

IA645: Data Analysis for Cybersecurity

Linux syslog Data Analysis and Machine Learning

Team: Aaron Liske, Mia Jones, William Smith III, Yogesh Chavarkar

Faculty Guidance: Dr Omar Darwish

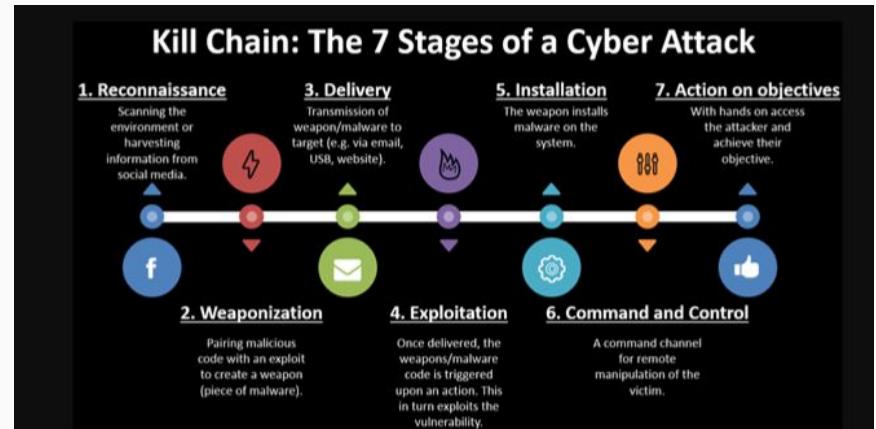
Game Above College of Engineering and Technology
Eastern Michigan University

Agenda

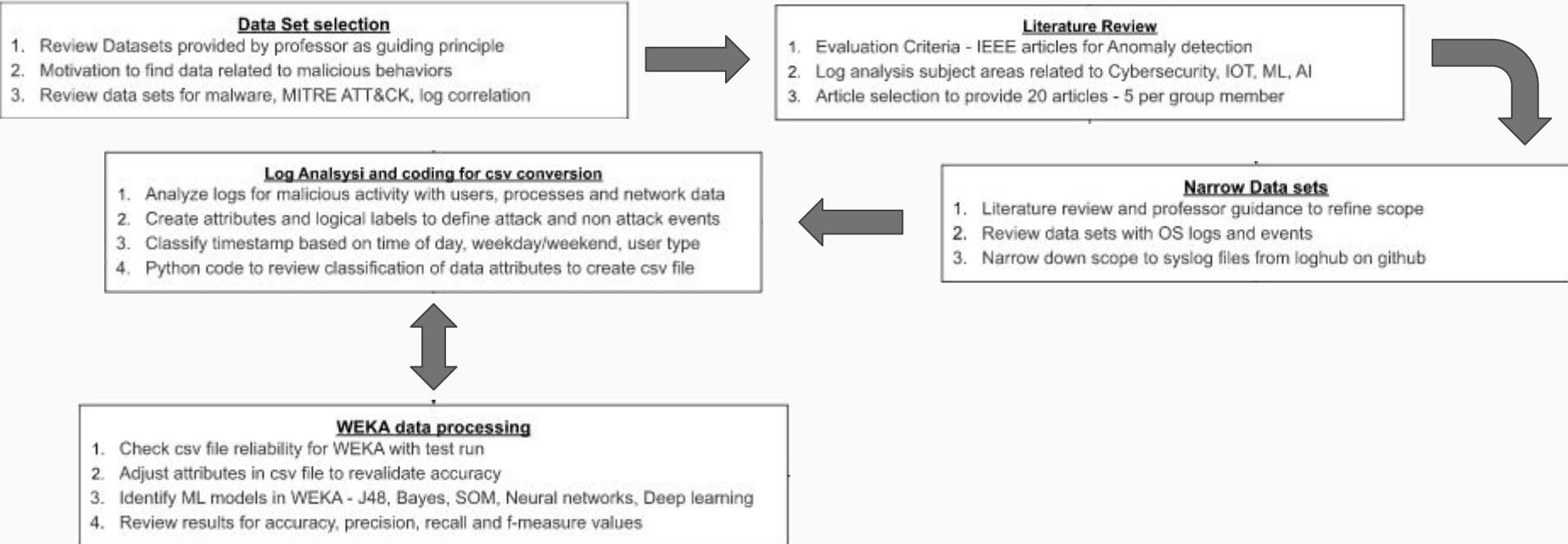
- Introduction
- Proposed Work
- Results
- Limitation
- Conclusion
- Future Work

Introduction

- Goal is to aid detection and classification of malware using log file analysis using ML.
- Classification of outbound connections from compromised systems



Workflow

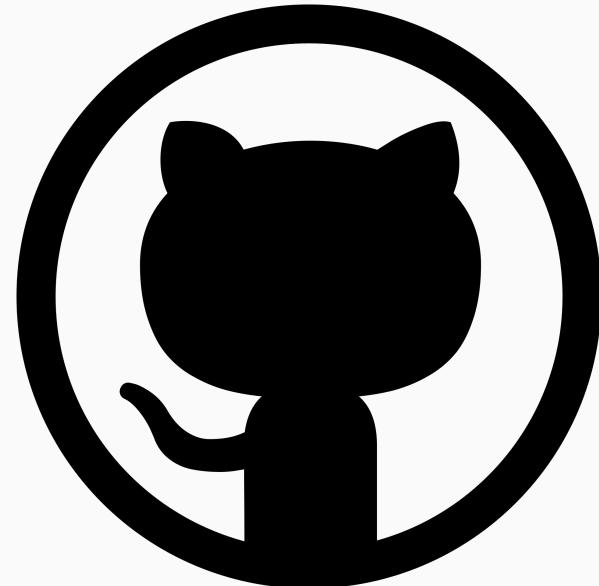


Dataset Selection

- Objective: identify malicious patterns found in authentication failures, ftp, https and nfs-rpc connections.
- Scope of the project is to focus on a single area compromising of operating system
- Dataset used was linux syslog data from var/log/messages.

