Behavioral Security:
10 steps forward & 5 steps backward
DeepSec 2011

Sourabh Satish

Distinguished Engineer/ Chief Architect, Symantec

Agenda

1. Threat Landscape
2. Behavioral Security Overview
3. Traditional rules based behavioral security
4. Machine Learning – Supervised and Unsupervised
5. Machine Learning for behavioral security
6. Real world examples
7. Conclusion

Threat Landscape

Motivation?

Threat Landscape

2010-2011 Trends

1. **Targeted Attacks**
continued to evolve

2. **Social Networking**
+ social engineering = compromise

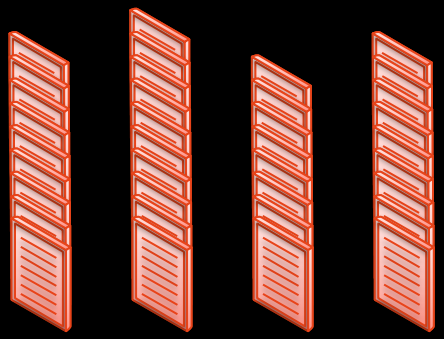
3. **Attack Kits**
get a caffeine boost

4. **Hide and Seek**
(zero-day vulnerabilities and rootkits)

5. **Mobile Threats**
increase

Why is it hard to stop attacks?

Many reasons, one being: Malware authors have switched tactics



From:

A mass distribution of a relatively few threats e.g.

- Storm made its way onto millions of machines across the globe

To:

A micro distribution model e.g.

- The average Vundo variant is distributed to 18 Symantec users!
- The average Harakit variant is distributed to 1.6 Symantec users!

What are the odds a security vendor will discover all these threats?

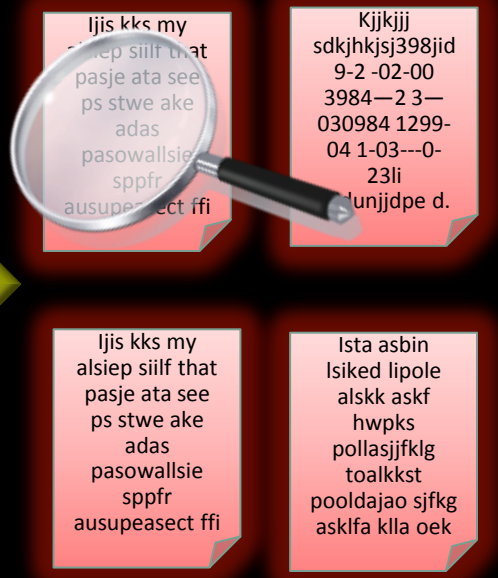
Analyzing the Problem

"Unique" threats are unique at the byte-level

This is my first virus that I plan to use to steal key and passwords from unsuspecting victims.

Hacker develops threat

Hacker uses Tool to obfuscate executable



Tool generates clones that differ at the byte-level

Changes at the byte-level evade traditional file-based pattern-matching engines


v 1.0
Icy Virus Gen
Shut Down Computer
Log Off Computer
Restart Computer
Hide Taskbar
Deactivate Taskmanager
Open CD Tray
Anti Anubis
Anti Kaspersky
Anti AVG
Kill Process
Start Webpage
Change Clock
Create Folder On Desktop
Create File On Desktop
Spam Files On Desktop
Spam Folders On Desktop
Spam Eicar
Process Name
www.Google.com
12
35
Folder Name
File Name
Example.exe
Folders Name
Example.exe
Creation Options
Filename
Created By Icy Wind
Build
About
Exit
Generate
Browse
Crypt/Protect
Use Icon Changer
ResHack
fLaSh.dll
Buy Unique Stub

ONLY $1 FOR 24 HOURS OF CRYPTING!

PURCHASE NOW!

March 25, 2011

Welcome to OnlineCrypter.com !

March 15, 2011

"Professionals. Need I say more?...

Read More

-daemon10, USA

OC Licensing System

Protect your programs with our hardware id/licensing system. With your own full

Examples of Threat Cloning

Number of Clones: 49

Analyzing the Problem

Are these “unique” threats really unique ?

- Bytes change. But how about the behaviors of these threats ?

Password Stealers

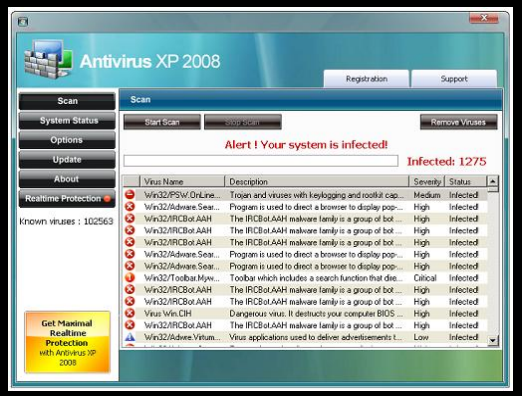
will continue to steal passwords

Spam Bots

will continue to send Spam

Rogue AntiVirus

will continue to popup misleading messages



…behaviors don’t change..

Solving the Problem

Behavior-based Detection

Engine that ignores what the threat **looks like**




But detects threats based on what the threat **does**

Clarifying the terminology

Detection vs. Prevention vs. Protection

- **Detection is "after" the fact**
  - After the sample has run on the system, you analyze the impact and conclude if the action taken was malicious and then remediate the threat and reverse its persistent system changes.
- **Prevention is "before" the fact**
  - You conclude that the action that a sample is about to take is malicious and hence prevent the action from happening in the first place. You remediate the threat and minimal system settings change(restore) is needed.
- **Protection**
  - Both detection based and prevention based technologies can offer protection.
- **Challenges:**
  - **Detection based approach**: Can all changes be reversed? File modified on disk?
  - **Prevention based approach**: Which action do you block and inspect? What is the performance overhead?
- **Debatable!**
  - Blocked the 5th event and hence prevented 6th most impactful event!

Legacy rules based behavioral security

The Legacy Solution

Rules based behavioral security

- Rules to identify malicious activity and take action

The legacy solution
Rules based behavioral security

- **Simple and intuitive model (Expert System)**
  - Domain Experts know how to distinguish between good and bad
  - They analyze the malware, spot the trends/patterns and write rules
  - Product ships with default set of rules & rules are updated regularly
  - The product may also have an ability to let users express new rules in the product
- **Applicability**
  - Many security products, especially enterprise products use this model
  - Maybe the only answer for some threat scenarios
- **Pros**
  - Broader coverage for variants, Precise reasoning for detection, Name the threat, Relevant Actions
- **Cons**
  - Scalability, Domain Expertise

Low error rate?

Addressing the challenge
Scalability

- **Fact:**
  - Behavioral variants are far less than file variants
- **New SHA256 = a file variant OR really a new malware?**
  - Same malware may be packed differently
  - Same malware may be skinned differently
- **Answer:**
  - Analyze the threat?

AUTOMATION **COLLECT DATA** **DATA MINING**

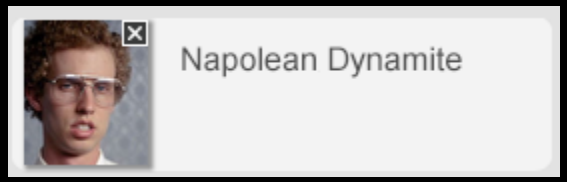
Machine Learning - Basics

Machine Learning

Learning by Example

- **New** approach to AI is to get the computer to program itself by showing it examples (data or past experiences) of behavior we want!
  - This is the ***learning*** approach to AI

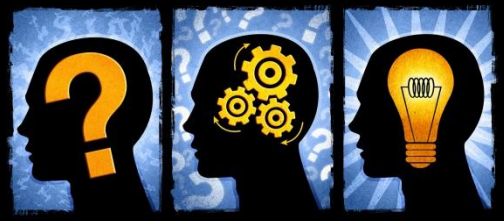
Face



Name

  - Often hand programming is not possible or not a feasible answer like face detectors, handwriting reader, etc.

Machine Learning
What is Machine Learning?



- **Central Question**

  "How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning process?"

- **What is the learning problem?**

  A process learns with respect to **<T, P, E>** if it

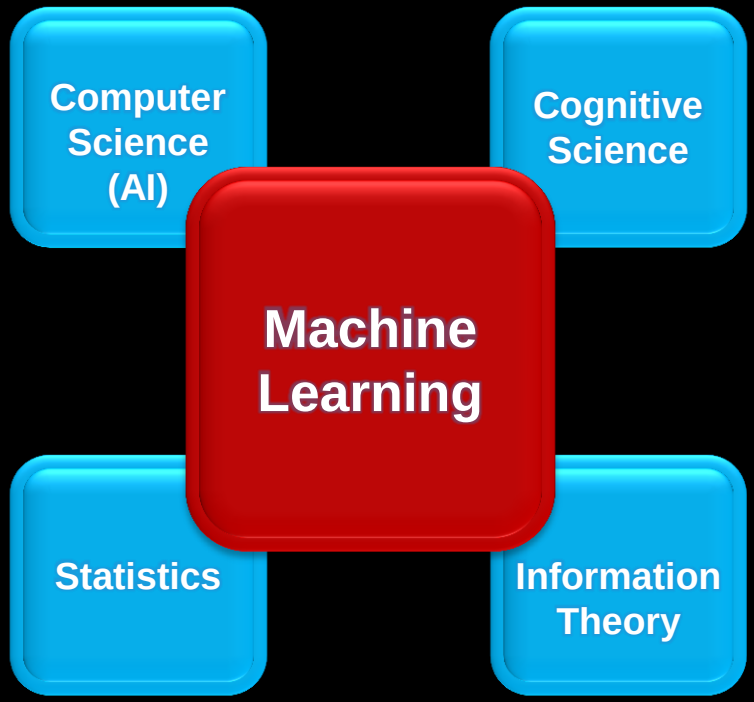
  - Improves its performance **P**
  - At task **T**
  - Through experience **E**

  "The Discipline of Machine Learning" T. Mitchell (2006)

- Machine Learning algorithms discover the relationships between the variables of a system (input, output and hidden) from direct samples of the system

Machine Learning
Building Blocks

- **Computer Science**
  - How can we build machines that solve problems, and which problems are inherently tractable/intractable?
- **Statistics**
  - What can be inferred from data plus a set of modeling assumptions, with what reliability?
- **Cognitive Science**
  - How does the mind process information in faculties such as perception, language, memory, reasoning and emotion?
- **Information Theory**
  - How can we quantify, process, store and communicate data efficiently?

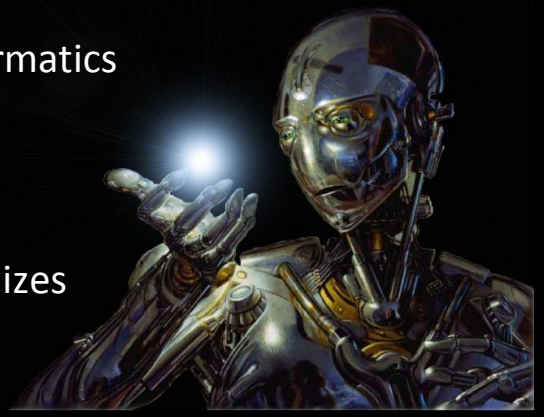


ML builds on all these questions but is a distinct question

Machine Learning

Categories of Machine Learning

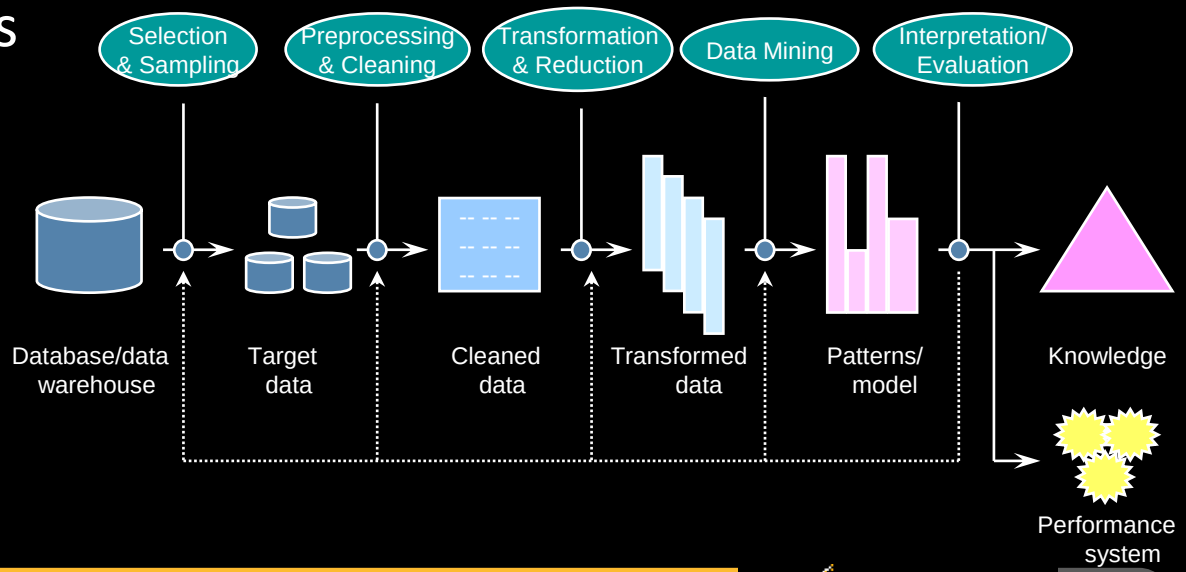
- **Supervised Learning**
  - Given example of inputs and corresponding desired outputs, predict outputs on future inputs
    - Given input output pairs $<x_i, y_i>$, learn a function $f(x_i) = y_i$ for all i that makes a good guess at y for unseen x
    - **Labeled Data***
  - Example: Classification, Regression
- **Unsupervised Learning**
  - Given only inputs, automatically discover representations, features, structure, etc.
    - **Unlabeled Data***
  - Example: Clustering, Outlier detection
- **Semi Supervised Learning**
  - Learning from a combination of **labeled** and **unlabeled data**
  - Example: supervised learning problems like video indexing, bioinformatics
    - Applied where there is **less labeled data and abundance of unlabeled data** *
- **Reinforcement Learning**
  - Given sequence of inputs, actions from a fixed set, and scalar rewards/punishments, learn to select action sequences that maximizes expected reward
  - Example: Robotics



