BHASKARACHARYA NATIONAL INSTITUTE FOR SPACE APPLICATIONS AND GEO-INFORMATICS

WEEKLY PROGRESS REPORT (27/02/2023 – 05/03/2023)

WEEK 6

PROJECT NAME	MALWARE DETECTION USING ML

PROJECT DESCRIPTION :	DESIGN AND IMPLEMENT ML MODEL TO DETECT MALWARE IN SYSTEM
GROUP MEMBER :	PRATHAM PATEL, Shubham Patel, Shashank Sharma, Yash Soni
GROUP ID :	I2
GROUP GUIDE :	HARSH KIRATSATA
GROUP COORDINATOR :	SIDHDHARTH PATEL
COLLEGE GUIDE NAME :	MAYANK PATEL, RITIKA LADHA, KINJAL CHAUDHARI
COLLEGE GUIDE No. :	+91 9033280445,8866438698, 9723173567
COLLEGE GUIDE EMAIL. :	mayank.patel@adaniuni.ac.in, ritika.ladha@adaniuni.ac.in, kinjal.chaudhari@adaniuni.ac.in
COLLEGE NAME :	ADANI INSTITUTE OF INFRASTRUCTURE AND ENGINEERING

27/02/2023 TILL 05/03/2023 (7 DAYS)

Using weka for ML and data visualization.

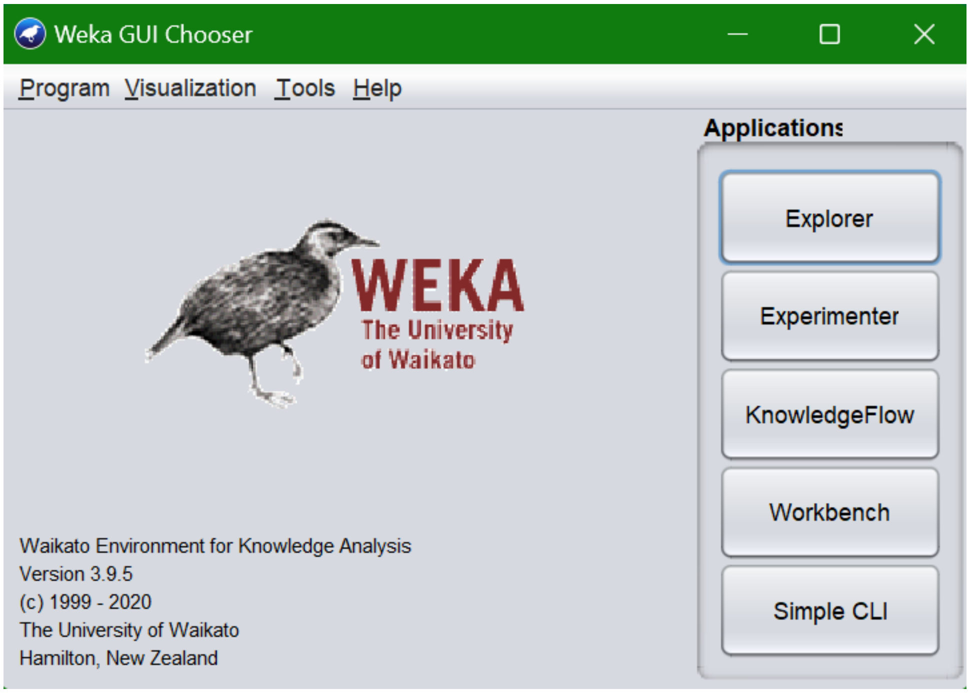
27/02/2023	Collecting real data from cuckoo sandbox.
28/02/2023	Exploring the dataset manually.
01/03/2023	Exploring Weka dashboard and its functions.
02/03/2023	Classifiers selection, data attributes changes.
03/03/2023	Using cross validation for feature selection.
04/03/2023	Getting insights for weka data visualization and features.
05/03/2023	(Holiday Sunday)

WEEK 6(PLAN)	We are planning to explore more ML algorithms in next week in order to fine accurate and perfect result.