Tools Used

- Visual Studio Code
- EMU Azure Tenant
- Log analytics workspace
 - Custom tables linuxlogs_CL and ia645nixlogs_CL
- SQL database
- Python



Codes and Scripts Used

- Extract, Transform, and Load (ETL) methods were used in a Python script to transform the data from the logs into a usable CSV that Weka could readily use.

```
import os
import re
import datetime
import calendar

directory = 'logs'
for filename in os.listdir(directory):
    if filename.endswith('.log'):
        directory = os.path.dirname(filename)
        file = open(directory + '/' + filename, 'r')
        file.seek(0, 0)
        lines = file.readlines()
        year = filename.split('.')[0]
        year = int(year)
        for line in lines:
            date = line[0:10]
            date = datetime.datetime.strptime(date, '%Y-%m-%d')
            day_of_week = date.strftime("%A")
            time_of_day = date.strftime("%H:%M:%S")
            if time_of_day <= "06:00": #if hour < 6
                time_of_day = 'early morning'
            if time_of_day >= "06:00" and time_of_day < "12:00": #if hour > 6 and hour < 12
                time_of_day = 'morning'
            if time_of_day >= "12:00" and time_of_day < "18:00": #if hour > 12 and hour < 21
                time_of_day = 'afternoon'
            if time_of_day >= "18:00": #if hour > 21
                time_of_day = 'evening'
            print("%s (%s) %s (%s) %s (%s)" % (date, day_of_week, a.group(1), time_of_day, a.group(4), a.group(3)))
        file.close()
```

```
#TODO: move user level to appropriate function for output instead of dual outputs for brevity
def message_parse(svc, msg, timestamp):
    #gain access to the variables needed
    global prev_service
    global prev_host
    global prev_timestamp
    global prev_timestamp

    #set variable defaults
    suspicious = 'NON-ATTACK'
    user_level = 'no_euid'

    #because sshd and samba have such similar messages, processing is identical
    if svc == 'sshd' or svc == 'samba':
        m = re.search('^(.*?)(\s+euid=\w+)(\.\w+)(\.\w+)(\.\w+)$', msg)
        if(m):
            user_level = user_type(m.group(3))
            if user_level == "root":
                suspicious = 'ATTACK'
            else:
                suspicious = 'NON-ATTACK'
        else:
            suspicious = 'NON-ATTACK'

    #process log for ftpd. mark as suspicious if line matches previous entry's remote host or is within 2 seconds per timestamp
    if svc == 'ftpd' or svc == 'ftp':
        m = re.search('^(.*?)(\.\w+)(\.\w+)(\.\w+)(\.\w+)$', msg.replace('.','.'))
        curr_host = ''
        if(m):
            curr_host = m.group(2)
            if prev_service == 'ftpd' or prev_service == 'ftp':
                if prev_timestamp != '':
                    #parse out the seconds for the timestamp
                    #TODO: add logic for seconds under 2 to loop back to 59 for timestamp checking
                    prev_seconds = int(prev_timestamp.split(":")[2])
                    curr_seconds = int(timestamp.split(":")[2])
                    delta_time = curr_seconds - prev_seconds
                    if (curr_seconds - prev_seconds < 2 or curr_host == prev_host) and (msg.strip() != "FTP session closed"):
                        suspicious = 'ATTACK'
                    else:
                        suspicious = 'NON-ATTACK'
                #set host and timestamp variables for comparison later
                prev_host = curr_host
                prev_timestamp = timestamp
            else:
                suspicious = 'NON-ATTACK'
```

Log Analysis

Python Used

- Utilized regex to pull each log message apart to find the relevant sections for processing

```
^(\w\w\w\s+\d+) (\d\d:\d\d:\d\d) (.*?)  
(.*) (\W) (.*?) (.*?) (\W) (.*?) (.*?)$
```

Main Regex for Processing Log Line

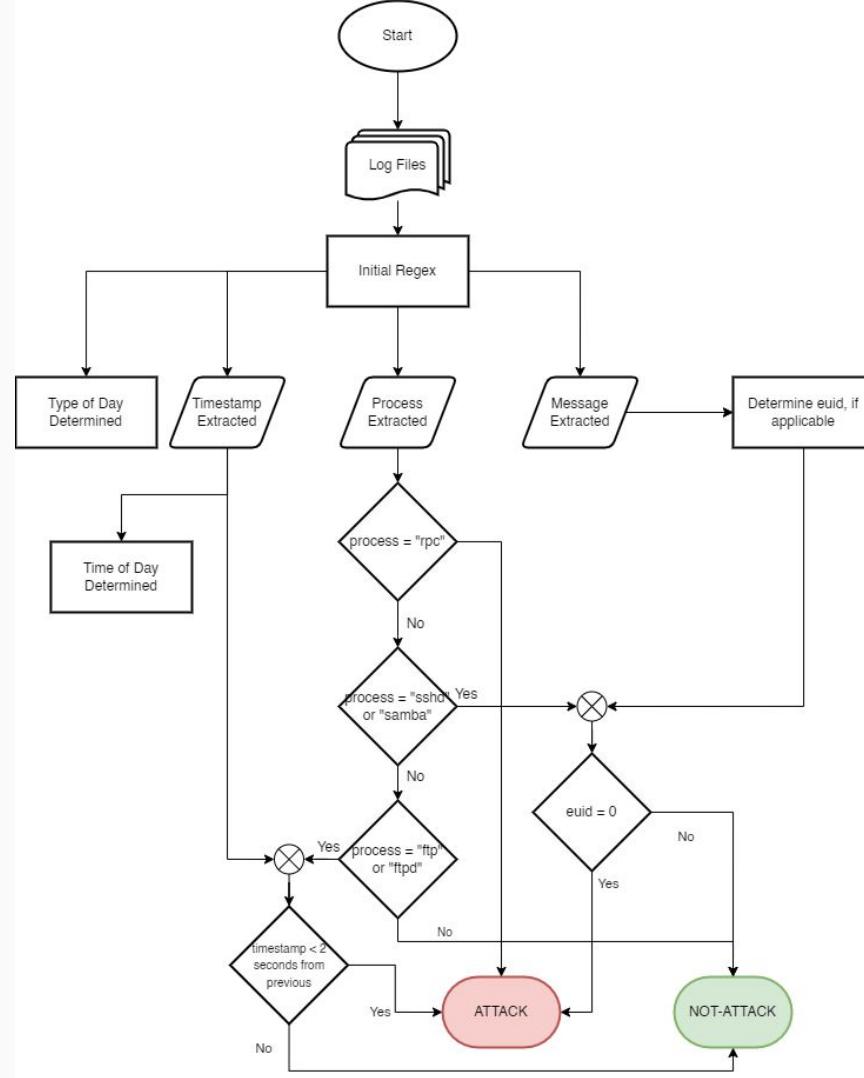
Regex Used

- Match groups used are the date, time, process, and the message

```
if hour >= 0 and hour <= 6:  
    time_of_day = 'early morning'  
if hour >= 7 and hour <= 11:  
    time_of_day = 'morning'  
if hour >= 12 and hour <= 17:  
    time_of_day = 'afternoon'  
if hour >= 18 and hour <= 21:  
    time_of_day = 'evening'  
if hour > 21:  
    time_of_day = 'night'
```

Determination of Time of Day

Individual Log Entry Flowchart



Weka Data Processing

- CSV below shows accuracy in a range of 93.65-90.03% running classifiers J48, BayesNet, and Naive Bayes.. .
- Data set further refined in the CSV below to higher consistency 90.02-90.03%.
- User level category was removed and suspicious changed to attack or non-attack.