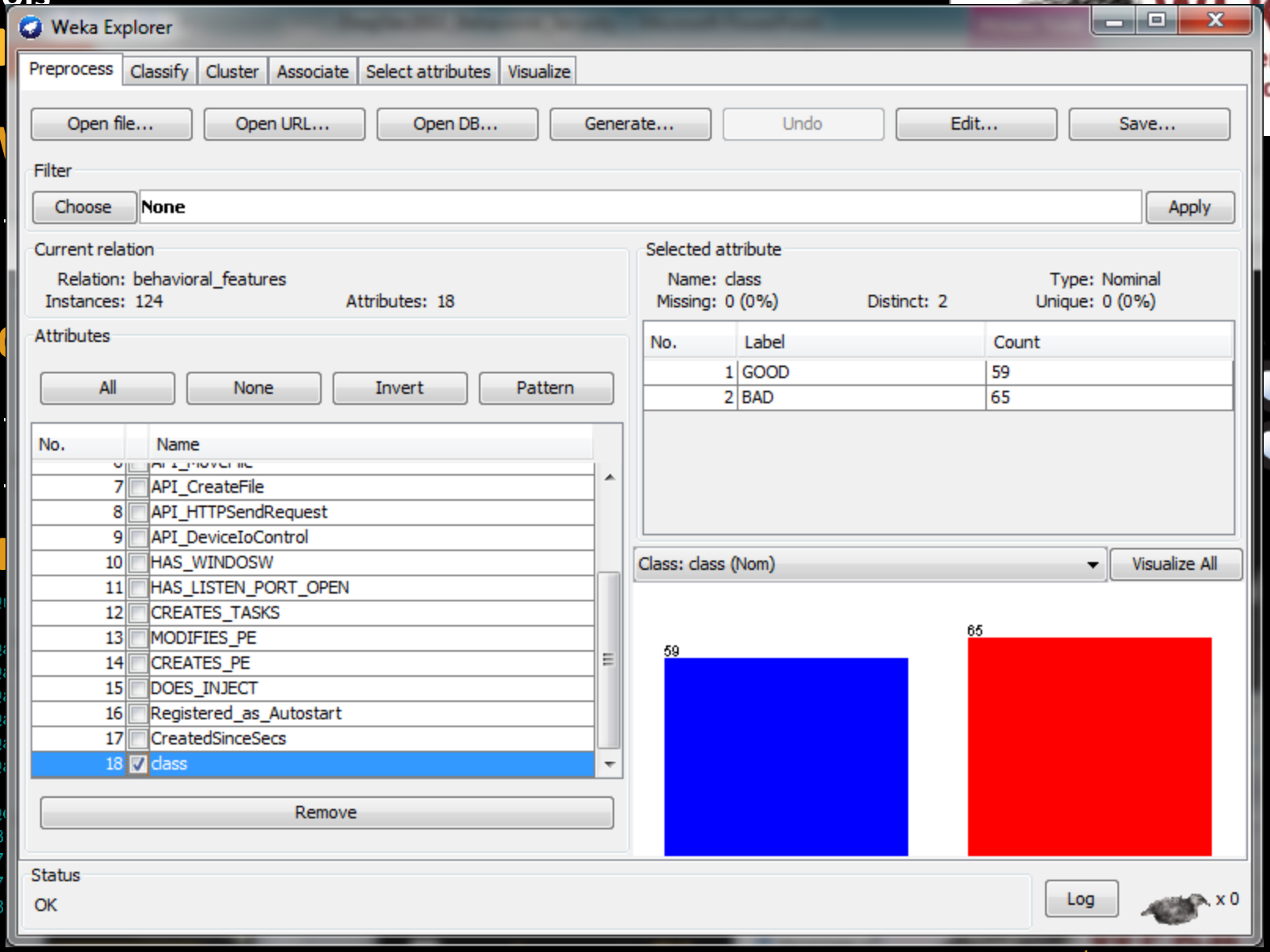
Machine Learning
Steps

1) Pick a feature representation for your task
   - Inputs and Outputs, Feature identification (power to discriminate)
2) Compile data
3) Choose a machine learning algorithm
4) Train the algorithm
5) Evaluate the results

Probably: go to (1)



Tools


Weka 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: behavioral_features
Instances: 124
Attributes: 18
Attributes
All
None
Invert
Pattern
No.
Name
7 API_CreateFile
8 API_HTTPSendRequest
9 API_DeviceIoControl
10 HAS_WINDOSW
11 HAS_LISTEN_PORT_OPEN
12 CREATES_TASKS
13 MODIFIES_PE
14 CREATES_PE
15 DOES_INJECT
16 Registered_as_Autostart
17 CreatedSinceSecs
18 class
Remove
Selected attribute
Name: class
Type: Nominal
Missing: 0 (0%)
Distinct: 2
Unique: 0 (0%)
No.
Label
Count
1 GOOD 59
2 BAD 65
Class: class (Nom)
Visualize All
59
65
Status
OK
Log
x 0

Machine Learning for behavioral security

Machine Learning for behavioral security

Overview

- **Goal**
  - Train a model to provide automated meaningful information about unknown samples
    - **Identify class/label** (Supervised Machine Learning) → **Classification**
    - **Identify association** (Unsupervised Machine Learning) → **Clustering**

- **Application of information extracted**
  - Classify the sample or provide information to analysts for labeling and writing definitions for detection
  - Real time protection

Machine Learning for behavioral security

Overall process

- **Steps**
  - Collect samples
  - Setup a VM with *monitoring framework*
  - Push and run samples in a farm of virtual machines
  - Collect sample behavior data
  - Recycle the VM
  - Extract information into format suitable for data mining
  - Train the models
  - Test and deploy the models

Supervised Machine Learning
For real-time protection

- **Monitoring framework**
  - Data Collection
  - User mode hooking API: *Detours (Microsoft)*
    - Hook the APIs
    - Collect the data in the context of the API Hook
      - API Info(Name, Parameters), Called-from API, State of the process, etc.
      - Log the information
    - Extract features: Logs → ARFF files
      - API Called
      - Has UI/Window
      - Does Network Communication
        - IRC
        - HTTP
      - Registered in AutoStart locations
      - Creates Windows Tasks (jobs)
      - Modifies PE Files
      - Creates PE Files
      - Injects into Trusted Processes

G:\DeepSecDemo\behavioral.arff - Notepad++

File Edit Search View Encoding Language Settings Macro Run Plugins Window ?

behavioral.arff

```
@RELATION behavioral_security

@ATTRIBUTE API_CreateWindow {T,F}
@ATTRIBUTE API_ShowWindow   {T,F}
@ATTRIBUTE API_FindWindow   {T,F}
@ATTRIBUTE API_ShellExecute {T,F}
@ATTRIBUTE API_DeleteFile   {T,F}
@ATTRIBUTE API_MoveFile {T,F}
@ATTRIBUTE API_CreateFile   {T,F}
@ATTRIBUTE API_HTTPSendRequest  {T,F}
@ATTRIBUTE API_DeviceIoControl  {T,F}
@ATTRIBUTE HAS_WINDOSW  {T,F}
@ATTRIBUTE HAS_LISTEN_PORT_OPEN     {T,F}
@ATTRIBUTE CREATES_TASKS    {T,F}
@ATTRIBUTE MODIFIES_PE  {T,F}
@ATTRIBUTE CREATES_PE   {T,F}
@ATTRIBUTE DOES_INJECT  {T,F}
@ATTRIBUTE Registered_as_Autostart  {T,F}
@ATTRIBUTE CreatedSinceSecs real
@ATTRIBUTE class    {GOOD,BAD}

@DATA
T,F,T,T,T,T,F,F,T,T,T,T,T,T,F,F,734,GOOD
T,F,T,T,F,T,T,T,F,T,F,T,F,T,F,F,22,BAD
F,F,T,T,F,T,F,T,F,T,T,T,T,T,F,F,34,BAD
F,F,T,T,F,F,T,T,T,F,T,F,F,T,F,566,BAD
F,F,T,F,F,F,F,T,F,F,T,T,F,T,T,F,2,BAD
T,F,T,T,T,T,F,F,T,T,T,T,T,T,F,F,629,GOOD
F,F,T,F,T,F,T,T,F,T,T,T,T,F,T,F,6587,BAD
F,F,F,F,T,F,T,T,T,F,T,F,T,T,F,F,232,BAD
F,F,F,F,T,F,T,F,T,T,F,T,T,T,F,F,1864,BAD
F,F,F,F,T,T,F,T,T,T,F,F,T,T,F,F,654,BAD
T,F,T,T,F,F,F,F,T,T,F,F,F,T,F,F,45,BAD
T,F,T,F,F,F,F,F,F,T,F,F,F,T,T,T,95,BAD
T,F,T,F,F,F,F,T,T,F,F,F,F,T,T,3,BAD
T,F,T,T,F,F,T,F,F,F,T,F,T,T,F,T,51,BAD
T,F,T,T,T,T,T,F,F,T,T,T,T,T,F,F,528,GOOD
T,F,T,T,T,T,T,F,F,T,T,T,T,T,F,F,734,GOOD
T,F,F,T,F,F,F,F,T,F,T,T,F,F,F,T,65,BAD
T,F,T,T,T,F,F,F,T,F,T,F,F,T,F,T,57,BAD
F,F,F,T,F,F,T,F,T,F,T,F,F,T,F,F,2354,GOOD
F,F,T,T,F,F,T,F,F,T,F,T,T,F,T,654,BAD
F,F,T,T,T,F,T,T,F,T,T,T,F,T,F,6,GOOD
F,F,T,F,T,F,F,F,T,T,F,T,F,T,F,F,9,GOOD
F,F,F,T,T,F,F,F,T,T,T,F,F,T,T,F,645,GOOD
T,F,F,F,T,F,F,F,T,F,T,F,F,T,T,F,534,GOOD
F,F,F,F,F,F,F,F,T,F,T,F,F,T,F,F,56,BAD
F,F,F,T,T,F,F,F,T,F,T,F,F,T,F,F,89,GOOD
T,F,T,F,F,F,F,T,T,F,F,F,F,F,F,F,15,BAD
T,F,F,F,T,F,F,T,T,F,F,F,F,F,F,68,BAD
T,F,F,F,T,F,T,T,T,T,F,F,F,F,F,29,BAD
T,F,T,F,T,F,T,T,T,F,F,T,F,T,T,T,F,37,GOOD
F,F,T,F,F,F,T,T,T,F,F,T,F,F,F,F,234,GOOD
T,F,T,F,T,F,T,T,T,T,F,F,F,F,F,24178,GOOD
F,F,T,F,T,F,T,F,T,T,T,T,F,F,F,F,4705,GOOD
T,F,T,T,T,F,T,F,T,T,F,F,T,F,F,47893,GOOD
T,F,T,T,T,F,F,F,T,T,F,T,F,T,T,F,5832,GOOD
```

Normal text file | length : 5746 | lines : 150 | Ln : 1 | Col : 30 | Sel : 0 | Dos\Windows | ANSI | INS

Example

Data to Models

...DEMO...

...click here if demo GODs act up!..