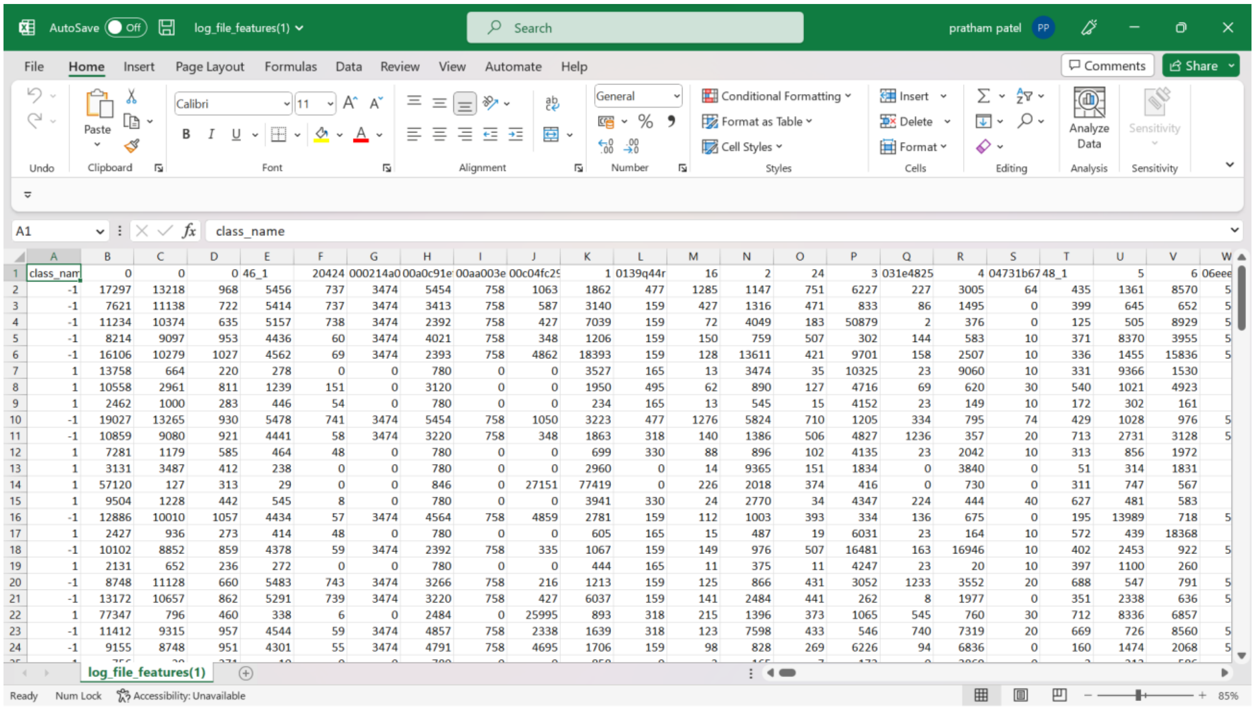
REFERENCE :

- https://github.com/topics/cuckoo-sandbox?o=desc&s=updated
- https://www.cs.waikato.ac.nz/ml/weka/
- https://www.weka.io/
- https://www.youtube.com/watch?v=XaEWGijbDOU&feature=youtu.be

Screenshots :