TYPE_OF_DAY	TIME_OF_DAY	PROCESS	USER_LEVEL	SUSPICIOUS
NON_WORKING	early morning	syslogd	no_euid	FALSE
NON_WORKING	early morning	syslogd	no_euid	FALSE
WORKING	morning	syslogd	no_euid	FALSE
WORKING	morning	syslog	no_euid	FALSE
WORKING	morning	syslog	no_euid	FALSE
WORKING	morning	kernel	no_euid	FALSE

TYPE_OF_DAY	TIME_OF_DAY	PROCESS	SUSPICIOUS
NON_WORKING	early morning	syslogd	NON-ATTACK
NON_WORKING	early morning	syslogd	NON-ATTACK
WORKING	morning	syslogd	NON-ATTACK
WORKING	morning	syslog	NON-ATTACK
WORKING	morning	syslog	NON-ATTACK
WORKING	morning	kernel	NON-ATTACK

Dataset Summary

Data Type: Linux syslog files labeled for suspicious activity
File Type: CSV
Name: WEKA Dataset
Instances: 71704
Attributes: 4
TYPE_OF_DAY
TIME_OF_DAY
PROCESS
SUSPICIOUS

Algorithms

Decision Tree - J48

Accuracy	Precision	Recall	F1-Measure
90.03	1.0, 0.77, 0.923	0.850, 1.0, 0.9	0.92, 0.87, 0.90
Inference: Folds 2, 5, and 10 with the same accuracy. Split 66% is slightly higher with 93.9% accuracy.			

Naive Bayes

Accuracy	Precision	Recall	F1-Measure
90.02	1.0, 77, .923	0.850, 1.0, 090	0.919, 0.870, 0.903
Inference: Folds 2, 5, and 10 with the same accuracy. Split 66% is slightly higher with 90.32% accuracy			

Algorithms

Bayes Net

Accuracy	Precision	Recall	F1-Measure
90.02	1.0, 0.77, 0.92	0.85, 1.0, 0.90	0.92, 0.87, 0.90
Inference: Folds 2, 5, and 10 with the same accuracy. Split 66% is slightly higher with 90.32% accuracy			

Random Tree

Accuracy	Precision	Recall	F1-Measure
90.03	1.0, 0.770, 0.92	0.85, 1.0, 0.900	0.92, 0.87, 0.90
Inference: Folds 2, 5, and 10 with the same accuracy. Split 66% is slightly higher with 90.32% accuracy			

SMO Support Vector Machine

Accuracy	Precision	Recall	F1-Measure
90.03	1.0, 0.77, 0.92	0.85, 1.0, 0.9	0.92, 0.87, 0.90
Inference: Folds 2, 5, and 10 with the same accuracy. Split 66% is slightly higher with 90.32% accuracy			

Multilayer Perceptron (Deep Learning Neural Network)

Accuracy	Precision	Recall	F1-Measure
90.03	1.0, 0.77, .92	0.95, 1.0, 0.9	0.92, 0.87, 0.90
Inference: Similar results using the “Use training set” test option.			

Limitation

- Services in a Linux System
- Retaining User Data



Conclusion

- “Stepping stone” in predicting attacks in near-real time
- Learns from behaviors and patterns, and not dependant on hard coded signatures
- Classifies outbound connections from compromised systems

Future Work

This Machine Learning tool can effectively work in two different cybersecurity functions

- Compliance using syslogs showing exposure of misconfigured assets
- Attack prediction for threat hunting and collecting Cyber threat Intelligence from observed attack campaigns

Appendix

Weka Data Processing

➤ BayesNet Classifier

Classifier **Choose** **NaiveBayes**

Test options

- Use training set
- Supplied test set
- Cross-validation Folds 10
- Percentage split % 66
- More options...

(Nom) SUSPICIOUS

Start Stop

Classifier output

```
PROCESS(81) - SUSPICIOUS
USER_LEVEL = root: SUSPICIOUS
SUSPICIOUS(2):
LogScore Bayes: -313675.9772954669
LogScore BDeu: -314646.899977337
LogScore MDL: -314755.4247124071
LogScore ENTROPY: -313777.1483837663
LogScore AIC: -313952.1483037663
```

Time taken to build model: 0.04 seconds

== Stratified cross-validation ==

== Summary ==

	Correctly Classified Instances	64554	90.0285 %
Incorrectly Classified Instances	7150	9.9715 %	
Kappa statistic	0.7915		
Mean absolute error	0.1294		
Root mean squared error	0.2568		
Relative absolute error	29.0513 %		
Root relative squared error	54.4279 %		
Total Number of Instances	71704		

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.858	0.000	1.000	0.850	0.919	0.889	0.959	0.981	0.981	FALSE
1.000	0.150	0.770	1.000	0.870	0.889	0.959	0.917	0.917	TRUE
Weighted Avg.	0.900	0.050	0.923	0.900	0.903	0.889	0.959	0.960	

== Confusion Matrix ==

a	b	<-- classified as
40551	7150	a = FALSE
0	24003	b = TRUE

13:48:43 - bayes.BayesNet
13:48:49 - bayes.NaiveBayes

➤ J48 Classifier

Preprocess **Classify** **Cluster** **Associate** **Select attributes** **Visualize**

Classifier **Choose** **NaiveBayes**

Test options

- Use training set
- Supplied test set
- Cross-validation Folds 10
- Percentage split % 66
- More options...