Lab to field

Apply Classifiers

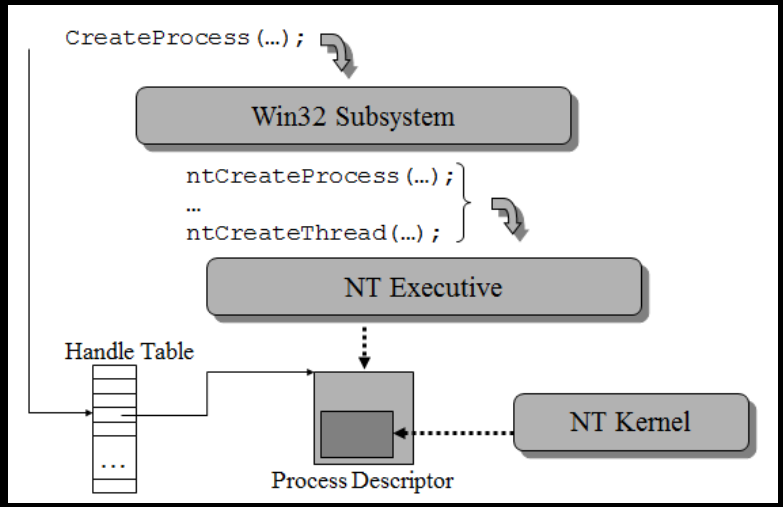
- **Monitoring and Blocking hook points**
  - May or may not be the same
    - Some hooks points are merely for state/information collection
  - Work done in API Hooks
    - Collect information
    - Transform information into feature vector
    - Evaluate against model
    - Allow or Deny

Lab to field

Apply Classifiers

- **Which APIs to hook?**
  - Higher level API (CreateProcess @ kernel32.dll)
  - Lower level API (NtCreateProcess? NtCreateThread?, Ldrpxxx?)
  - Higher level APIs (exports by kernel32) provide fine grain control
  - Many high level APIs map to few lower level APIs (functionally)
  - Lower level APIs provide a more comprehensive view

- **Block Action:**
  - Failing an API
    - Out parameter
    - Return code
  - Terminate Thread/process

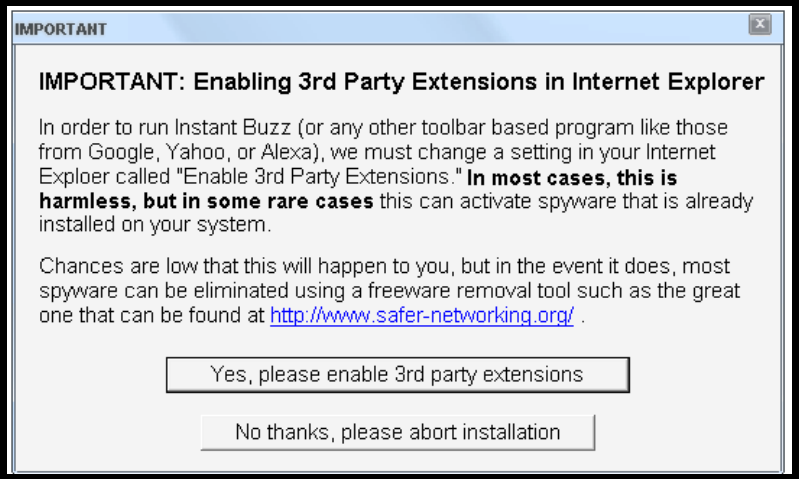


Machine Learning for behavioral security

Reality check…

Automation
Reality check

- **Practical Challenges**
  - Samples fail to run in automation
    - Good Samples fail to run in automation
      - ***more commonly than Malicious samples***
      - Dependency, Configuration, etc.
      - GUI automation
    - Malicious Samples deliberately fail to run in automation
      - VM Aware
      - Automation Aware
        - Check own file name (example: sysdat
        - Check parent process (Threat: Trojan.
        - Check application settings (Threat: Ad
        - Check commonly used applications (MS Office)
    - Samples may be stale: C & C Down
  - System state sensitivity
    - Valid Samples: Missing depencies like Java, .NET, etc.
    - Malicious Samples: Missing targeted applications like Ado




```
mov eax, 564d5868h  ;'VMXh'
mov ecx, 0ah        ;get VMware version
mov dx, 5658h       ;'VX'
in  eax, dx
cmp ebx, 564d5868h  ;'VMXh'
je  detected
```

VMWare

```
0F 3F x1 x2
0F C7 C8 y1 y2
```

VirtualPC

```
xor     eax, eax
cpuid
cmp     ecx, 444d4163h
jne     exit
mov     eax, 80000000h
cpuid
cmp     eax, 2
jb      exit
mov     eax, 80000002h
cpuid
cmp     eax, 554d4551h
je      $ ;detected
```

QEMU

Machine Learning
Reality check

- **Machine Learning Challenges**
  - Imbalanced data sets
  - Missing features
  - Anomalous feature values
    - outlier or deliberately manufactured?
    - Some tricks observed in malware*:
      - Non-standard ImageBase
      - Large values in .DATA/SizeofRawdata
      - Bogus values in LoaderFlags

- **Section Modification: Or how to kill many tools.**

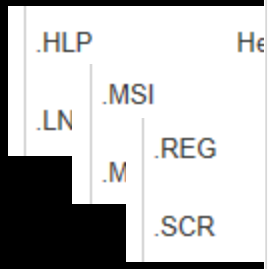
```
->Section Header Table
   1. item:
    Name:                  CODE
    VirtualSize:           0x00001000
    VirtualAddress:        0x00001000
    SizeOfRawData:         0x00001000
    PointerToRawData:      0x00001000
    PointerToRelocations:  0x00000000
    PointerToLinenumbers:  0x00000000
    NumberOfRelocations:   0x0000
    NumberOfLinenumbers:   0x0000
    Characteristics:       0xE0000020
    (CODE, EXECUTE, READ, WRITE)

   2. item:
    Name:                  DATA
    VirtualSize:           0x00045000
    VirtualAddress:        0x00002000
    SizeOfRawData:         0x00045000
    PointerToRawData:      0x00002000
    PointerToRelocations:  0x00000000
    PointerToLinenumbers:  0x00000000
    NumberOfRelocations:   0x0000
    NumberOfLinenumbers:   0x0000
    Characteristics:       0xC0000040
    (INITIALIZED_DATA, READ, WRITE)

   3. item:
    Name:                  NicolasB
    VirtualSize:           0x00001000
    VirtualAddress:        0x00047000
    SizeOfRawData:         0xEFEFADFF   <--- BIG Size of section on the disk.
    PointerToRawData:      0x00047000
    PointerToRelocations:  0x00000000
    PointerToLinenumbers:  0x00000000
    NumberOfRelocations:   0x0000
    NumberOfLinenumbers:   0x0000
    Characteristics:       0xC0000040
    (INITIALIZED_DATA, READ, WRITE)
```

Scan of the Month 33: Anti Reverse Engineering Uncovered By Nicolas Brulez - 0x90(at)Rstack(dot)org

Stealthy Malware

Malicious Payloads

- **NPTs (Non Process Threat)**
  - Trusted process -> Malicious Behavior
  - File vs. Process
  - Code vs. Data
    - Malicious PDF → Browser or Adobe reader
    - Malicious JAR files → Browser or java.exe
    - Malicious MSI files → msiexec.exe
  - DLLs
    - Regsvr32
    - Rundll32
    - Svchost.exe
    - IE/Explorer Extensions



How to automate these?

How/where is protection enforced?

What is remediated?

Infectable Object	Description
.ASP	Active-X components of Active server pages
.BAT	DOS Batch files
.CSH	C Shell Script
.CHM	MS Compiled HTML Help
.CLA(SS)	JAVA file
.CSC	Corel Script
.CSS	Cascading Style Sheet
.HT?	HTML variant
.HTM	HTML variant
.HTA	HTML variant
.HTML	HTML variant
.HTT	Hypertext Template
.INI	mIRC - SCRIPT.INI
.INI	pIRCH - EVENTS.INI
.JS	JavaScript source
.JSE	JavaScript Encoded Script File
.MHT	HTML code
.MHTML	HTML code
.SCT	Windows Script Component
.sh	Bourne shell or Korn Script
.SHB	Shell Scrap object
.SHTML	HTML file
.VB	VBScript File
.VBS	Visual Basic script file
.VBE	Visual Basic encoded script
.VBX	Visual Basic Extension
.WBT	Windows Batch file
.WSC	Windows Script Component
.WSF	Windows Script File

Conclusion & Food for thought!

Recap

Scaling to the Malware population

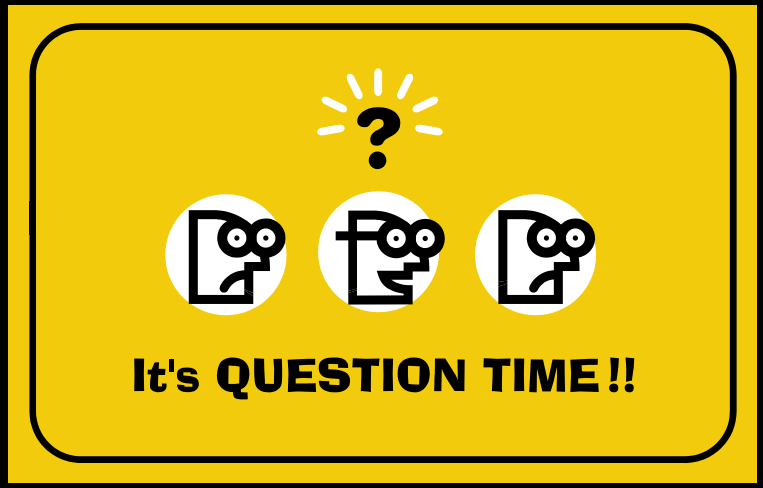
- **Volume of malware by unique file fingerprint != New Malware**
  - Behaviorally malware is not evolving at every instance
  - Scalability can be handled with Automation
  - Be aware of pitfalls of automation
  - Automation + domain knowledge
  - Use domain experts effectively
- **Challenge**
  - What if the Malware is a valid application with configuration file?

  - Solution: Opportunity for Creative Feature Engineering?

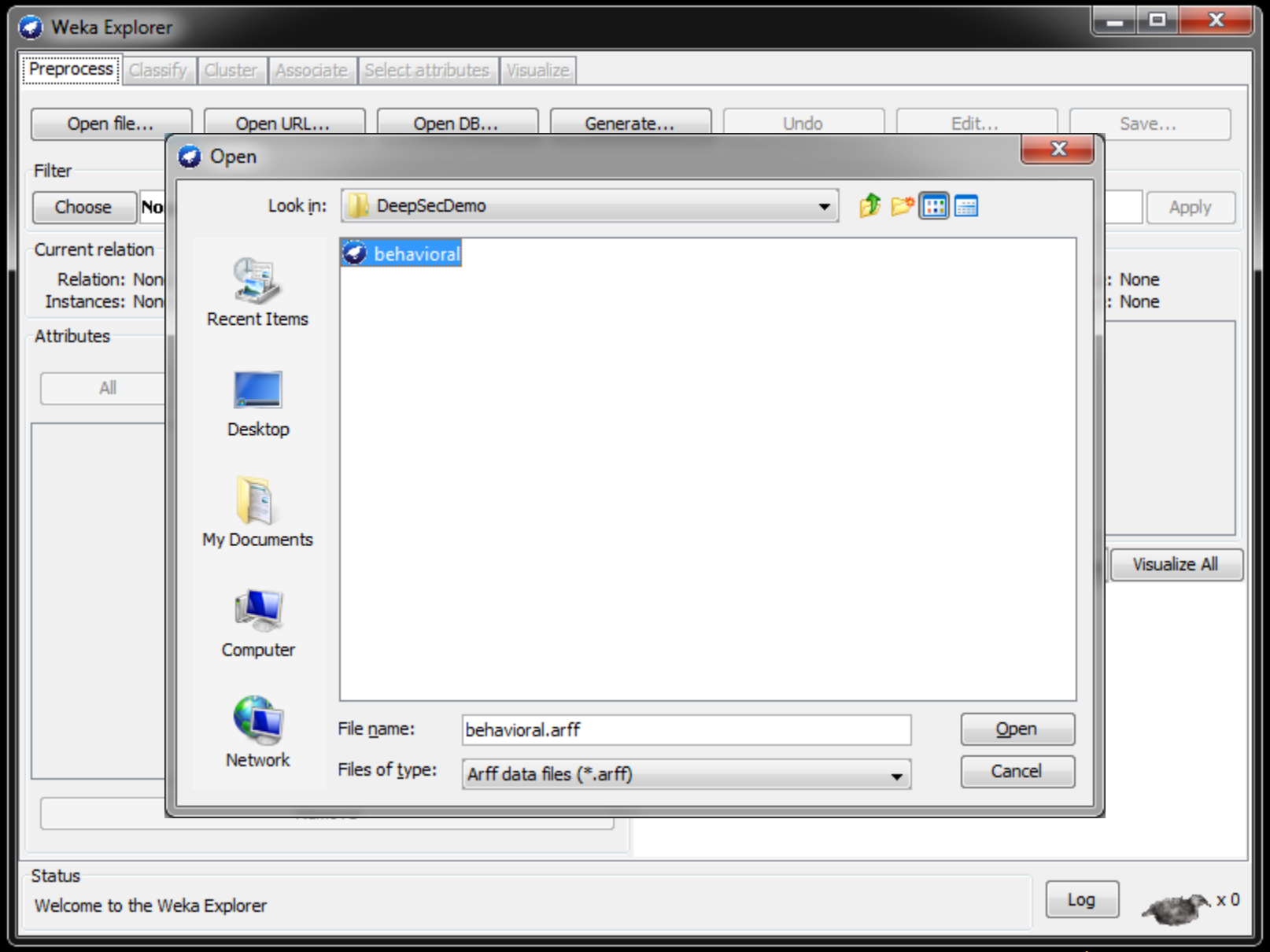
Thank You!