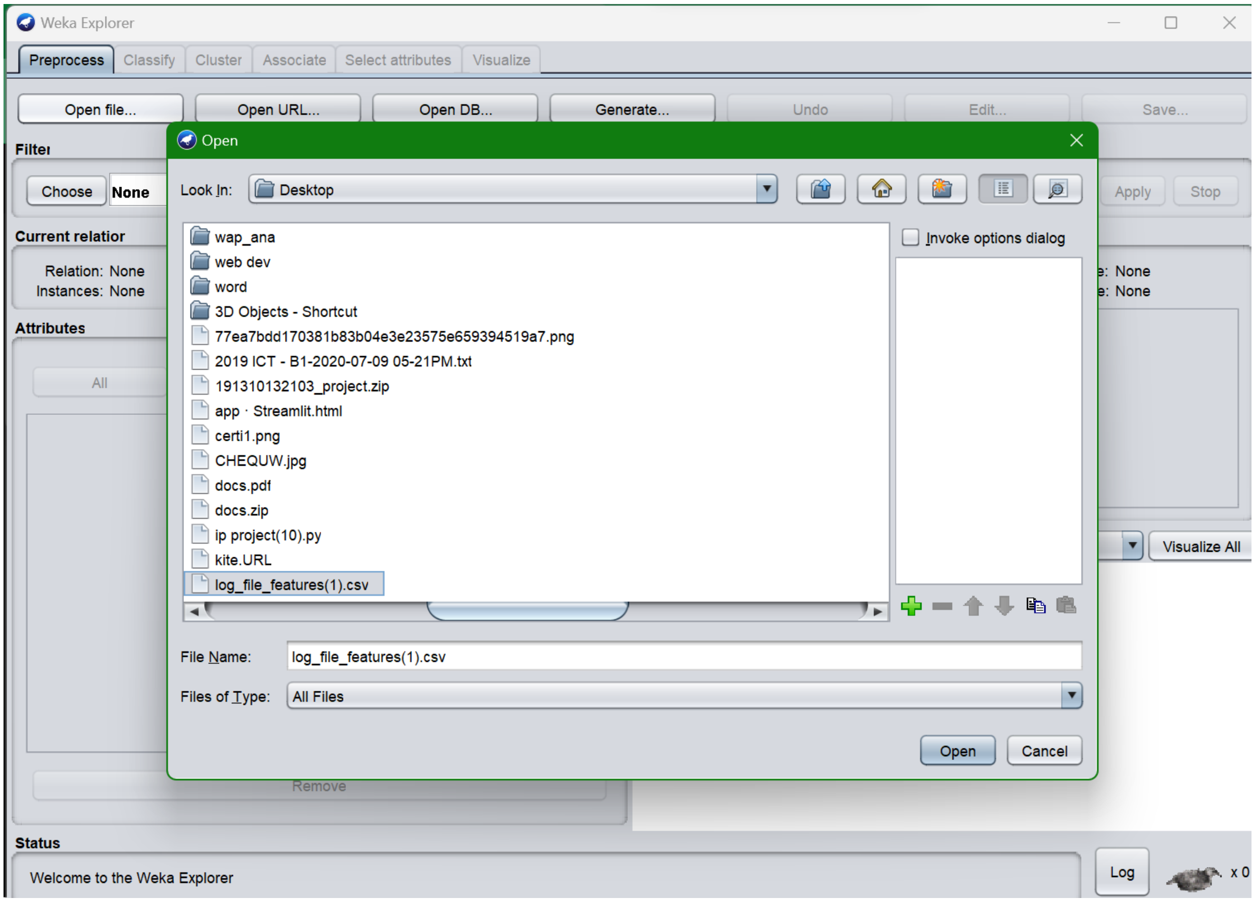


Waikato Environment for Knowledge Analysis
Version 3.9.5
(c) 1999 - 2020
The University of Waikato
Hamilton, New Zealand



A1: class_name

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	class_nam	0	0	0	46_1	20424	000214a0	00a0c91e	00aa003e	00c04fc29	1	0139q44r	16	2	24	3	031e4825	4	04731b67	48_1	5	6	06eee
2	-1	17297	13218	968	5456	737	3474	5454	758	1063	1862	477	1285	1147	751	6227	227	3005	64	435	1361	8570	5
3	-1	7621	11138	722	5414	737	3474	3413	758	587	3140	159	427	1316	471	833	86	1495	0	399	645	652	5
4	-1	11234	10374	635	5157	738	3474	2392	758	427	7039	159	72	4049	183	50879	2	376	0	125	505	8929	5
5	-1	8214	9097	953	4436	60	3474	4021	758	348	1206	159	150	759	507	302	144	583	10	371	8370	3955	5
6	-1	16106	10279	1027	4562	69	3474	2393	758	4862	18393	159	128	13611	421	9701	158	2507	10	336	1455	15836	5
7	1	13758	664	220	278	0	0	780	0	0	3527	165	13	3474	35	10325	23	9060	10	331	9366	1530	
8	1	10558	2961	811	1239	151	0	3120	0	0	1950	495	62	890	127	4716	69	620	30	540	1021	4923	
9	1	2462	1000	283	446	54	0	780	0	0	234	165	13	545	15	4152	23	149	10	172	302	161	
10	-1	19027	13265	930	5478	741	3474	5454	758	1050	3223	477	1276	5824	710	1205	334	795	74	429	1028	976	5
11	-1	10859	9080	921	4441	58	3474	3220	758	348	1863	318	140	1386	506	4827	1236	357	20	713	2731	3128	5
12	1	7281	1179	585	464	48	0	780	0	0	699	330	88	896	102	4135	23	2042	10	313	856	1972	
13	1	3131	3487	412	238	0	0	780	0	0	2960	0	14	9365	151	1834	0	3840	0	51	314	1831	
14	1	57120	127	313	29	0	0	846	0	27151	77419	0	226	2018	374	416	0	730	0	311	747	567	
15	1	9504	1228	442	545	8	0	780	0	0	3941	330	24	2770	34	4347	224	444	40	627	481	583	
16	-1	12886	10010	1057	4434	57	3474	4564	758	4859	2781	159	112	1003	393	334	136	675	0	195	13989	718	5
17	1	2427	936	273	414	48	0	780	0	0	605	165	15	487	19	6031	23	164	10	572	439	18368	
18	-1	10102	8852	859	4378	59	3474	2392	758	335	1067	159	149	976	507	16481	163	16946	10	402	2453	922	5
19	1	2131	652	236	272	0	0	780	0	0	444	165	11	375	11	4247	23	20	10	397	1100	260	
20	-1	8748	11128	660	5483	743	3474	3266	758	216	1213	159	125	866	431	3052	1233	3552	20	688	547	791	5
21	-1	13172	10657	862	5291	739	3474	3220	758	427	6037	159	141	2484	441	262	8	1977	0	351	2338	636	5
22	1	77347	796	460	338	6	0	2484	0	25995	893	318	215	1396	373	1065	545	760	30	712	8336	6857	
23	-1	11412	9315	957	4544	59	3474	4857	758	2338	1639	318	123	7598	433	546	740	7319	20	669	726	8560	5
24	-1	9155	8748	951	4301	55	3474	4791	758	4695	1706	159	98	828	269	6226	94	6836	0	160	1474	2068	5