(Nom) SUSPICIOUS

Start Stop

Classifier output

```
PROCESS = recall: FALSE (1.0)
| PROCESS = PAM: FALSE (2.0)
| PROCESS = unix_ckpwd: FALSE (306.2)
USER_LEVEL = root: TRUE (10038.0)
USER_LEVEL = other: FALSE (1.0)
```

Number of Leaves : 83

Size of the tree : 85

Time taken to build model: 0.13 seconds

== Stratified cross-validation ==

== Summary ==

	Correctly Classified Instances	67153	93.6531 %
Incorrectly Classified Instances	4551	6.3469 %	
Kappa statistic	0.8639		
Mean absolute error	0.0946		
Root mean squared error	0.2174		
Relative absolute error	21.2474 %		
Root relative squared error	46.0625 %		
Total Number of Instances	71704		

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class	
0.905	0.000	1.000	0.905	0.905	0.905	0.950	0.872	0.973	0.987	FALSE
1.000	0.095	0.841	1.000	0.913	0.872	0.973	0.937	0.957	0.957	TRUE
Weighted Avg.	0.937	0.052	0.947	0.937	0.937	0.958	0.872	0.973	0.971	

== Confusion Matrix ==

a	b	<-- classified as
43150	4551	a = FALSE
0	24003	b = TRUE

13:46:01 - trees.J48
13:48:43 - bayes.BayesNet
13:48:49 - bayes.NaiveBayes

Weka Data Processing

➤ SMO- Support Vector Machine Classifier

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **NaiveBayes**

Test options

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

More options...

Number of kernel evaluations: 80505499 (28.778% cached)

(Nom) SUSPICIOUS

Start Stop

Result list (right-click for options)

```
18:00:19 - functions.MultilayerPerceptron
18:01:45 - functions.MultilayerPerceptron
18:02:32 - functions.MultilayerPerceptron
18:04:28 - functions.MultilayerPerceptron
18:06:09 - functions.SMO
18:06:35 - functions.SMO
18:06:52 - functions.SMO
18:10:37 - functions.SMO
18:15:20 - functions.SMO
18:24:46 - functions.SMO
07:15:15 - bayes.BayesNet
07:15:37 - bayes.NaiveBayesMultinomialText
07:16:04 - trees.J48
07:16:22 - trees.HoeffdingTree
07:17:30 - trees.RandomTree
07:18:26 - trees.RandomForest
13:46:01 - trees.J48
13:48:43 - bayes.BayesNet
13:48:49 - bayes.NaiveBayes
```

Time taken to build model: 131.68 seconds

== Stratified cross-validation ==

== Summary ==

	Correctly Classified Instances	64554	90.0285 %
Incorrectly Classified Instances	7150	9.9715 %	
Kappa statistic	0.7915		
Mean absolute error	0.0997		
Root mean squared error	0.3158		
Relative absolute error	22.3884 %		
Root relative squared error	66.9157 %		
Total Number of Instances	71704		

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.850	0.000	1.000	0.850	0.919	0.889	0.925	0.950	NON-ATTACK
	1.000	0.150	0.770	1.000	0.870	0.889	0.925	0.770	ATTACK

== Confusion Matrix ==

a	b	<-- classified as
40551	7150	a = NON-ATTACK
0	24003	b = ATTACK

➤ Multilayer Perceptron

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **NaiveBayes**

Test options

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

More options...

Input
Node 0
Class ATTACK
Input
Node 1

Time taken to build model: 97.65 seconds

== Evaluation on training set ==

Time taken to test model on training data: 1.73 seconds

== Summary ==

	Correctly Classified Instances	64554	90.0285 %
Incorrectly Classified Instances	7150	9.9715 %	
Kappa statistic	0.7915		
Mean absolute error	0.1367		
Root mean squared error	0.2767		
Relative absolute error	30.6891 %		
Root relative squared error	58.6296 %		
Total Number of Instances	71704		

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.850	0.000	1.000	0.919	0.889	0.941	0.974	0.950	NON-ATTACK
	1.000	0.150	0.770	1.000	0.870	0.889	0.941	0.839	ATTACK

== Confusion Matrix ==

a	b	<-- classified as
40551	7150	a = NON-ATTACK
0	24003	b = ATTACK

Weka Data Processing

➤ J48

Test options

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

Classifier output

```

PROCESS = modprobe: NON-ATTACK (6.0)
PROCESS = net: NON-ATTACK (1.0)
PROCESS = PAM: NON-ATTACK (2.0)
PROCESS = unix_chkpwd: NON-ATTACK (306.17)

```

Number of Leaves : 81

Size of the tree : 82

Time taken to build model: 0.07 seconds

== Stratified cross-validation ==

== Summary ==

	Correctly Classified Instances	64554	90.0285 %
Incorrectly Classified Instances	7150	9.9715 %	
Kappa statistic	0.7915		
Mean absolute error	0.1517		
Root mean squared error	0.2752		
Relative absolute error	34.0493 %		
Root relative squared error	58.3228 %		
Total Number of Instances	71704		

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.850	0.000	1.000	0.850	0.919	0.809	0.932	0.971		NON-ATTACK
1.000	0.150	0.770	1.000	0.870	0.809	0.932	0.810		ATTACK
Weighted Avg.	0.900	0.050	0.923	0.900	0.903	0.809	0.932	0.917	

== Confusion Matrix ==

	a	b	← classified as
40551	7150	1	a = NON-ATTACK
0 24003	1	b = ATTACK	

➤ HoefttingTree Classifier

Test options

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

Classifier output

```

PROCESS = walli: NON-ATTACK (3.000)
PROCESS = fingerid: NON-ATTACK (3.000)
PROCESS = rbf: NON-ATTACK (3.000)
PROCESS = rshd: NON-ATTACK (3.000)
PROCESS = rlojind: NON-ATTACK (3.000)
PROCESS = modprobe: NON-ATTACK (7.000)
PROCESS = net: NON-ATTACK (2.000)
PROCESS = PAM: NON-ATTACK (3.000)
PROCESS = unix_chkpwd: NON-ATTACK (307.000)

```

(Nom) SUSPICIOUS

Time taken to build model: 0.1 seconds

== Stratified cross-validation ==

== Summary ==

Result list (right-click for options)

	Correctly Classified Instances	64554	90.0285 %
Incorrectly Classified Instances	7150	9.9715 %	
Kappa statistic	0.7915		
Mean absolute error	0.1503		
Root mean squared error	0.2736		
Relative absolute error	33.7438 %		
Root relative squared error	57.9721 %		
Total Number of Instances	71704		

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.850	0.000	1.000	0.850	0.919	0.809	0.932	0.971		NON-ATTACK
1.000	0.150	0.770	1.000	0.870	0.809	0.932	0.810		ATTACK
Weighted Avg.	0.900	0.050	0.923	0.900	0.903	0.809	0.932	0.917	

== Confusion Matrix ==

	a	b	← classified as
40551	7150	1	a = NON-ATTACK
0 24003	1	b = ATTACK	

Weka Data Processing

➤ RandomTree Classifier

Test options

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

(Nom) SUSPICIOUS

Start Stop

Result list (right-click for options)