Sourabh Satish

sourabh_satish@symantec.com




Weka GUI Chooser
Program Visualization Tools Help
WEKA
The University of Waikato
Applications
Explorer
Experimenter
KnowledgeFlow
Simple CLI
Waikato Environment for Knowledge Analysis
Version 3.6.6
(c) 1999 - 2011
The University of Waikato
Hamilton, New Zealand


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
Apply
Current relation
Relation: Non
Instances: Non
Attributes
All
: None
: None
Visualize All
Open
Look in:
DeepSecDemo
behavioral
Recent Items
Desktop
My Documents
Computer
Network
File name:
behavioral.arff
Open
Files of type:
Arff data files (*.arff)
Cancel
Status
Welcome to the Weka Explorer
Log
x 0

Weka 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: behavioral_features
Instances: 124 Attributes: 18

Attributes

All | None | Invert | Pattern

No.	Name
1	API_CreateWindow
2	API_ShowWindow
3	API_FindWindow
4	API_ShellExecute
5	API_DeleteFile
6	API_MoveFile
7	API_CreateFile
8	API_HTTPSendRequest
9	API_DeviceIoControl
10	HAS_WINDOSW
11	HAS_LISTEN_PORT_OPEN
12	CREATES_TASKS
13	MODIFIES_PE
14	CREATES_PE
15	DOES_INJECT
16	Registered_as_Autostart
17	CreatedSinceSecs
18	class

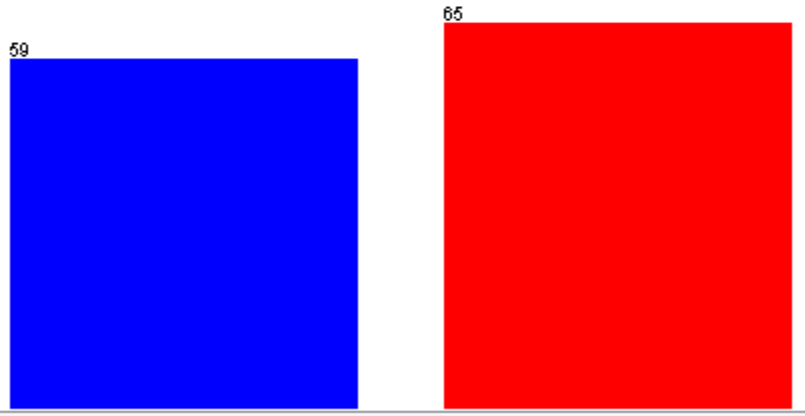
Remove

Selected attribute

Name: class Type: Nominal
Missing: 0 (0%) Distinct: 2 Unique: 0 (0%)

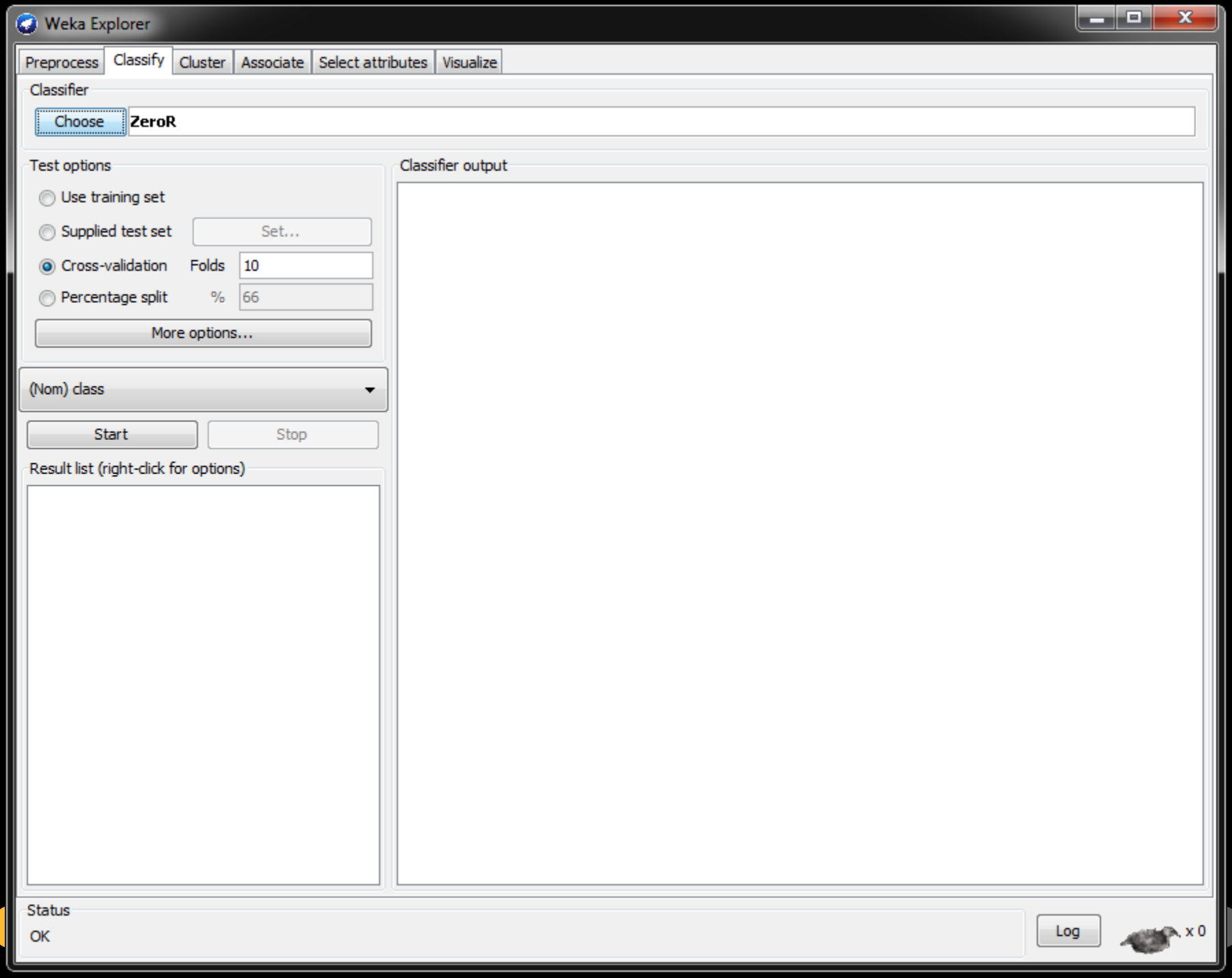
No.	Label	Count
1	GOOD	59
2	BAD	65

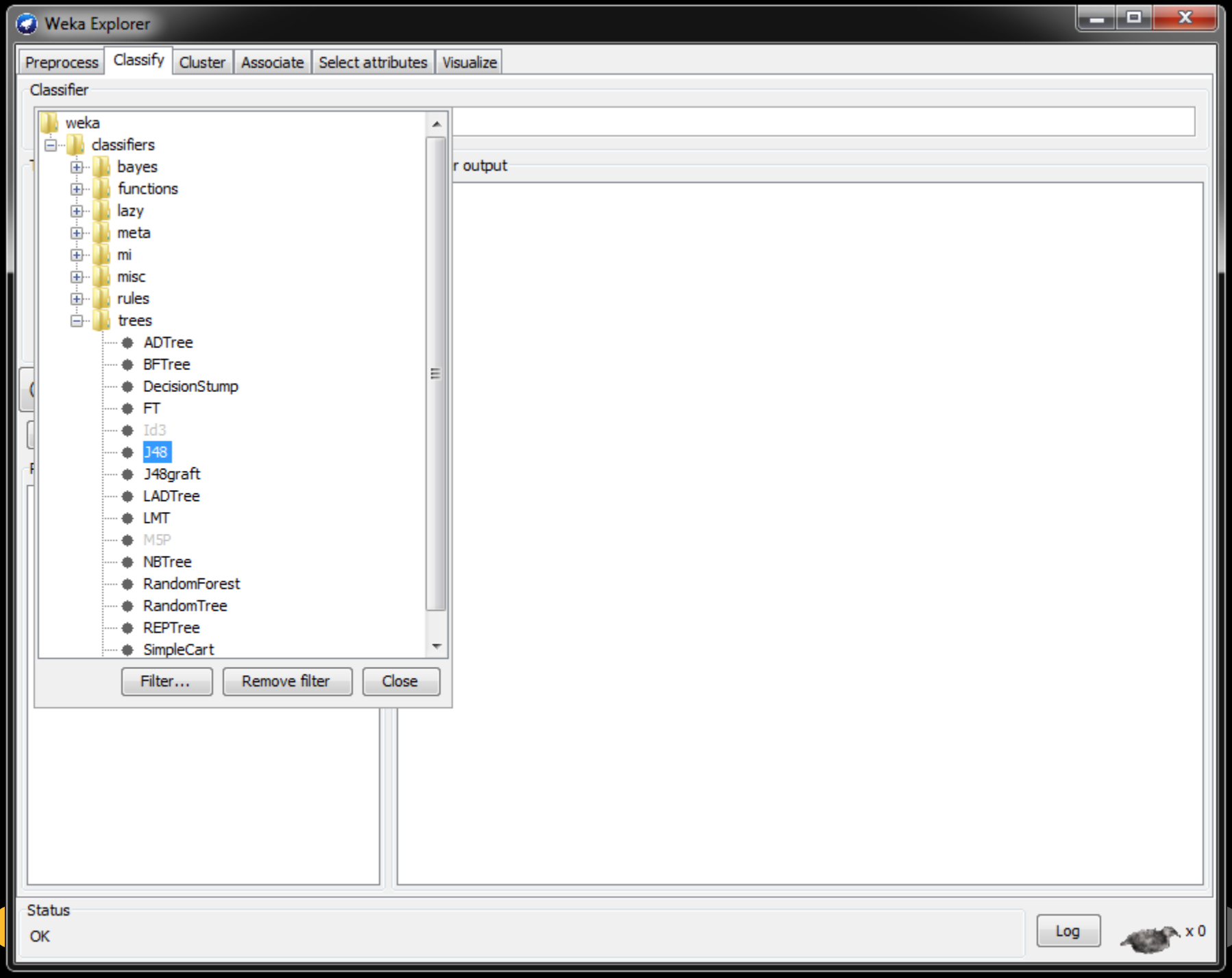
Class: class (Nom) Visualize All

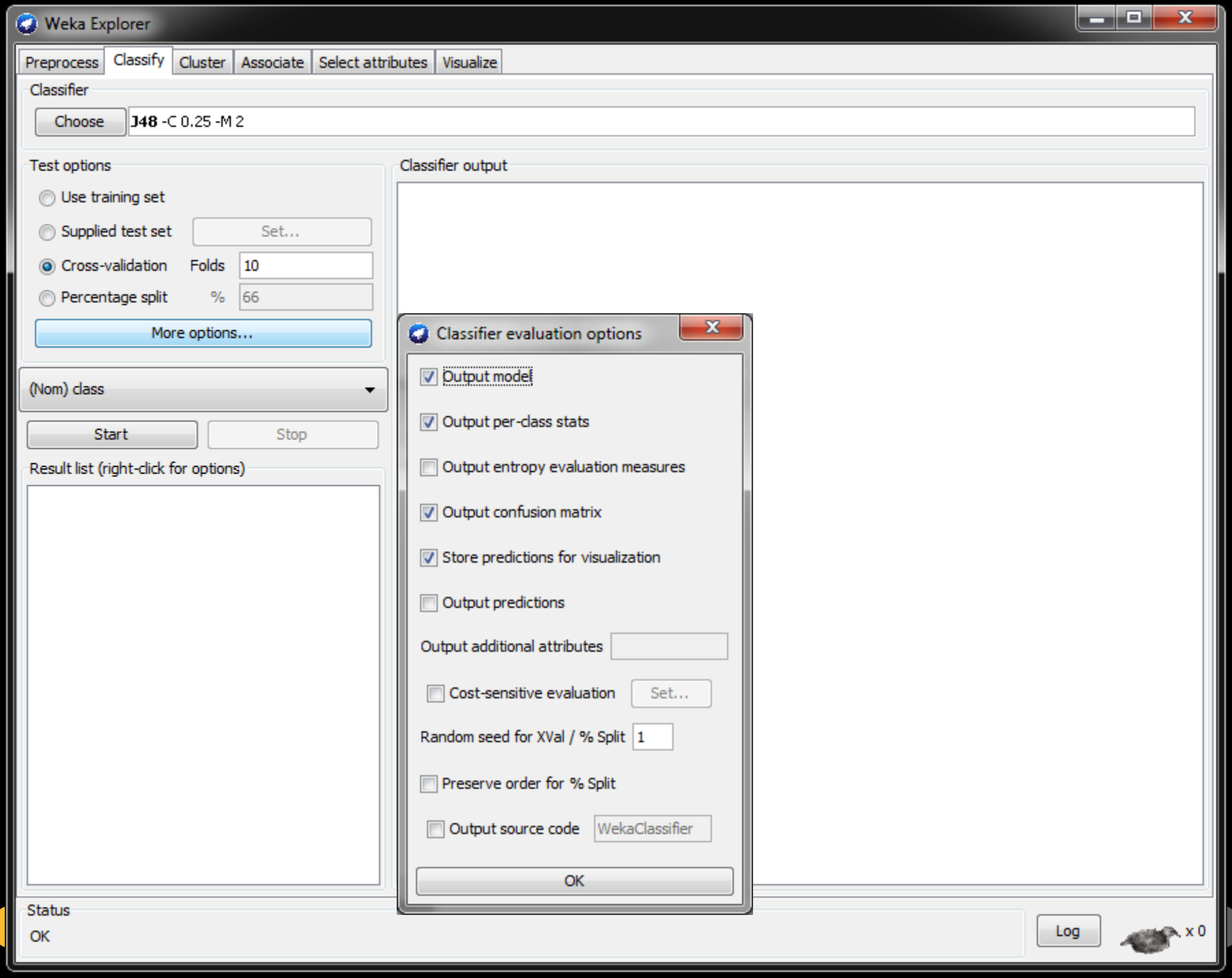


Status

OK Log x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
ZeroR
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
Classifier output
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
weka
classifiers
bayes
functions
lazy
meta
mi
misc
rules
trees
ADTree
BFTree
DecisionStump
FT
Id3
J48
J48graft
LADTree
LMT
M5P
NBTree
RandomForest
RandomTree
REPTree
SimpleCart
Filter...
Remove filter
Close
r output
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click for options)
Classifier evaluation options
Output model
Output per-class stats
Output entropy evaluation measures
Output confusion matrix
Store predictions for visualization
Output predictions
Output additional attributes
Cost-sensitive evaluation
Set...
Random seed for XVal / % Split
1
Preserve order for % Split
Output source code
WekaClassifier
OK
Status
OK
Log
x 0

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose | J48 -C 0.25 -M 2

Test options

- Use training set
- Supplied test set — Set...
- Cross-validation — Folds 10
- Percentage split — % 66

More options...