Filter: None

Relation: None
Instances: None

Open

Look In: Desktop

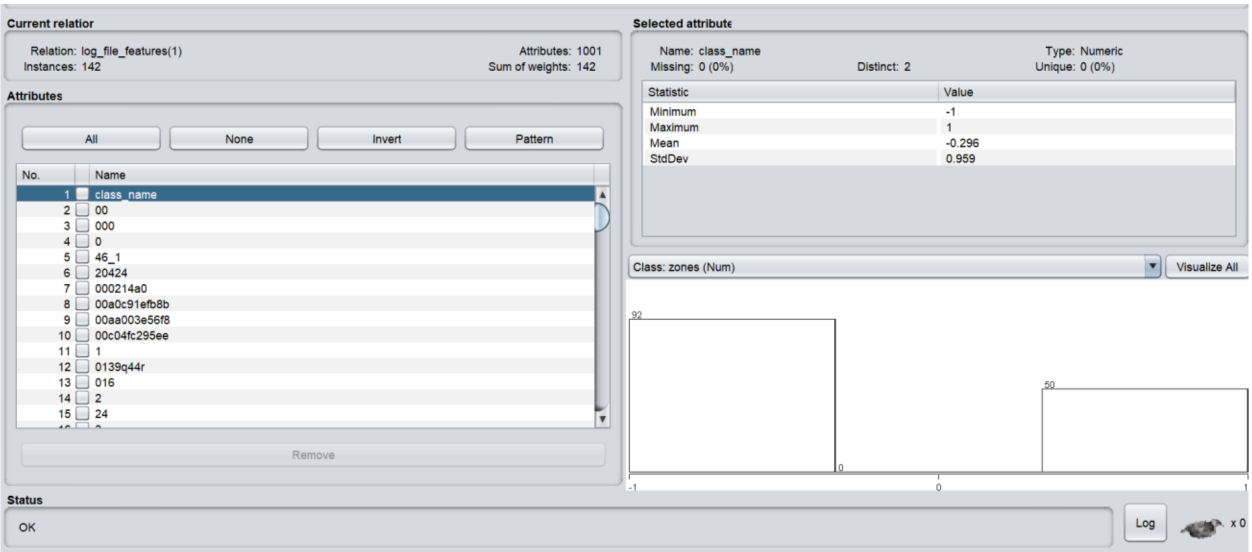
- wap_ana
- web dev
- word
- 3D Objects - Shortcut
- 77ea7bdd170381b83b04e3e23575e659394519a7.png
- 2019 ICT - B1-2020-07-09 05-21PM.txt
- 191310132103_project.zip
- app · Streamlit.html
- certi1.png
- CHEQUW.jpg
- docs.pdf
- docs.zip
- ip project(10).py
- kite.URL
- log_file_features(1).csv

□ Invoke options dialog

File Name: log_file_features(1).csv

Files of Type: All Files

Status
Welcome to the Weka Explorer



Relation: log_file_features(1)
Instances: 142
Attributes: 1001
Sum of weights: 142

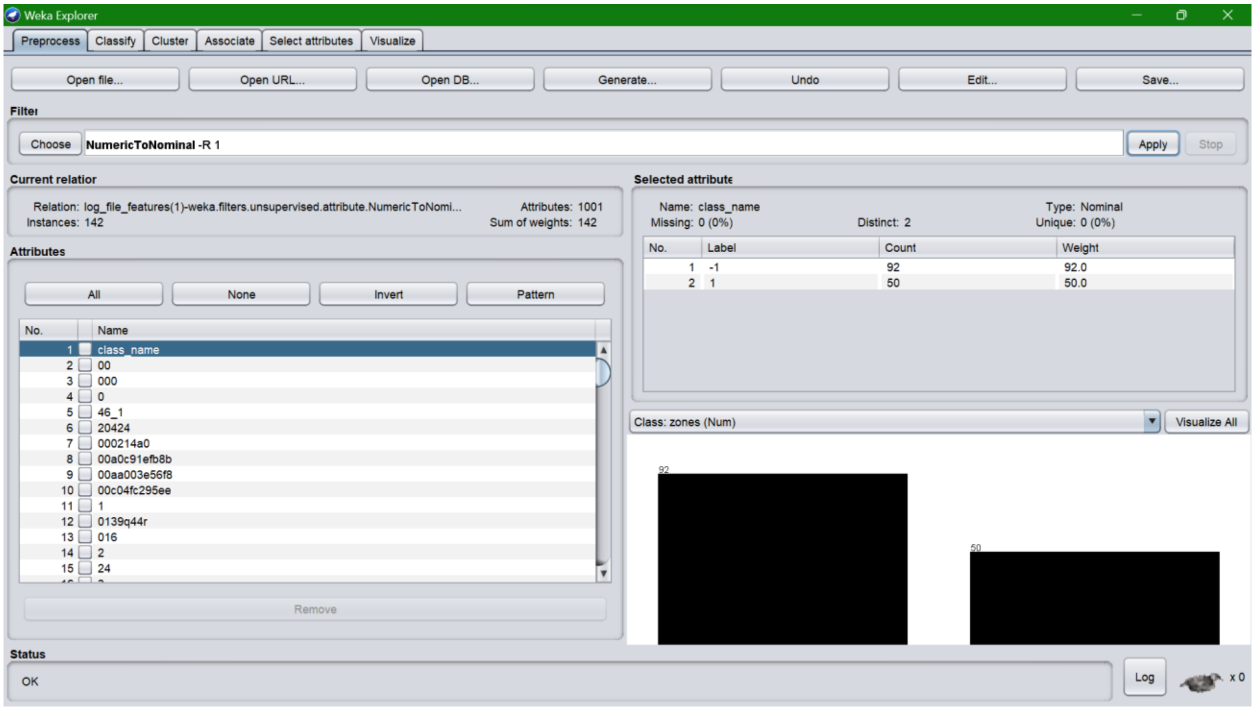
No.		Name
1	□	class_name
2	□	00
3	□	000
4	□	0
5	□	46_1
6	□	20424
7	□	000214a0
8	□	00a0c91efb8b
9	□	00aa003e56f8
10	□	00c04fc295ee
11	□	1
12	□	0139q44r
13	□	016
14	□	2
15	□	24