- 18:00:19 - functions.MultilayerPerceptron
- 18:01:45 - functions.MultilayerPerceptron
- 18:02:32 - functions.MultilayerPerceptron
- 18:04:28 - functions.MultilayerPerceptron
- 18:06:09 - functions.SMO
- 18:06:35 - functions.SMO
- 18:06:52 - functions.SMO
- 18:10:37 - functions.SMO
- 18:15:20 - functions.SMO
- 18:24:46 - functions.SMO
- 07:15:15 - bayes.BayesNet
- 07:15:37 - bayes.NaiveBayesMultinomialText
- 07:16:04 - trees.J48
- 07:16:22 - trees.HoeffdingTree
- 07:17:30 - trees.RandomTree**
- 07:18:26 - trees.RandomForest
- 13:46:01 - trees.J48
- 13:48:43 - bayes.BayesNet
- 13:48:49 - bayes.NaiveBayes

Classifier output

```
PROCESS = in : NON-ATTACK (2/0)
PROCESS = rshd : NON-ATTACK (2/0)
PROCESS = rlogin : NON-ATTACK (2/0)
PROCESS = modprobe : NON-ATTACK (6/0)
PROCESS = net : NON-ATTACK (1/0)
PROCESS = PAM : NON-ATTACK (2/0)
PROCESS = unix_chkpwd : NON-ATTACK (306.17/0)
```

Size of the tree : 135

Time taken to build model: 0.07 seconds

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

Correctly Classified Instances	64554	90.0285 %
Incorrectly Classified Instances	7150	9.9715 %
Kappa statistic	0.7915	
Mean absolute error	0.1495	
Root mean squared error	0.2733	
Relative absolute error	33.5565 %	
Root relative squared error	57.9206 %	
Total Number of Instances	71704	

==== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.850	0.000	1.000	0.850	0.919	0.889	0.941	0.975		NON-AT
1.000	0.150	0.770	1.000	0.870	0.889	0.941	0.843		ATTACK
Weighted Avg.	0.900	0.050	0.923	0.900	0.903	0.889	0.941	0.931	

==== Confusion Matrix ===

		a b	<-- classified as
a	b	40551 7150	a = NON-ATTACK
b	a	0 24003	b = ATTACK

Results

Deep learning multilayer perceptron

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 -G -R

Test options

Use training set
 Supplied test set Set...
 Cross-validation Folds 2
 Percentage split % 66
More options...

(Nom) SUSPICIOUS

Start Stop

Result list (right-click for options)

09:27:06 - trees.J48
09:27:14 - trees.J48
09:50:59 - trees.J48
09:51:33 - trees.J48
17:05:38 - trees.J48
17:07:21 - trees.J48
17:07:28 - trees.J48
19:27:14 - bayes.NaiveBayes
19:27:27 - bayes.NaiveBayes
19:27:37 - bayes.NaiveBayes
19:31:17 - bayes.NaiveBayes
19:32:31 - bayes.BayesNet
19:32:36 - bayes.BayesNet
19:32:41 - bayes.BayesNet
19:32:43 - bayes.BayesNet
19:37:08 - trees.RandomTree
19:37:13 - trees.RandomTree
19:37:18 - trees.RandomTree
19:37:20 - trees.RandomTree
19:46:35 - trees.RandomTree
19:46:49 - bayes.NaiveBayesMultinomialText
19:46:56 - bayes.NaiveBayesMultinomialText
19:46:59 - bayes.NaiveBayesMultinomialText
19:47:04 - bayes.NaiveBayesMultinomialText
19:48:07 - functions.SMO
19:49:11 - trees.HoeffdingTree
19:49:20 - trees.HoeffdingTree
19:49:24 - trees.HoeffdingTree
19:49:28 - trees.HoeffdingTree
19:51:58 - trees.REPTree
20:47:11 - functions.MultilayerPerceptron
20:47:55 - functions.MultilayerPerceptron
20:48:11 - functions.MultilayerPerceptron
20:48:25 - functions.MultilayerPerceptron
20:48:56 - functions.MultilayerPerceptron
20:49:22 - functions.MultilayerPerceptron
20:57:07 - functions.MultilayerPerceptron

Classifier output

Attrib PROCESS=monitor 0.08829930874919698
Attrib PROCESS=ndmpd 0.0036504741945590113
Attrib PROCESS=Logrotate 0.08763506866974166
Attrib PROCESS=gdm 0.040382284987981396
Attrib PROCESS=klogind 0.034131835687536944
Attrib PROCESS=shutdown 0.018926431696239911
Attrib PROCESS=FreeWnn 0.08037490365747429
Attrib PROCESS=cannerver 0.051953790545390949
Attrib PROCESS=last 0.016915150732850632
Attrib PROCESS=passwd 0.02470280296123396
Attrib PROCESS=wall 0.065969674266873599
Attrib PROCESS=fingerd 0.09007716812929337
Attrib PROCESS=in 0.09150187517989697
Attrib PROCESS=rshd 0.0900780285880985224
Attrib PROCESS=flogin 0.07036775104545434
Attrib PROCESS=modprobe 0.020206739731949256
Attrib PROCESS=net 0.08025703750995031
Attrib PROCESS=PAM 0.02373671307804558
Attrib PROCESS=unix_chkpwd 0.04447707154418355

Sigmoid Node 46

Inputs Weights
Threshold 0.04280732331727072

Class NON-ATTACK

Input
Node 0

Class ATTACK

Input
Node 1

Time taken to build model: 107.72 seconds

== Evaluation on training set ==

Time taken to test model on training data: 1.89 seconds

== Summary ==

Correctly Classified Instances	64554	90.0285 %
Incorrectly Classified Instances	7150	9.9715 %
Kappa statistic	0.7915	
Mean absolute error	0.1372	
Root mean squared error	0.2768	
Relative absolute error	30.799 %	
Root relative squared error	58.6488 %	
Total Number of Instances	71704	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.850	0.000	1.000	0.850	0.919	0.809	0.940	0.974	NON-ATTACK
	1.000	0.150	0.770	1.000	0.870	0.809	0.940	0.838	ATTACK

== Confusion Matrix ==

		a b	← classified as
a	b	40551 7150	a = NON-ATTACK
0	24003	0	b = ATTACK

Status

Results

Decision Tree - J48

Weka Workbench

Program Preprocess Classify Cluster Associate Select attributes Visualize Experiment Data mining processes Simple CLI