(Nom) class

Start | Stop

Result list (right-click for options)

12:32:59 - trees.J48

Classifier output

```
=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     behavioral_features
Instances:    124
Attributes:   18
              API_CreateWindow
              API_ShowWindow
              API_FindWindow
              API_ShellExecute
              API_DeleteFile
              API_MoveFile
              API_CreateFile
              API_HTTPSendRequest
              API_DeviceIoControl
              HAS_WINDOSW
              HAS_LISTEN_PORT_OPEN
              CREATES_TASKS
              MODIFIES_PE
              CREATES_PE
              DOES_INJECT
              Registered_as_Autostart
              CreatedSinceSecs
              class
Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
------------------
```

Status

OK

Log | x 0

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose **J48** -C 0.25 -M 2

Test options

- Use training set
- Supplied test set — Set...
- Cross-validation — Folds 10
- Percentage split — % 66

More options...

(Nom) class

Start | Stop

Result list (right-click for options)

12:32:59 - trees.J48

Classifier output

```
J48 pruned tree
------------------

CreatedSinceSecs <= 6587
|   CREATES_TASKS = T
|   |   CreatedSinceSecs <= 9: BAD (18.0/3.0)
|   |   CreatedSinceSecs > 9
|   |   |   API_ShellExecute = T
|   |   |   |   API_DeleteFile = T: GOOD (9.0)
|   |   |   |   API_DeleteFile = F: BAD (12.0/2.0)
|   |   |   API_ShellExecute = F
|   |   |   |   API_DeleteFile = T
|   |   |   |   |   API_MoveFile = T: GOOD (2.0)
|   |   |   |   |   API_MoveFile = F
|   |   |   |   |   |   MODIFIES_PE = T: BAD (3.0)
|   |   |   |   |   |   MODIFIES_PE = F: GOOD (2.0)
|   |   |   |   API_DeleteFile = F: GOOD (13.0)
|   CREATES_TASKS = F
|   |   API_CreateFile = T
|   |   |   API_MoveFile = T: GOOD (8.0/1.0)
|   |   |   API_MoveFile = F: BAD (10.0/2.0)
|   |   API_CreateFile = F
|   |   |   CreatedSinceSecs <= 654: BAD (32.0/4.0)
|   |   |   CreatedSinceSecs > 654: GOOD (2.0)
CreatedSinceSecs > 6587: GOOD (13.0)

Number of Leaves  :     12

Size of the tree :      23
```

Status

OK

Log

x 0

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose: J48 -C 0.25 -M 2

Test options

- Use training set
- Supplied test set — Set...
- Cross-validation — Folds: 10
- Percentage split — %: 66

More options...

(Nom) class

Start | Stop

Result list (right-click for options)

12:32:59 - trees.J48

Classifier output

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         101               81.4516 %
Incorrectly Classified Instances        23               18.5484 %
Kappa statistic                          0.6249
Mean absolute error                      0.2555
Root mean squared error                  0.4142
Relative absolute error                 51.1945 %
Root relative squared error             82.9103 %
Total Number of Instances              124

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.712     0.092      0.875     0.712     0.785      0.757    GOOD
                 0.908     0.288      0.776     0.908     0.837      0.757    BAD
Weighted Avg.    0.815     0.195      0.823     0.815     0.812      0.757

=== Confusion Matrix ===

  a  b   <-- classified as
 42 17 |  a = GOOD
  6 59 |  b = BAD
```

Status

OK

Log

x 0

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose **RandomForest** -I 10 -K 0 -S 1

Test options

- Use training set
- Supplied test set — Set...
- Cross-validation — Folds 10
- Percentage split — % 66

More options...

(Nom) class

Start | Stop

Result list (right-click for options)

- 12:32:59 - trees.J48
- 12:39:17 - trees.RandomForest

Classifier output

```
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         102               82.2581 %
Incorrectly Classified Instances        22               17.7419 %
Kappa statistic                          0.6432
Mean absolute error                      0.2519
Root mean squared error                  0.3774
Relative absolute error                 50.48   %
Root relative squared error             75.5492 %
Total Number of Instances              124

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.78      0.138      0.836      0.78      0.807      0.871    GOOD
                 0.862     0.22       0.812      0.862     0.836      0.871    BAD
Weighted Avg.    0.823     0.181      0.823      0.823     0.822      0.871