Name: class_name
Type: Numeric
Missing: 0 (0%)
Distinct: 2
Unique: 0 (0%)

Statistic	Value
Minimum	-1
Maximum	1
Mean	-0.296
StdDev	0.959

Class: zones (Num)

Status
OK



Filter

NumericToNominal -R 1

Current relation

Relation: log_file_features(1)-weka.filters.unsupervised.attribute.NumericToNomi... Attributes: 1001
Instances: 142 Sum of weights: 142

Attributes

No.		Name
1	□	class_name
2	□	00
3	□	000
4	□	0
5	□	46_1
6	□	20424
7	□	000214a0
8	□	00a0c91efb8b
9	□	00aa003e56f8
10	□	00c04fc295ee
11	□	1
12	□	0139q44r
13	□	016
14	□	2
15	□	24

Selected attribute

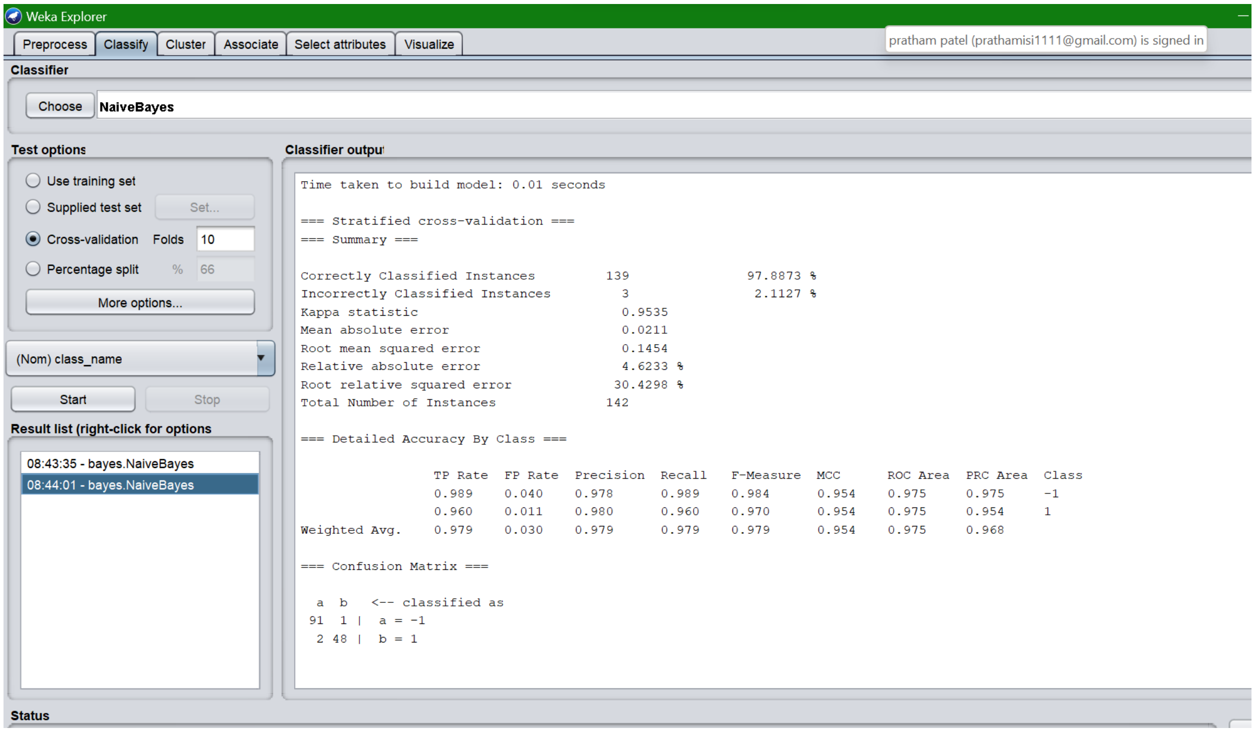
Name: class_name Type: Nominal
Missing: 0 (0%) Distinct: 2 Unique: 0 (0%)

No.	Label	Count	Weight
1	-1	92	92.0
2	1	50	50.0

Class: zones (Num)