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) SUSPICIOUS

Start Stop

Result list (right-click for options)

11:10:11 - xtree_346

```
PROCESS = wod: NON-ATTACK (22.01)
PROCESS = identd: NON-ATTACK (24.01)
PROCESS = rsyncd: NON-ATTACK (38.02)
PROCESS = telnetd: NON-ATTACK (45.66)
PROCESS = httpd: NON-ATTACK (179.79/4515.79)
PROCESS = loginet: NON-ATTACK (440.09)
PROCESS = ftp: ATTACK (736.41/0.41)
PROCESS = Font: NON-ATTACK (7.0)
PROCESS = ddi: NON-ATTACK (30.02)
PROCESS = exiting: NON-ATTACK (35.01)
PROCESS = rsh: NON-ATTACK (324.20)
PROCESS = samba: ATTACK (544.3/0.3)
PROCESS = adduser: NON-ATTACK (4.0)
PROCESS = ircblanlone: NON-ATTACK (20.01)
PROCESS = rpcidmapd: NON-ATTACK (19.01)
PROCESS = httpd: NON-ATTACK (38.02)
PROCESS = hcid: NON-ATTACK (19.01)
PROCESS = aspd: NON-ATTACK (19.01)
PROCESS = smarid: NON-ATTACK (214.12)
PROCESS = cups: NON-ATTACK (194.11)
PROCESS = spamsassini: NON-ATTACK (28.02)
PROCESS = proxmox: NON-ATTACK (19.01)
PROCESS = savor: NON-ATTACK (1443.64)
PROCESS = IJimi: NON-ATTACK (19.01)
PROCESS = http_server: NON-ATTACK (19.01)
PROCESS = canna: NON-ATTACK (38.02)
PROCESS = retheaded: NON-ATTACK (19.01)
PROCESS = messagebus: NON-ATTACK (37.02)
PROCESS = mDNSResponder: NON-ATTACK (19.01)
PROCESS = mdmonitrc: NON-ATTACK (19.01)
PROCESS = mdmpd: NON-ATTACK (19.01)
PROCESS = logrotate: NON-ATTACK (530.3)
PROCESS = gdm: NON-ATTACK (57.01)
PROCESS = klogd: NON-ATTACK (146.09)
PROCESS = cronand: NON-ATTACK (6.0)
PROCESS = FreeRdp: NON-ATTACK (8.0)
PROCESS = cannaexecv: NON-ATTACK (8.0)
PROCESS = last: NON-ATTACK (33.02)
PROCESS = passwd: NON-ATTACK (12.01)
PROCESS = ntp: NON-ATTACK (1.0)
PROCESS = fingerd: NON-ATTACK (2.0)
PROCESS = in: NON-ATTACK (3.0)
PROCESS = rshd: NON-ATTACK (2.0)
PROCESS = clogind: NON-ATTACK (2.0)
PROCESS = modprobe: NON-ATTACK (4.0)
PROCESS = cron: NON-ATTACK (1.0)
PROCESS = RAM: NON-ATTACK (2.0)
PROCESS = unix_chipped: NON-ATTACK (306.17)

Number of Leaves : 81
Size of the tree : 62

Time taken to build model: 0.17 seconds
--- Evaluation on test split ---
Time taken to test model on test split: 0.23 seconds
--- Summary ---
Correctly Classified Instances 22020 90.3236 %
Incorrecly Classified Instances 2358 9.6764 %
Kappa statistic 0.7975
Mean absolute error 0.1498
Root mean squared error 0.2721
Relative absolute error 33.433 %
Root relative squared error 57.6127 %
Total Number of Instances 24379

--- Detailed Accuracy By Class ---

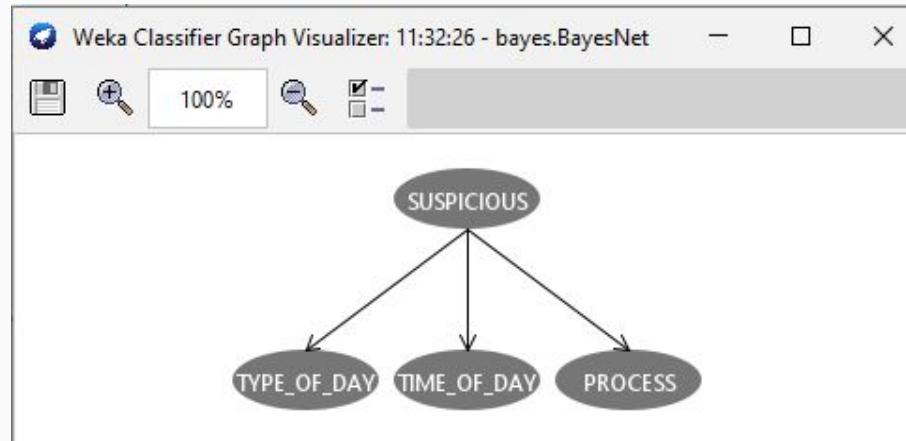
```

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	Roc Area	Class
0.854	0.000	1.000	0.854	0.921	0.814	0.935	0.962	NON-ATTACK
1.000	0.148	0.776	1.000	0.874	0.814	0.938	0.964	ATTACK
Weighted Avg.	0.903	0.049	0.925	0.903	0.906	0.914	0.935	0.909

Status OK Log x0

Results

Graph



Results

Naive Bayes results

Weka Workbench

Classifier: NaiveBayes

Test options:

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

More options...

(Nom) SUSPICIOUS

Start Stop

Result list (right-click for options)

11:10:11 - trees.748
11:25:50 - hayes.NaiveBayes

input	error	run
ab	38.0	1.0
ucd	23.0	1.0
identd	25.0	1.0
raymond	39.0	1.0
telnetd	1191.0	1.0
tcpdump	4889.0	1300.0
login	141.0	1.0
ftp	1.0	737.0
Font	8.0	1.0
dd	31.0	1.0
existing	18.0	1.0
www	284.0	1.0
sema	1.0	545.0
adduser	5.0	1.0
irgbalance	21.0	1.0
rpcidmpd	20.0	1.0
bluetooth	39.0	1.0
hcid	20.0	1.0
sdpd	20.0	1.0
smartd	215.0	1.0
cups	195.0	1.0
spansassassin	28.0	1.0
proxy	47.0	1.0
udev	1144.0	1.0
flim	20.0	1.0
http_server	20.0	1.0
canna	29.0	1.0
resolvconf	20.0	1.0
messagebus	28.0	1.0
mysqld	35.0	1.0
mdmonitor	20.0	1.0
ndmpd	20.0	1.0
logrotate	531.0	1.0
ym	55.0	1.0
klogind	147.0	1.0
shutdown	9.0	1.0
FreeWin	9.0	1.0
canaserver	9.0	1.0
last	34.0	1.0
passwd	15.0	1.0
wall	3.0	1.0
fingerd	3.0	1.0
in	3.0	1.0
rshd	3.0	1.0
logind	3.0	1.0
modprobe	7.0	1.0
net	2.0	1.0
FNM	3.0	1.0
unix_chkpwd	307.0	1.0
[total]	47742.0	24084.0