=== Confusion Matrix ===

  a  b   <-- classified as
 46 13 |  a = GOOD
  9 56 |  b = BAD
```

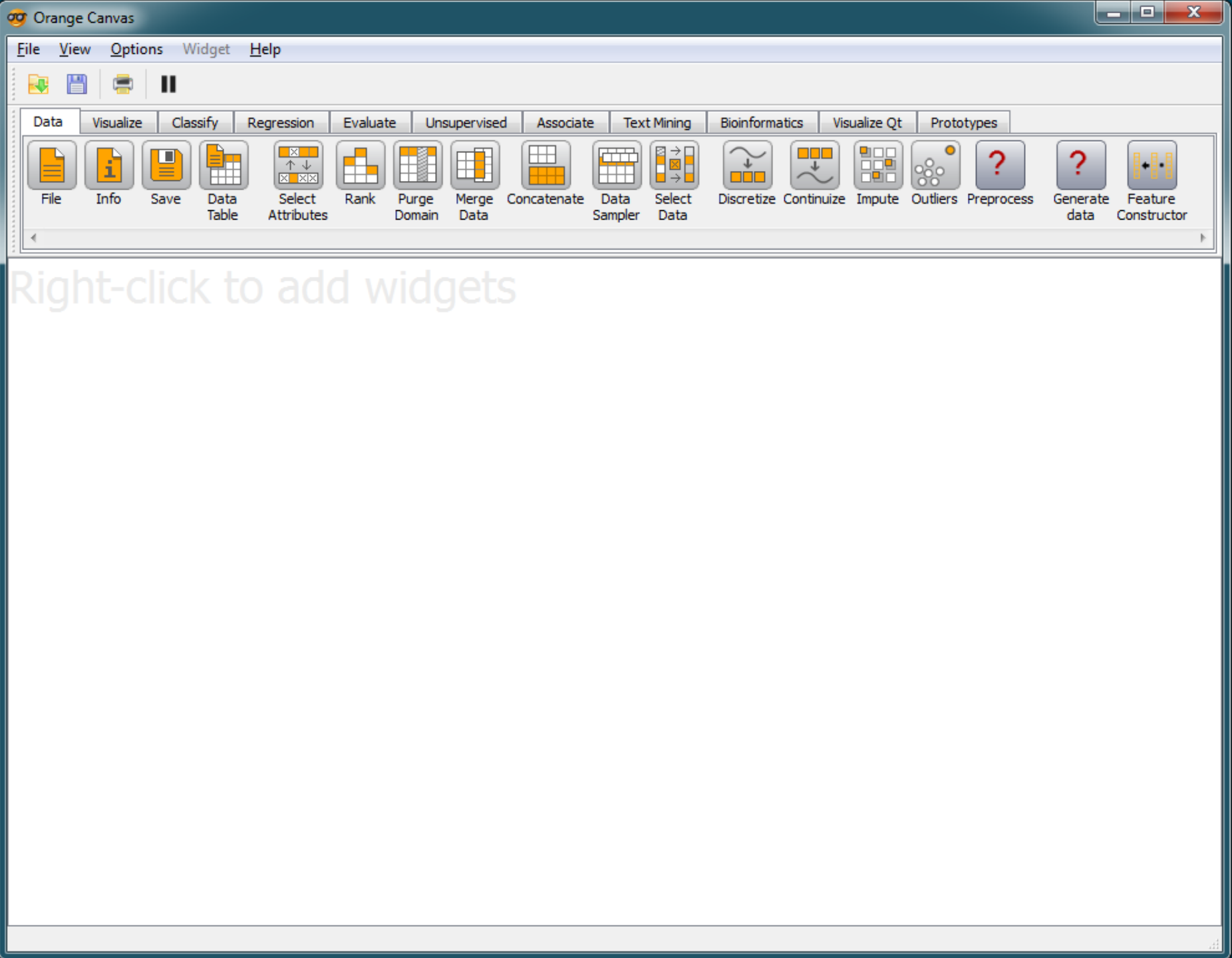
Status

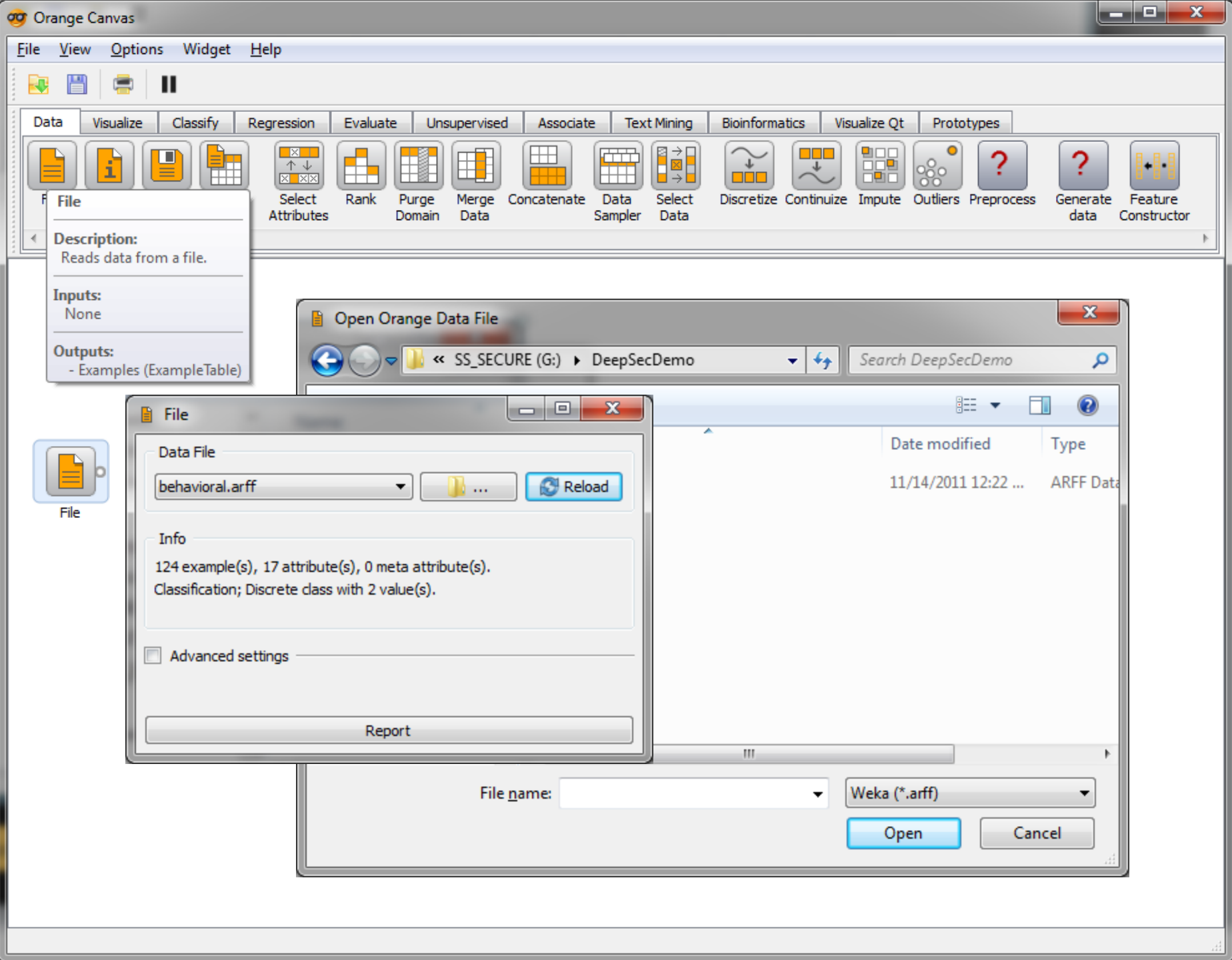
OK

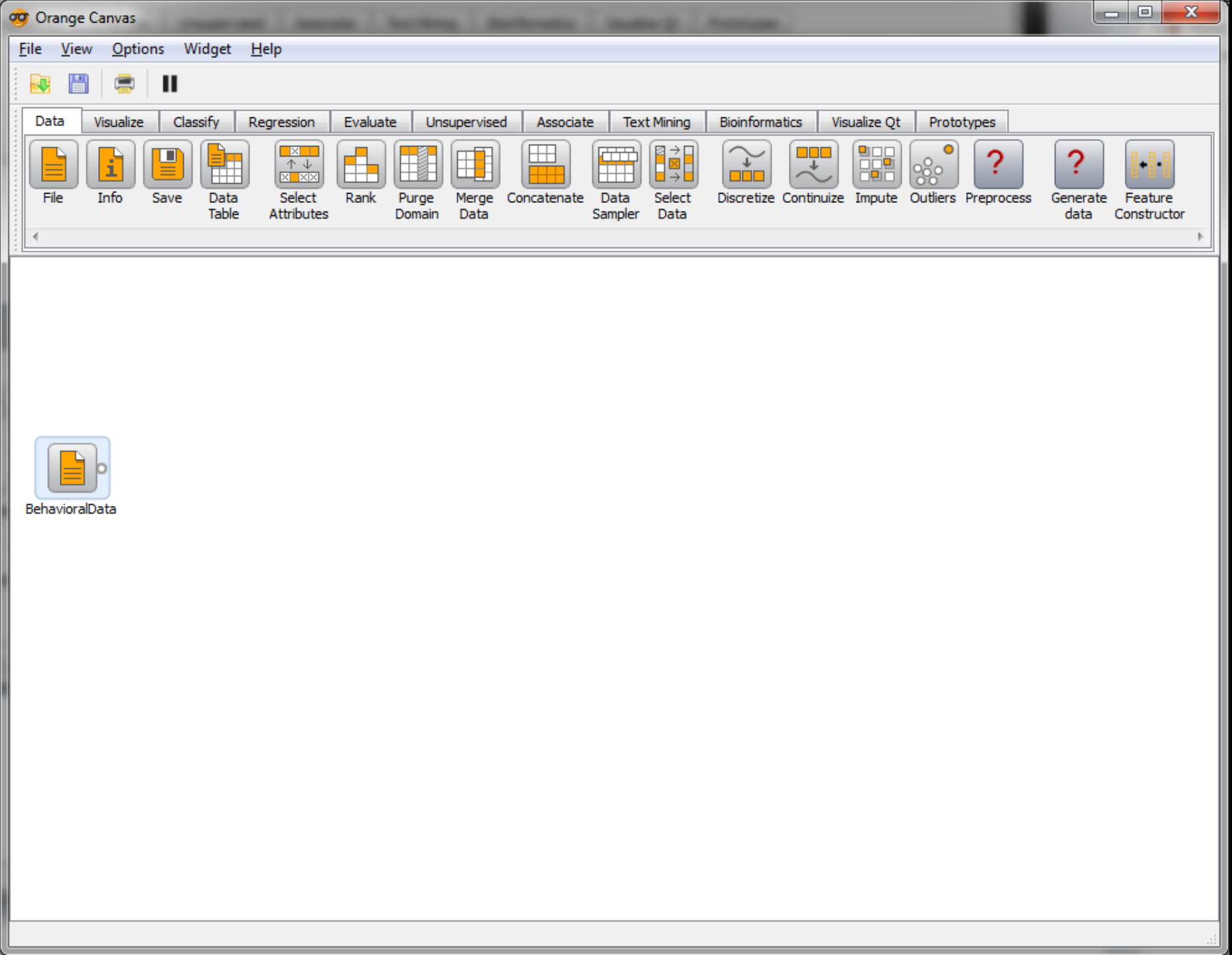
Log

x 0


About Orange
orange
Orange
version 2.0b
(built Jul 26 2011)
Close


Orange Canvas
File View Options Widget Help
Data Visualize Classify Regression Evaluate Unsupervised Associate Text Mining Bioinformatics Visualize Qt Prototypes
File Info Save Data Table Select Attributes Rank Purge Domain Merge Data Concatenate Data Sampler Select Data Discretize Continuize Impute Outliers Preprocess Generate data Feature Constructor
Right-click to add widgets


Orange Canvas
File View Options Widget Help
Data Visualize Classify Regression Evaluate Unsupervised Associate Text Mining Bioinformatics Visualize Qt Prototypes
Select Attributes
Rank
Purge Domain
Merge Data
Concatenate
Data Sampler
Select Data
Discretize
Continuize
Impute
Outliers
Preprocess
Generate data
Feature Constructor
File
Description:
Reads data from a file.
Inputs:
None
Outputs:
- Examples (ExampleTable)
File
Data File
behavioral.arff
...
Reload
Info
124 example(s), 17 attribute(s), 0 meta attribute(s).
Classification; Discrete class with 2 value(s).
Advanced settings
Report
Open Orange Data File
SS_SECURE (G:) ▸ DeepSecDemo
Search DeepSecDemo
Date modified
Type
11/14/2011 12:22 ...
ARFF Data
File name:
Weka (*.arff)
Open
Cancel


Orange Canvas
File View Options Widget Help
Data Visualize Classify Regression Evaluate Unsupervised Associate Text Mining Bioinformatics Visualize Qt Prototypes
File Info Save Data Table Select Attributes Rank Purge Domain Merge Data Concatenate Data Sampler Select Data Discretize Continuize Impute Outliers Preprocess Generate data Feature Constructor
BehavioralData

Orange Canvas

File View Options Widget Help

Data

Distributio

Behavioral

Distributions

Variable

C CreatedSinceSecs

Displayed outcomes

GOOD

BAD

General graph settings

Main title

X axis title CreatedSinceSecs

Y axis title frequency

Show continuous class graph

Number of bars 5

Probability plot

Show probabilities

Target value

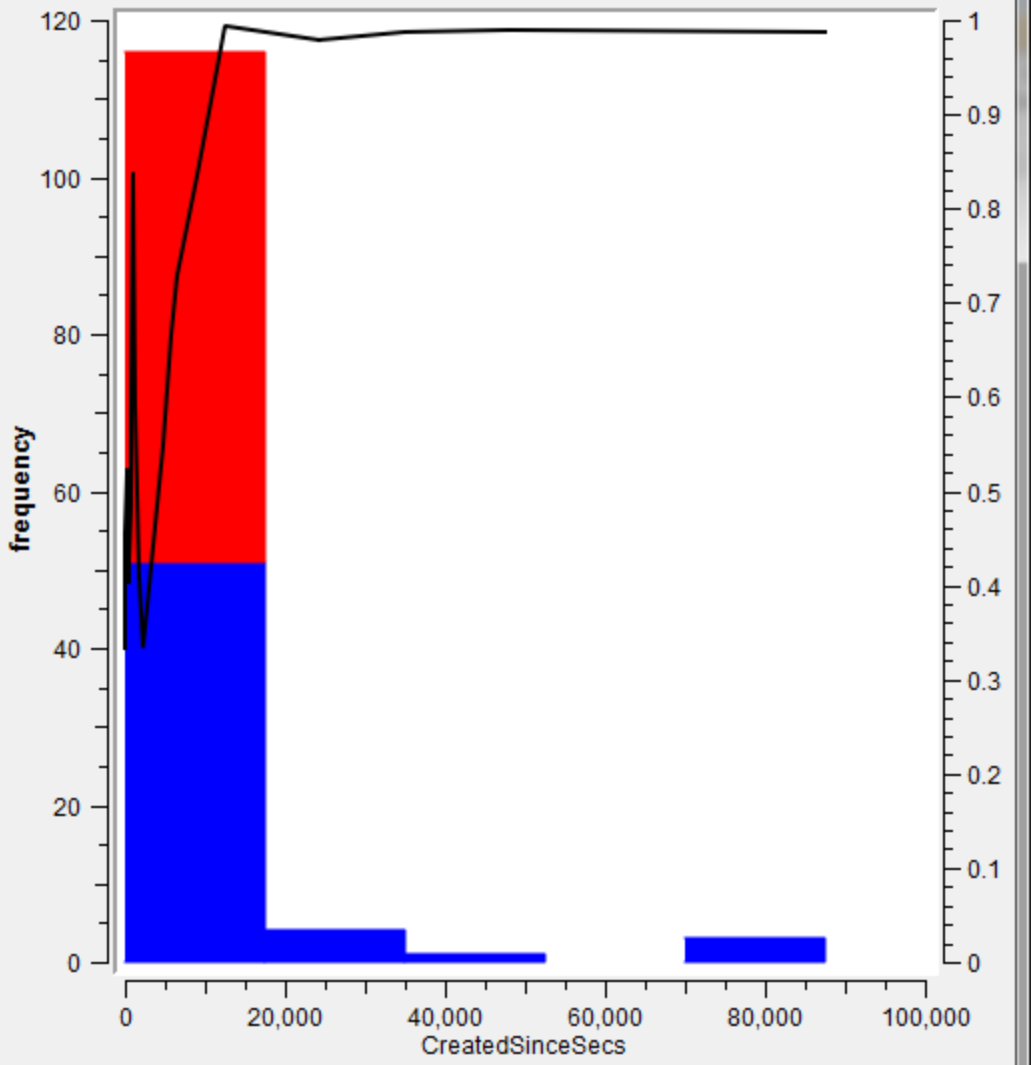
GOOD

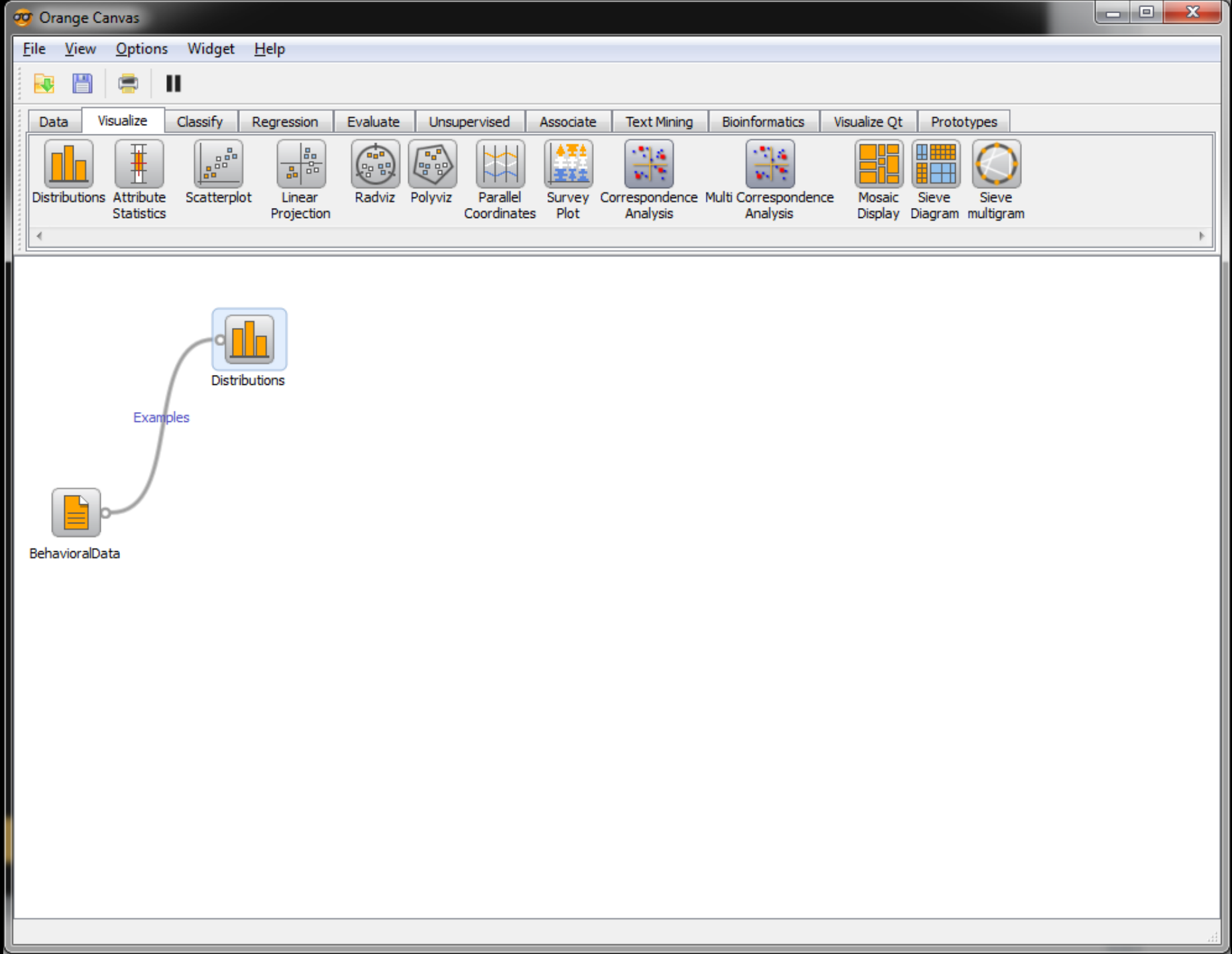
Axis title

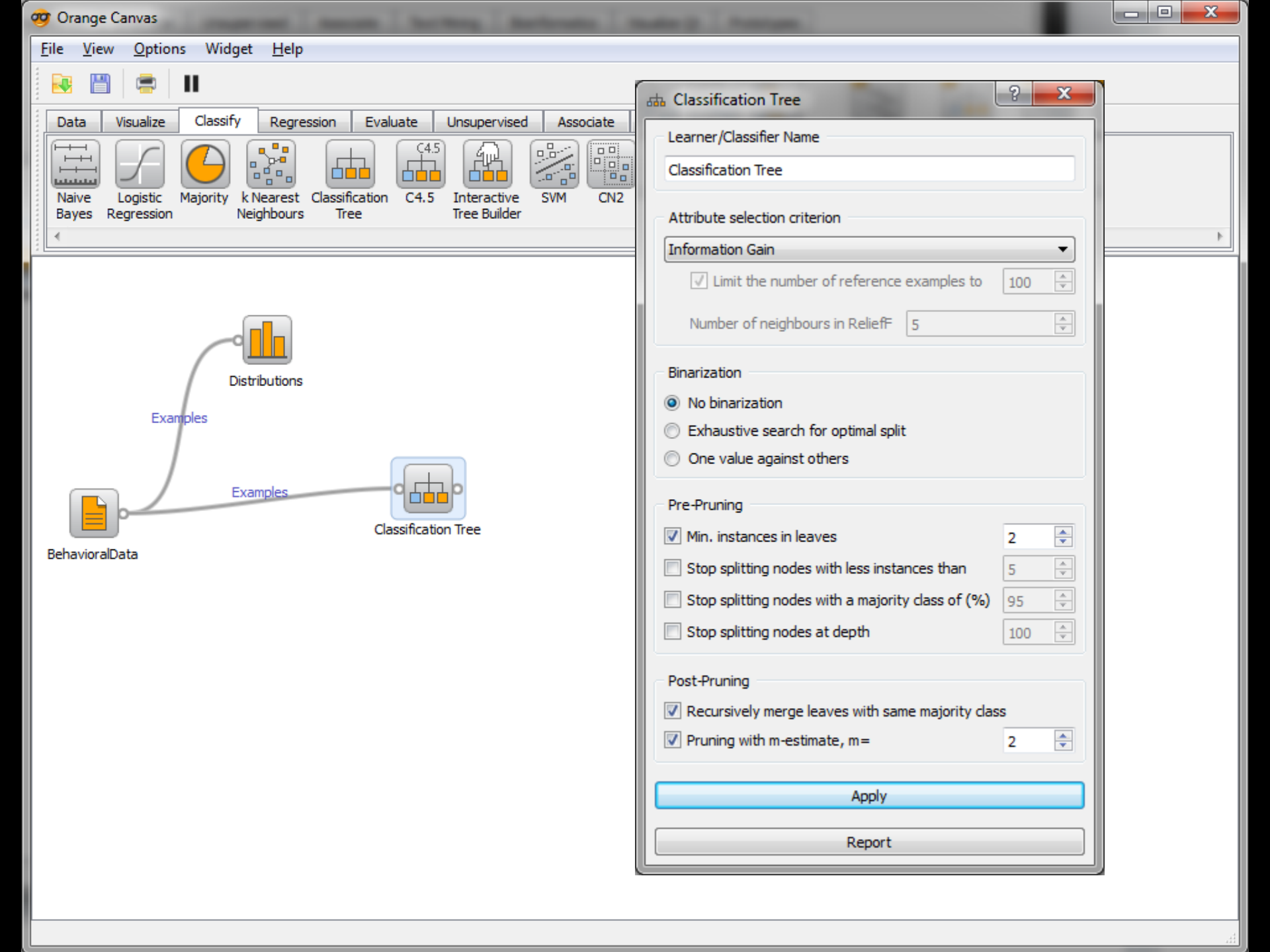
Show confidence intervals

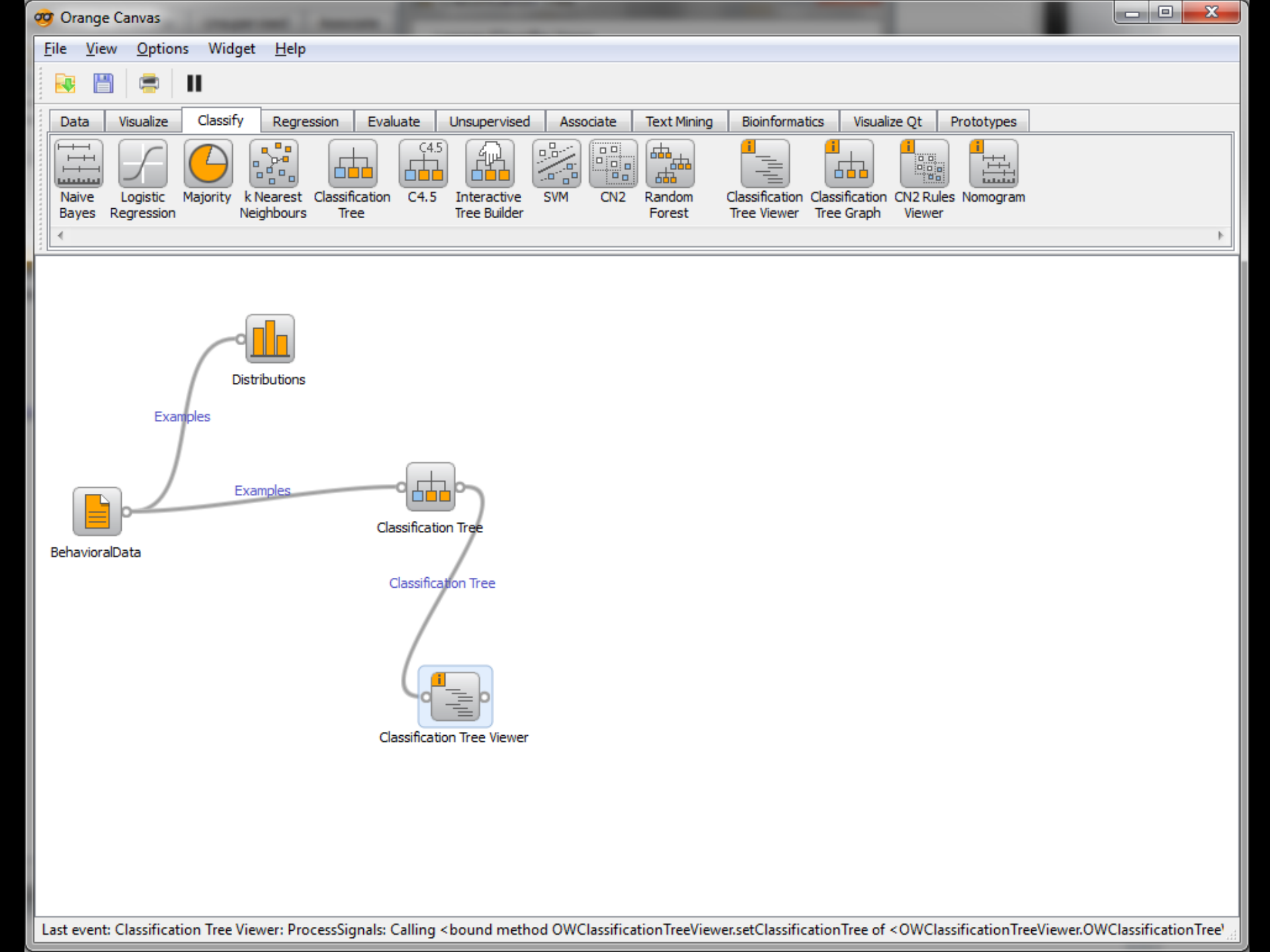
Smooth probability lines

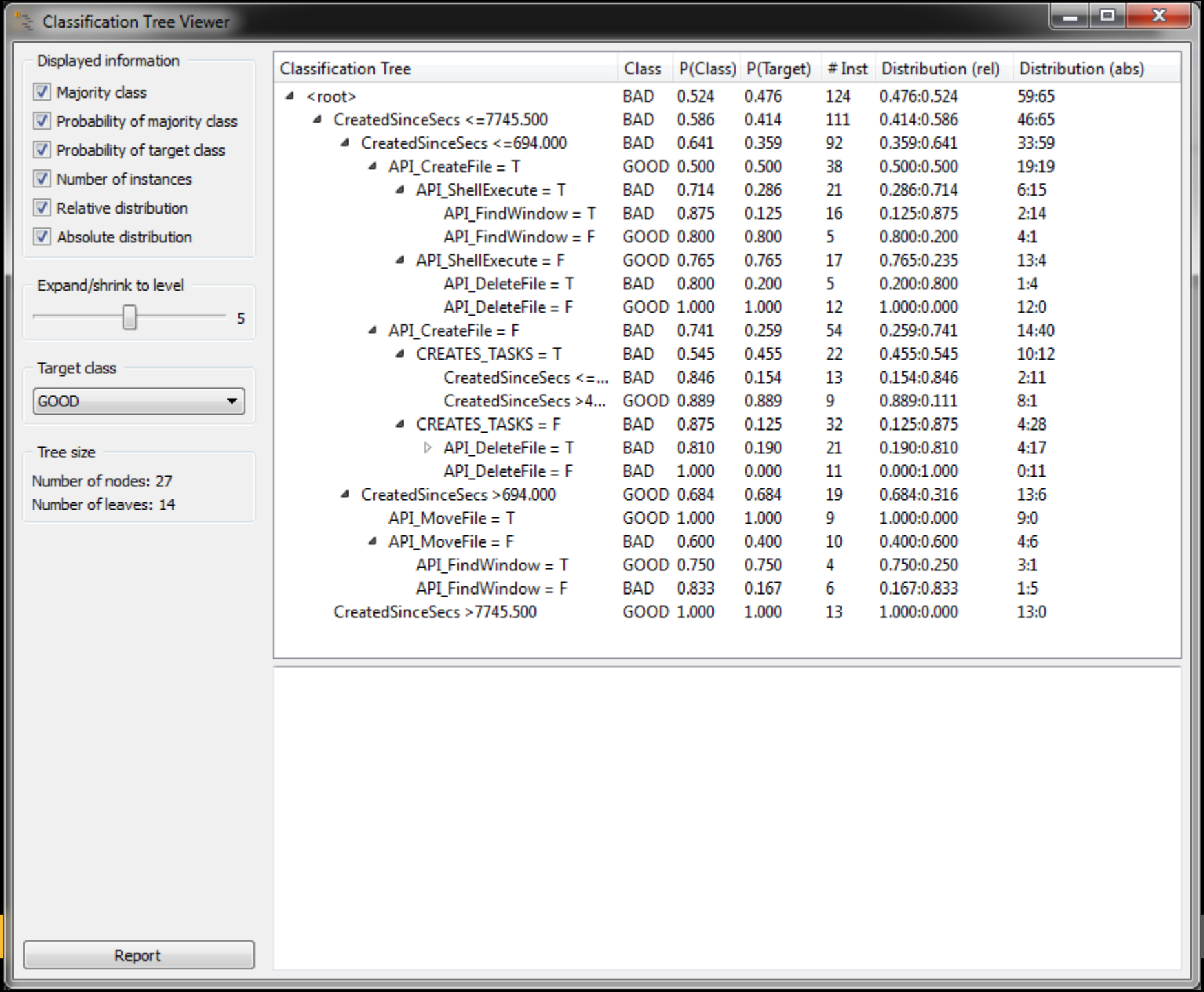
Save Graph Report

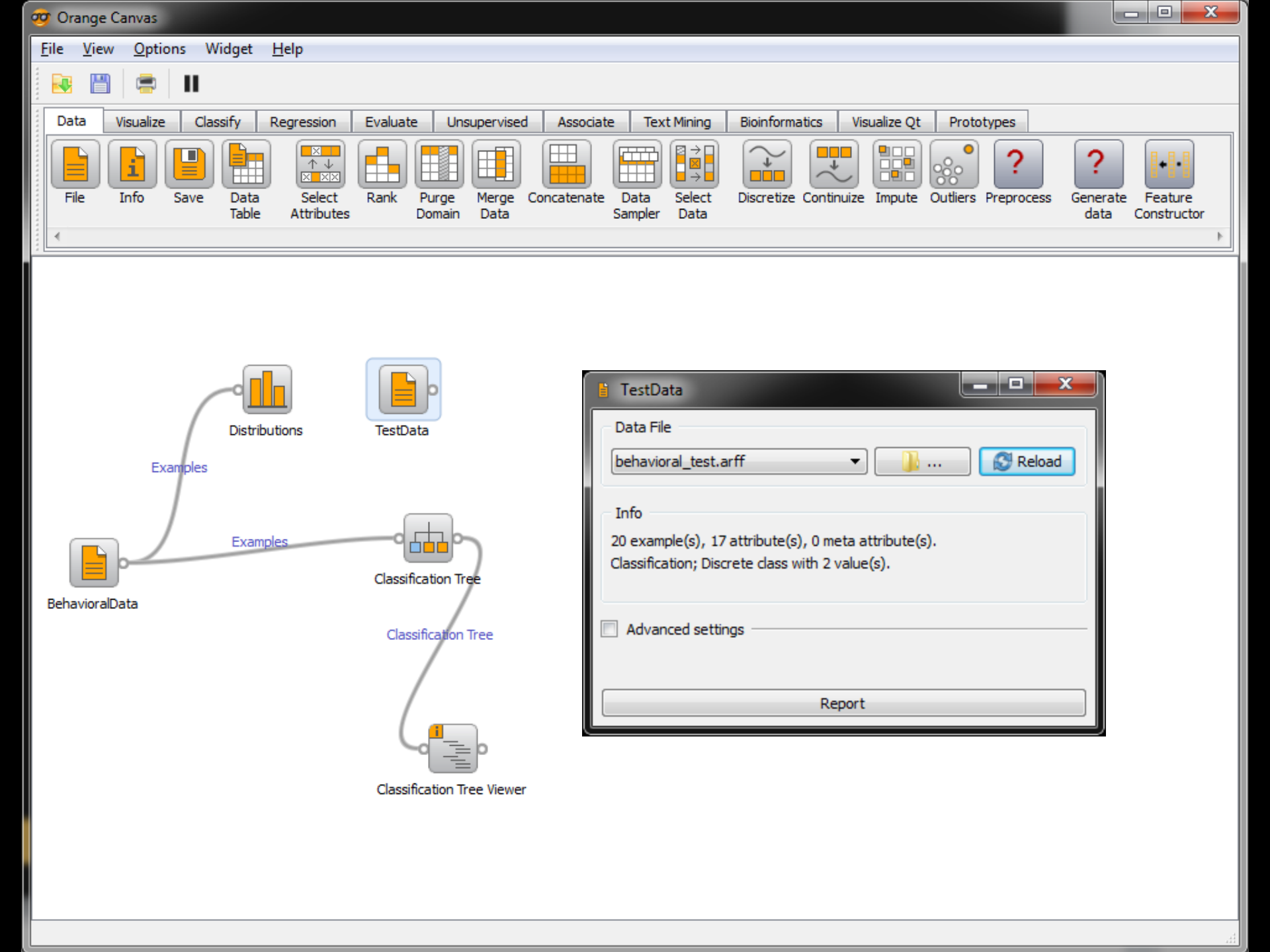


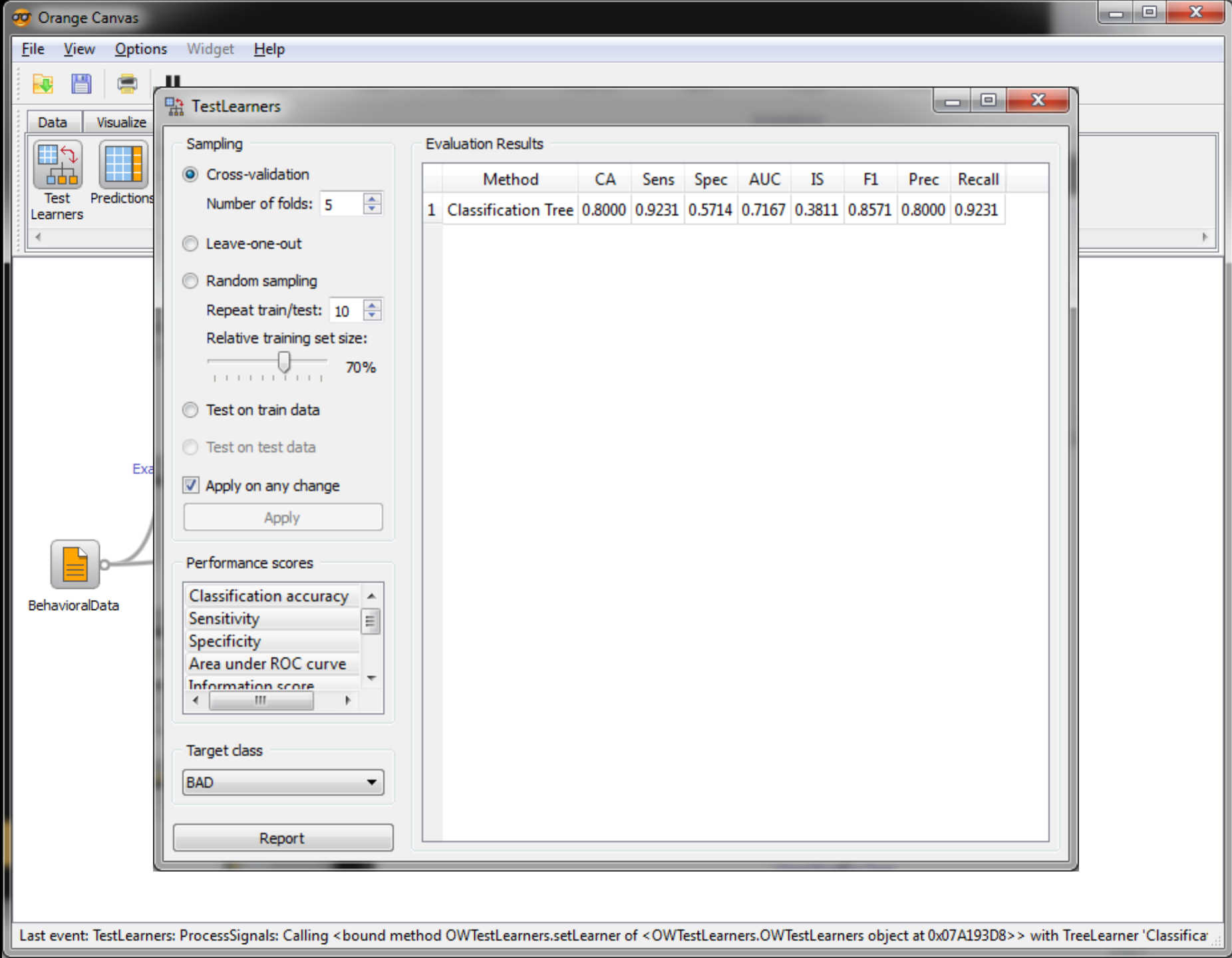

Orange Canvas
File View Options Widget Help
Data Visualize Classify Regression Evaluate Unsupervised Associate Text Mining Bioinformatics Visualize Qt Prototypes
Distributions
Attribute Statistics
Scatterplot
Linear Projection