Status

OK



Classifier

NaiveBayes

Test options

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

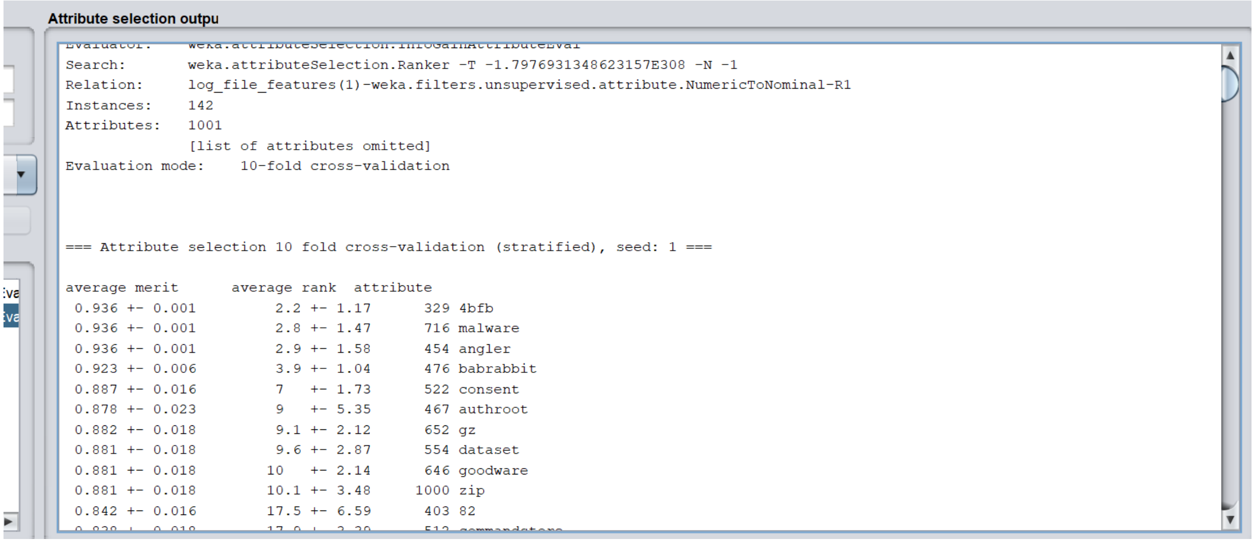
(Nom) class_name

Result list (right-click for options)

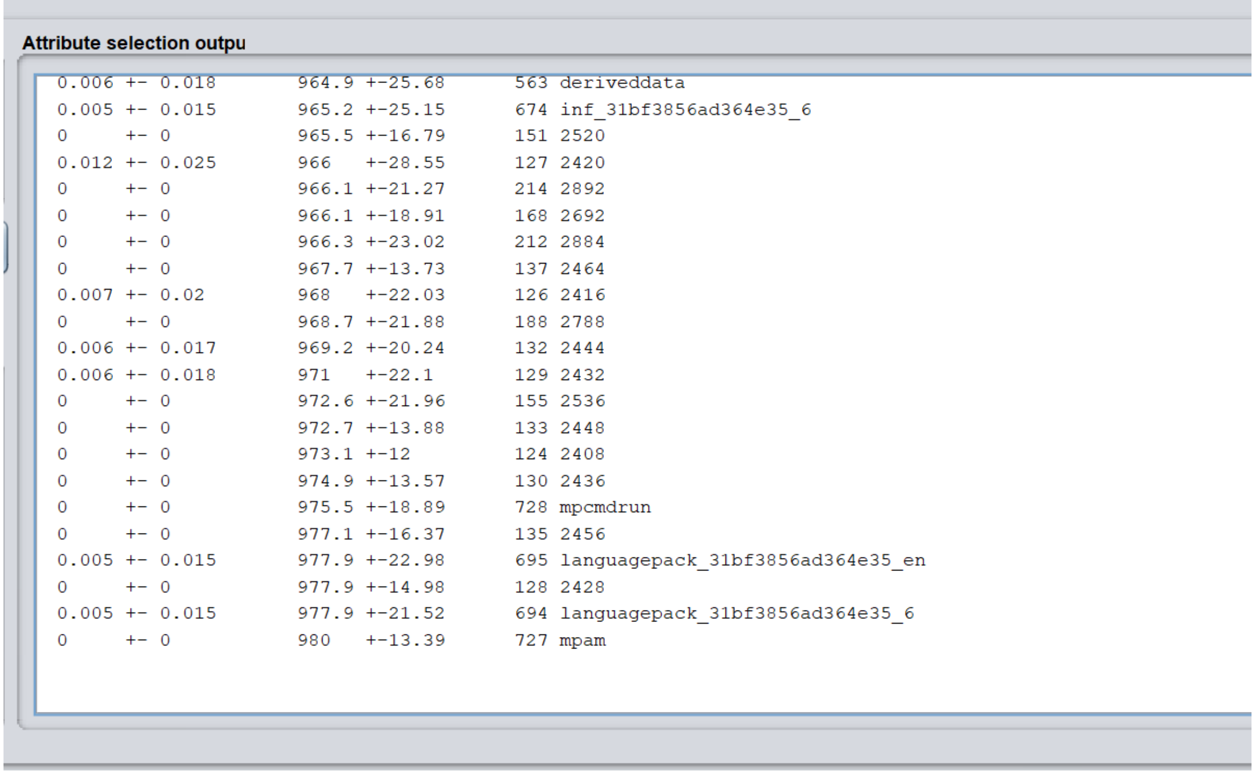
08:43:35 - bayes.NaiveBayes
08:44:01 - bayes.NaiveBayes