Time taken to build model: 0.04 seconds

*** Evaluation on test split ***

Time taken to test model on test split: 0.23 seconds

*** Summary ***

Correctly Classified Instances	22020	90.3236 %
Incorrectly Classified Instances	2359	9.6764 %
Mappa statistic	0.7975	
Mean absolute error	0.149	
Root mean squared error	0.2716	
Relative absolute error	33.4535 %	
Root relative squared error	01.5114 %	
Total Number of Instances	24779	

*** Detailed Accuracy By Class ***

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	FRC Area	Class
a	0.854	0.000	1.000	0.854	0.921	0.814	0.941	0.974	NON-ATTACK
b	1.000	0.146	0.776	1.000	0.874	0.814	0.941	0.836	ATTACK
Weighted Avg.	0.903	0.049	0.925	0.903	0.906	0.814	0.941	0.928	

*** Confusion Matrix ***

		a <- classified as	
		a = NON-ATTACK	b = ATTACK
a	18536	2359	1
b	0	8184	1

Status: OK

Log: x 0

Program Weka Workbench

Preprocess Classify Cluster Associate Select attributes Visualize Experiment Data mining processes Simple CLI

Classifier
Choose BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5

Test options

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

Classifier output

```
== Run information ==
Scheme: weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5
Relation: WEKA DATASET_V2
Instances: 71704
Attributes: 4
TYPE_OF_DAY
TIME_OF_DAY
PROCESS
SUSPICIOUS
Test mode: split 66.0% train, remainder test

== Classifier model (full training set) ==
Bayes Network Classifier
not using ADTree
#attributes=4 #classindex=3
Network structure (nodes followed by parents)
TYPE_OF_DAY(2): SUSPICIOUS
TIME_OF_DAY(5): SUSPICIOUS
PROCESS(1): SUSPICIOUS
SUSPICIOUS(2):
LogScore Bayes: -297327.6874413444
LogScore BDeu: -298292.33020662266
LogScore MDL: -298388.2584071655
LogScore ENTROPY: -297432.34260215075
LogScore AIC: -297603.34260215075

Time taken to build model: 0.19 seconds

== Evaluation on test split ==

Time taken to test model on test split: 0.11 seconds

== Summary ==
Correctly Classified Instances 22020 90.3236 %
Incorrectly Classified Instances 2359 9.6764 %
Kappa statistic 0.7975
Mean absolute error 0.1482
Root mean squared error 0.2714
Relative absolute error 33.2675 %
Root relative squared error 57.4789 %
Total Number of Instances 24379

== Detailed Accuracy By Class ==

```

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.854	0.000	1.000	0.854	0.921	0.814	0.941	0.974	NON-ATTACK
1.000	0.146	0.776	1.000	0.874	0.814	0.941	0.936	ATTACK
Weighted Avg.	0.903	0.049	0.925	0.903	0.906	0.914	0.941	0.928

```
== Confusion Matrix ==
a   b   <-- classified as
13836 2359 |   a = NON-ATTACK
0 8184 |   b = ATTACK
```

Status OK Log

Results

Bayes Net results

Results

SMO Results

Choose SMO -> C 1.0-L 0.001-P 1.0E-12-N 0-V 1-W 1-K 'weka.classifiers.functions.supportVector.PolyKernel'-E 1.0-C 250007->calibrator->weka.classifiers.functions.Logistic-R 1.0E-8-M 1-num-decimal-places 4"

Classifier output

```
+ -0.1243 * (normalized) PROCESS=cmd
+ -0.1243 * (normalized) PROCESS=ls
+ -0.1238 * (normalized) PROCESS=syncd
+ -0.1244 * (normalized) PROCESS=selected
1.9772 * (normalized) PROCESS=ftp
+ -0.1233 * (normalized) PROCESS=login
+ 1.9772 * (normalized) PROCESS=ls
+ -0.124 * (normalized) PROCESS=ps
+ -0.1233 * (normalized) PROCESS=ad
+ -0.1239 * (normalized) PROCESS=existing
+ -0.1225 * (normalized) PROCESS=eu
1.9772 * (normalized) PROCESS=webc
+ -0.1239 * (normalized) PROCESS=background
+ -0.1239 * (normalized) PROCESS=imbalance
+ -0.1238 * (normalized) PROCESS=pcmcia
+ -0.1238 * (normalized) PROCESS=bluetooth
-0.1235 * (normalized) PROCESS=child
+ -0.1244 * (normalized) PROCESS=script
+ -0.1244 * (normalized) PROCESS=smard
+ -0.123 * (normalized) PROCESS=cpu
+ -0.1238 * (normalized) PROCESS=spamassassin
+ -0.1241 * (normalized) PROCESS=privoxy
+ -0.1238 * (normalized) PROCESS=proxy
+ -0.1228 * (normalized) PROCESS=rlm
+ -0.1239 * (normalized) PROCESS=http_server
+ -0.1241 * (normalized) PROCESS=cname
+ -0.1238 * (normalized) PROCESS=needhead
+ -0.1232 * (normalized) PROCESS=imapd
+ -0.1237 * (normalized) PROCESS=pgsql
+ -0.1244 * (normalized) PROCESS=monitor
+ -0.123 * (normalized) PROCESS=mdpd
+ -0.1228 * (normalized) PROCESS=logrotate
+ -0.1231 * (normalized) PROCESS=telnetd
+ -0.1232 * (normalized) PROCESS=backgroundd
+ -0.1236 * (normalized) PROCESS=bind
+ -0.1238 * (normalized) PROCESS=freedns
+ -0.1238 * (normalized) PROCESS=FreeRn
+ -0.1241 * (normalized) PROCESS=cannaserver
+ -0.1242 * (normalized) PROCESS=last
+ -0.1238 * (normalized) PROCESS=scriptd
+ -0.1239 * (normalized) PROCESS=mail
+ -0.1232 * (normalized) PROCESS=tinderd
+ -0.1211 * (normalized) PROCESS=in
+ -0.1211 * (normalized) PROCESS=rshd
+ -0.1211 * (normalized) PROCESS=elogind
+ -0.1238 * (normalized) PROCESS=sshd
+ -0.1232 * (normalized) PROCESS=net
+ -0.122 * (normalized) PROCESS=fam
+ -0.1228 * (normalized) PROCESS=unix_chkpwd
- 0.8769
```