Radviz
Polyviz
Parallel Coordinates
Survey Plot
Correspondence Analysis
Multi Correspondence Analysis
Mosaic Display
Sieve Diagram
Sieve multigram
Distributions
Examples
BehavioralData

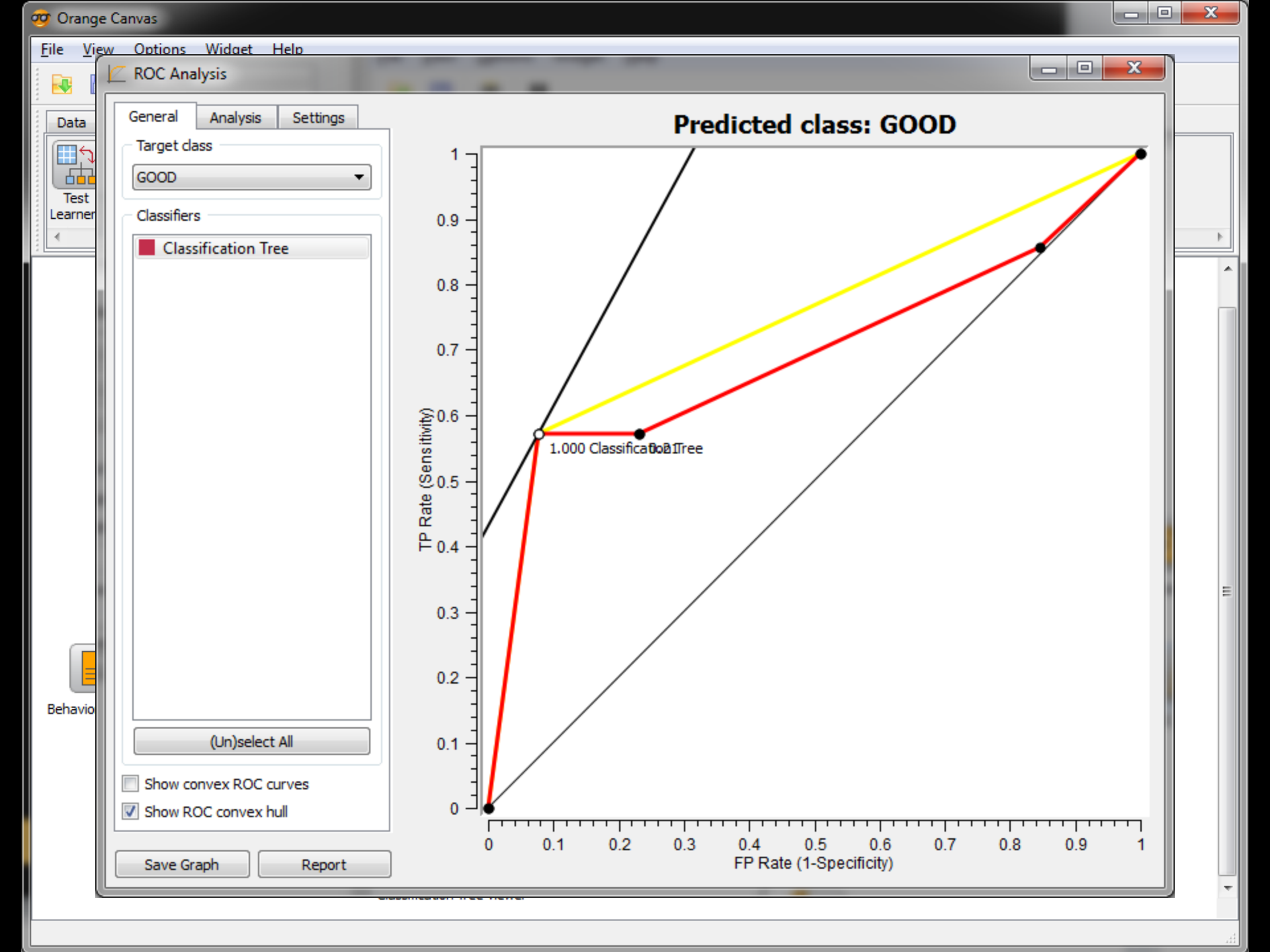

Orange Canvas
File View Options Widget Help
Data Visualize Classify Regression Evaluate Unsupervised Associate
Naive Bayes
Logistic Regression
Majority
k Nearest Neighbours
Classification Tree
C4.5
Interactive Tree Builder
SVM
CN2
Distributions
Examples
Examples
BehavioralData
Classification Tree
Classification Tree
Learner/Classifier Name
Classification Tree
Attribute selection criterion
Information Gain
Limit the number of reference examples to 100
Number of neighbours in ReliefF 5
Binarization
No binarization
Exhaustive search for optimal split
One value against others
Pre-Pruning
Min. instances in leaves 2
Stop splitting nodes with less instances than 5
Stop splitting nodes with a majority class of (%) 95
Stop splitting nodes at depth 100
Post-Pruning
Recursively merge leaves with same majority class
Pruning with m-estimate, m= 2
Apply
Report


Orange Canvas
File View Options Widget Help
Data Visualize Classify Regression Evaluate Unsupervised Associate Text Mining Bioinformatics Visualize Qt Prototypes
Naive Bayes
Logistic Regression
Majority
k Nearest Neighbours
Classification Tree
C4.5
Interactive Tree Builder
SVM
CN2
Random Forest
Classification Tree Viewer
Classification Tree Graph
CN2 Rules Viewer
Nomogram
Distributions
Examples
Examples
BehavioralData
Classification Tree
Classification Tree
Classification Tree Viewer
Last event: Classification Tree Viewer: ProcessSignals: Calling <bound method OWClassificationTreeViewer.setClassificationTree of <OWClassificationTreeViewer.OWClassificationTree


Classification Tree Viewer
Displayed information
Majority class
Probability of majority class
Probability of target class
Number of instances
Relative distribution
Absolute distribution
Expand/shrink to level
5
Target class
GOOD
Tree size
Number of nodes: 27
Number of leaves: 14
Report
Classification Tree | Class | P(Class) | P(Target) | # Inst | Distribution (rel) | Distribution (abs)
<root> | BAD | 0.524 | 0.476 | 124 | 0.476:0.524 | 59:65
CreatedSinceSecs <=7745.500 | BAD | 0.586 | 0.414 | 111 | 0.414:0.586 | 46:65
CreatedSinceSecs <=694.000 | BAD | 0.641 | 0.359 | 92 | 0.359:0.641 | 33:59
API_CreateFile = T | GOOD | 0.500 | 0.500 | 38 | 0.500:0.500 | 19:19
API_ShellExecute = T | BAD | 0.714 | 0.286 | 21 | 0.286:0.714 | 6:15
API_FindWindow = T | BAD | 0.875 | 0.125 | 16 | 0.125:0.875 | 2:14
API_FindWindow = F | GOOD | 0.800 | 0.800 | 5 | 0.800:0.200 | 4:1
API_ShellExecute = F | GOOD | 0.765 | 0.765 | 17 | 0.765:0.235 | 13:4
API_DeleteFile = T | BAD | 0.800 | 0.200 | 5 | 0.200:0.800 | 1:4