Classifier output

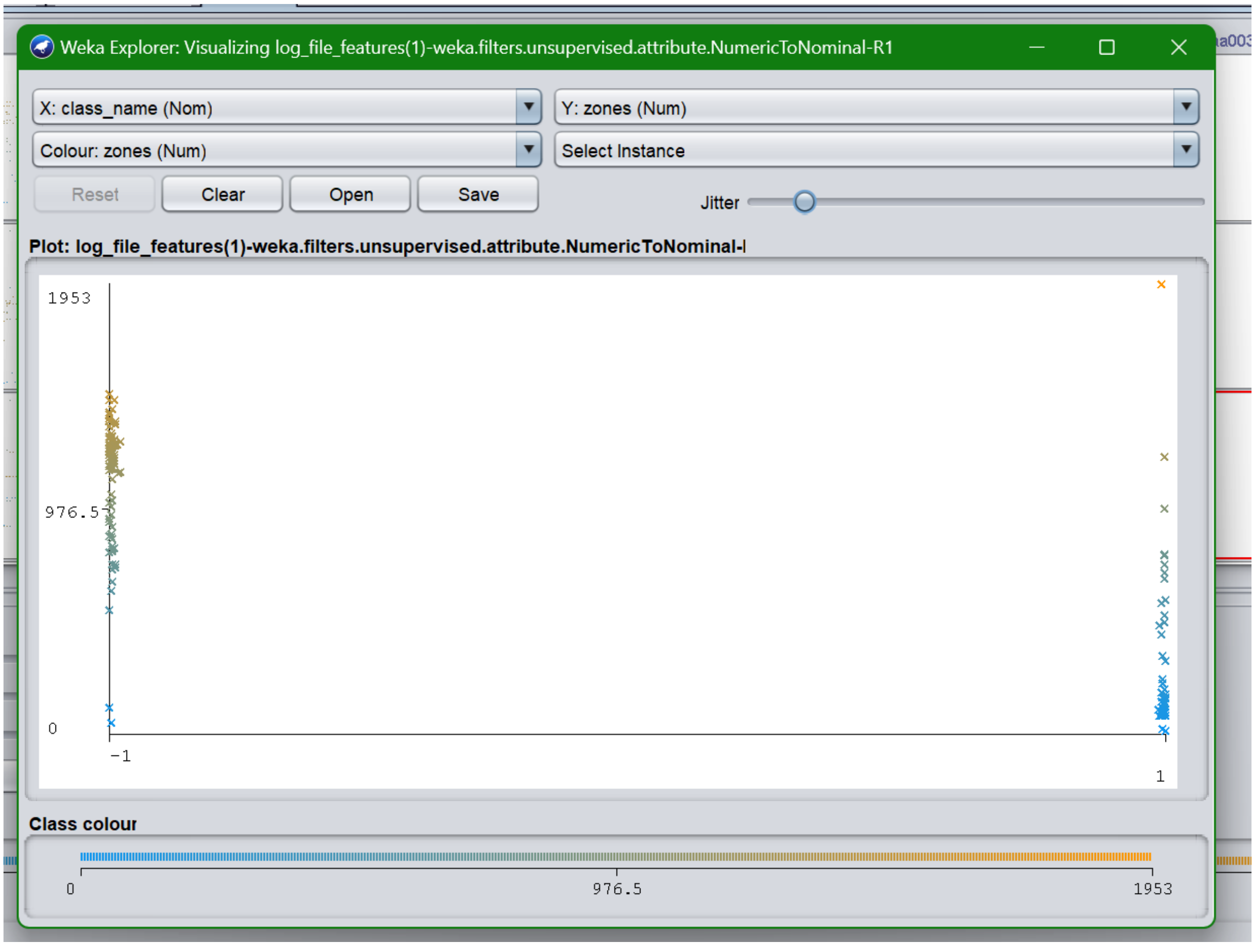
```
Time taken to build model: 0.01 seconds

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

Correctly Classified Instances         139               97.8873 %
Incorrectly Classified Instances         3                2.1127 %
Kappa statistic                          0.9535
Mean absolute error                      0.0211
Root mean squared error                  0.1454
Relative absolute error                  4.6233 %
Root relative squared error             30.4298 %
Total Number of Instances              142

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.989    0.040    0.978      0.989    0.984      0.954    0.975     0.975     -1
                 0.960    0.011    0.980      0.960    0.970      0.954    0.975     0.954     1
Weighted Avg.    0.979    0.030    0.979      0.979    0.979      0.954    0.975     0.968

=== Confusion Matrix ===

  a  b   <-- classified as
 91  1 |  a = -1
  2 48 |  b = 1
```



```
Evaluator:    weka.attributeSelection.InfoGainAttributeEval
Search:       weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation:     log_file_features(1)-weka.filters.unsupervised.attribute.NumericToNominal-R1
Instances:    142
Attributes:   1001
              [list of attributes omitted]
Evaluation mode:    10-fold cross-validation



=== Attribute selection 10 fold cross-validation (stratified), seed: 1 ===

average merit      average rank  attribute
 0.936 +- 0.001         2.2 +- 1.17      329 4bfb
 0.936 +- 0.001         2.8 +- 1.47      716 malware
 0.936 +- 0.001         2.9 +- 1.58      454 angler
 0.923 +- 0.006         3.9 +- 1.04      476 babrabbit
 0.887 +- 0.016         7   +- 1.73      522 consent
 0.878 +- 0.023         9   +- 5.35      467 authroot
 0.882 +- 0.018         9.1 +- 2.12      652 gz
 0.881 +- 0.018         9.6 +- 2.87      554 dataset
 0.881 +- 0.018        10   +- 2.14      646 goodware
 0.881 +- 0.018        10.1 +- 3.48     1000 zip
 0.842 +- 0.016        17.5 +- 6.59      403 82
 0.838 +- 0.018        17.9 +- 2.39      513 commandstore
```



```
0.006 +- 0.018        964.9 +-25.68      563 deriveddata
0.005 +- 0.015        965.2 +-25.15      674 inf_31bf3856ad364e35_6
0     +- 0            965.5 +-16.79      151 2520
0.012 +- 0.025        966   +-28.55      127 2420
0     +- 0            966.1 +-21.27      214 2892
0     +- 0            966.1 +-18.91      168 2692
0     +- 0            966.3 +-23.02      212 2884
0     +- 0            967.7 +-13.73      137 2464
0.007 +- 0.02         968   +-22.03      126 2416
0     +- 0            968.7 +-21.88      188 2788
0.006 +- 0.017        969.2 +-20.24      132 2444
0.006 +- 0.018        971   +-22.1       129 2432
0     +- 0            972.6 +-21.96      155 2536
0     +- 0            972.7 +-13.88      133 2448
0     +- 0            973.1 +-12         124 2408
0     +- 0            974.9 +-13.57      130 2436
0     +- 0            975.5 +-18.89      728 mpcmdrun
0     +- 0            977.1 +-16.37      135 2456
0.005 +- 0.015        977.9 +-22.98      695 languagepack_31bf3856ad364e35_en
0     +- 0            977.9 +-14.98      128 2428
0.005 +- 0.015        977.9 +-21.52      694 languagepack_31bf3856ad364e35_6
0     +- 0            980   +-13.39      727 mpam
```