Number of kernel evaluations: 805055499 (28.778% cached)

Time taken to build model: 362.01 seconds

--- Evaluation on test split ---

Time taken to test model on test split: 0.19 seconds

--- Summary ---

	22020	90.3236 %
Correctly Classified Instances	22020	90.3236 %
Incorrectly Classified Instances	2359	9.6764 %
Kappa statistic	0.975	
Mean absolute error	0.0005	
Root mean squared error	0.3111	
Relative absolute error	21.102 %	
Root relative squared error	65.8714 %	
Total Number of Instances	24379	

--- Detailed Accuracy By Class ---

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.854	0.000	1.000	0.854	0.921	0.814	0.927	0.951	0.951	NON-ATTACK
1.000	0.146	0.776	1.000	0.874	0.814	0.927	0.776	0.776	ATTACK
Weighted Avg.	0.903	0.049	0.925	0.903	0.906	0.914	0.927	0.927	

--- Confusion Matrix ---

	a	b	<- classified as
a	13856	2359	a = NON-ATTACK
b	0	8184	b = ATTACK

Status OK Log x 0

Results

Deep learning - Multilayer Perceptron results

Program Weka Workbench

Preprocess Classify Cluster Associate Select attribute Visualize Experiment Data mining processes Simple CLI

Classifier
Choose **MultilayerPerceptron - L: 0.3 - M: 0.2 - N: 500 - V: 0 - E: 20 - H: 6 - R**

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

(Nom) SUSPICIOUS
Start Stop
Result list (right-click for options)
11:25:50 - Hayes_HouseDays
11:12:12x - Hayes_Baptists
12:12x5 - functions_SBD
12:41x12 - functions_MultilayerPerceptron

Class output
Predicted PROCESS-execd 0.005100773114424891
Attrib PROCESS-execd 0.0354791593339753
Attrib PROCESS-date 0.018075025054548764
Attrib PROCESS-neefs -0.0173850941237219
Attrib PROCESS-fack 0.02970181560319918
Attrib PROCESS-end 0.0344468368351576
Attrib PROCESS-init 0.0344468368351576
Attrib PROCESS-indus -0.0374946206122455
Attrib PROCESS-network -0.03896497216077
Attrib PROCESS-autocomm -0.02984552352483
Attrib PROCESS-geoflow 0.0344468368351576
Attrib PROCESS-geoflow 0.0344468368351576
Attrib PROCESS-andh 0.021142310073079434
Attrib PROCESS-xinetd 0.0478395059072736
Attrib PROCESS-sendmail -0.015338518087176939
Attrib PROCESS-ypd 0.02984552352483
Attrib PROCESS-ypcd 0.0123399736017575
Attrib PROCESS-efds -0.011028752877040008
Attrib PROCESS-anacrc 0.024695321113951583
Attrib PROCESS-htpd -0.0264695321113951583
Attrib PROCESS-nginx 0.024695321113951583
Attrib PROCESS-nginx 0.024424205442196
Attrib PROCESS-ampd 0.00527070926707544
Attrib PROCESS-mpd 0.01055939741796609
Attrib PROCESS-mpd -0.01417409630840311
Attrib PROCESS-mpd 0.012894776282725
Attrib PROCESS-identd 0.02894776282725
Attrib PROCESS-sshd 0.02894776282725
Attrib PROCESS-sshd -0.0284535689170225704
Attrib PROCESS-telnetd 0.01830037475294107
Attrib PROCESS-ftp 0.024030037475294107
Attrib PROCESS-ftp 0.024662232176674745
Attrib PROCESS-ftp 0.02305443344594516
Attrib PROCESS-dd 0.00461033601280917
Attrib PROCESS-exit 0.046200177117246
Attrib PROCESS-ex 0.005114033382021144
Attrib PROCESS-ppc 0.0273029116211499
Attrib PROCESS-adducte 0.04072337113200776
Attrib PROCESS-lrpalance 0.0248528080342458
Attrib PROCESS-rcpidmapd 0.0021036035546052475
Attrib PROCESS-rcpidmapd 0.0248528080342458
Attrib PROCESS-hcid 0.0248528080342458
Attrib PROCESS-wqpd 0.0104561635509581
Attrib PROCESS-wqpd -0.04461420185634224
Attrib PROCESS-cups 0.01913797338137616
Attrib PROCESS-cups 0.01913797338137616
Attrib PROCESS-privacy 0.036521793031030
Attrib PROCESS-edev -0.02213343456050704
Attrib PROCESS-llm -0.0452649214912958
Attrib PROCESS-htr_server 0.00591109809748279
Attrib PROCESS-ctcdeauth -0.009723538477555554
Attrib PROCESS-ctcdeauth -0.00972353847755554
Attrib PROCESS-nemsgbus 0.01213447531519346
Attrib PROCESS-mysqld 0.04407923466030017
Attrib PROCESS-monitor 0.04284401197361744
Attrib PROCESS-mpmd -0.041412124485986
Attrib PROCESS-mpmd -0.041412124485986
Attrib PROCESS-pdm -0.02836674595659214
Attrib PROCESS-klogind -0.004579049536374016
Attrib PROCESS-shutdown -0.02730291169348359
Attrib PROCESS-exeewer 0.007153926012953
Attrib PROCESS-slart -0.02732390107243958
Attrib PROCESS-passw -0.020175732320342017
Attrib PROCESS-mail 0.0237993274372904
Attrib PROCESS-fingerd 0.044463524461960503
Attrib PROCESS-mail 0.044463524461960503
Attrib PROCESS-wndh 0.044463524461960503
Attrib PROCESS-wndh 0.044463524461960503
Attrib PROCESS-elogind 0.02424153917513898
Attrib PROCESS-madprobe -0.025004753693086947
Attrib PROCESS-ssh 0.0395245097424724
Attrib PROCESS-pAM -0.022854539335555
Attrib PROCESS-mail_chkdq 0.01259032327933768

Sigmoid Node 46
Inputs Weights
Threshold 0.04280732331727072
Class ATTACK
Input
Node 0
Class ATTACK
Input
Node 1

Time taken to build model: 55.15 seconds

Log x