API_DeleteFile = F | GOOD | 1.000 | 1.000 | 12 | 1.000:0.000 | 12:0
API_CreateFile = F | BAD | 0.741 | 0.259 | 54 | 0.259:0.741 | 14:40
CREATES_TASKS = T | BAD | 0.545 | 0.455 | 22 | 0.455:0.545 | 10:12
CreatedSinceSecs <=... | BAD | 0.846 | 0.154 | 13 | 0.154:0.846 | 2:11
CreatedSinceSecs >4... | GOOD | 0.889 | 0.889 | 9 | 0.889:0.111 | 8:1
CREATES_TASKS = F | BAD | 0.875 | 0.125 | 32 | 0.125:0.875 | 4:28
API_DeleteFile = T | BAD | 0.810 | 0.190 | 21 | 0.190:0.810 | 4:17
API_DeleteFile = F | BAD | 1.000 | 0.000 | 11 | 0.000:1.000 | 0:11
CreatedSinceSecs >694.000 | GOOD | 0.684 | 0.684 | 19 | 0.684:0.316 | 13:6
API_MoveFile = T | GOOD | 1.000 | 1.000 | 9 | 1.000:0.000 | 9:0
API_MoveFile = F | BAD | 0.600 | 0.400 | 10 | 0.400:0.600 | 4:6
API_FindWindow = T | GOOD | 0.750 | 0.750 | 4 | 0.750:0.250 | 3:1
API_FindWindow = F | BAD | 0.833 | 0.167 | 6 | 0.167:0.833 | 1:5
CreatedSinceSecs >7745.500 | GOOD | 1.000 | 1.000 | 13 | 1.000:0.000 | 13:0


Orange Canvas
File View Options Widget Help
Data Visualize Classify Regression Evaluate Unsupervised Associate Text Mining Bioinformatics Visualize Qt Prototypes
File Info Save Data Table Select Attributes Rank Purge Domain Merge Data Concatenate Data Sampler Select Data Discretize Continuize Impute Outliers Preprocess Generate data Feature Constructor
Distributions
TestData
Examples
Examples
BehavioralData
Classification Tree
Classification Tree
Classification Tree Viewer
TestData
Data File
behavioral_test.arff
...
Reload
Info
20 example(s), 17 attribute(s), 0 meta attribute(s).
Classification; Discrete class with 2 value(s).
Advanced settings
Report


Orange Canvas
File View Options Widget Help
Data Visualize
Test Learners
Predictions
TestLearners
Sampling
Cross-validation
Number of folds: 5
Leave-one-out
Random sampling
Repeat train/test: 10
Relative training set size:
70%
Test on train data
Test on test data
Apply on any change
Apply
Performance scores
Classification accuracy
Sensitivity
Specificity
Area under ROC curve
Information score
Target class
BAD
Report
Evaluation Results
Method CA Sens Spec AUC IS F1 Prec Recall
1 Classification Tree 0.8000 0.9231 0.5714 0.7167 0.3811 0.8571 0.8000 0.9231
BehavioralData
Last event: TestLearners: ProcessSignals: Calling <bound method OWTestLearners.setLearner of <OWTestLearners.OWTestLearners object at 0x07A193D8>> with TreeLearner 'Classifica


Orange Canvas
File View Options Widget Help
ROC Analysis
General Analysis Settings
Target class
GOOD
Classifiers
Classification Tree
(Un)select All
Show convex ROC curves
Show ROC convex hull
Save Graph
Report
Predicted class: GOOD
TP Rate (Sensitivity)
FP Rate (1-Specificity)
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
1.000 Classification Tree
Test Learner
Data

Go back…