Weka Explorer: Visualizing log_file_features(1)-weka.filters.unsupervised.attribute.NumericToNominal-R1

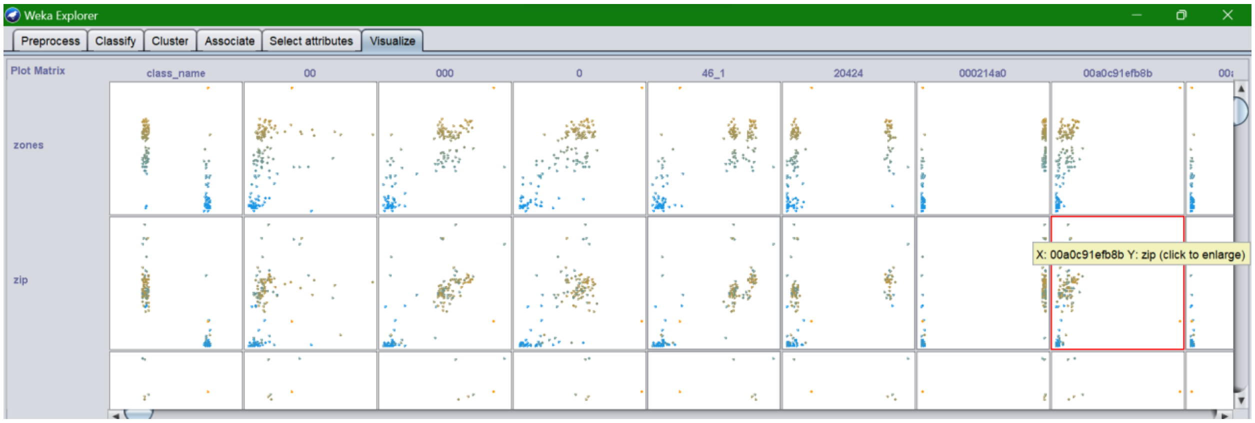
X: class_name (Nom)

Y: zones (Num)

Colour: zones (Num)

Select Instance

Plot: log_file_features(1)-weka.filters.unsupervised.attribute.NumericToNominal-I



Weka Explorer

X: 00a0c91efb8b Y: zip (click to enlarge)