffic



Projects based on WEKA

- 45 projects currently (30/01/07) listed on the <u>WekaWiki</u>
- Incorporate/wrap WEKA
 - GRB Tool Shed a tool to aid gamma ray burst research
 - YALE facility for large scale ML experiments
 - ♦ GATE NLP workbench with a WEKA interface
 - Judge document clustering and classification
 - RWeka an R interface to Weka
- Extend/modify WEKA
 - BioWeka extension library for knowledge discovery in biology
 - WekaMetal meta learning extension to WEKA
 - Weka-Parallel parallel processing for WEKA
 - Grid Weka grid computing using WEKA
 - Weka-CG computational genetics tool library

WEKA and PENTAHO

- <u>Pentaho</u> The leader in Open Source Business Intelligence (BI)
- September 2006 Pentaho <u>acquires</u> the Weka project (exclusive license and SF.net page)
- Weka will be used/integrated as data mining component in their BI suite
- Weka will be still available as GPL open source software
- Most likely to evolve 2 editions:
 - Community edition
 - BI oriented edition

Limitations of WEKA

- Traditional algorithms need to have all data in main memory
- ==> big datasets are an issue
- Solution:
 - Incremental schemes
 - Stream algorithms

MOA "Massive Online Analysis"

(not only a *flightless* bird, but also extinct!)

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
- (probably incomplete list of) 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