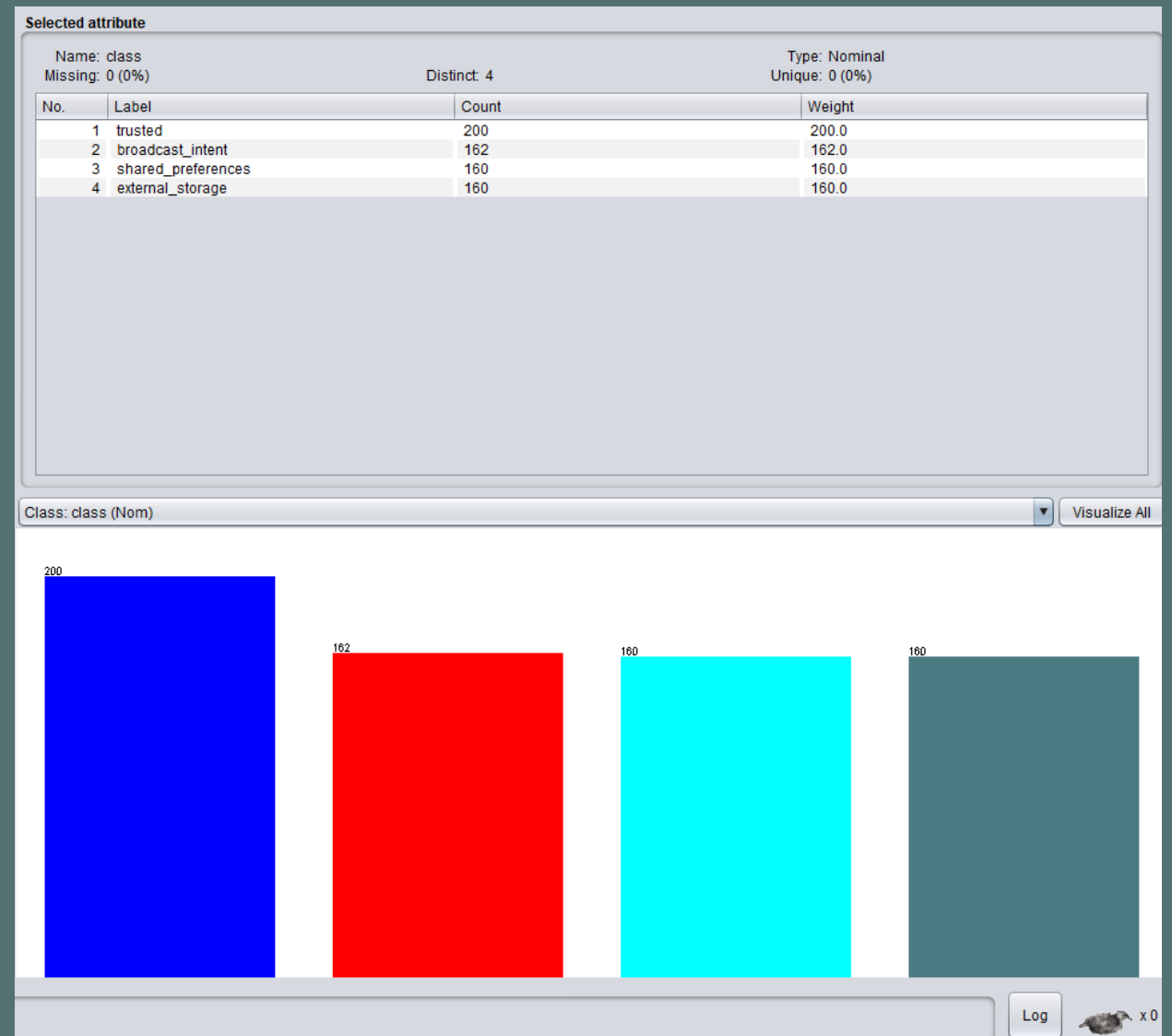


# OSSERVAZIONI DATASET CON 4 LABEL

*I due dataset contengono le features degli  
audio splittati.*

*Il primo dataset ha split di 2093 secondi  
il secondo dataset ha split di 1046  
secondi.*

*Le istanze delle bag sono quindi il doppio  
del primo dataset*



# L'unico algoritmo a dare risultati su una classe con 4 label e il TLC

Split 2093 secondi  
Circa 8 istanze per bag

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **TLC** -P "weka.classifiers.trees.J48" -C 0.25 -M 2 -W weka.classifiers.meta.LogitBoost -P 100 -L -1.7976931348623157E308 -H 1.0 -Z 3.0 -O 1 -E 1 -S 1 -I 10 -W weka.classifiers.trees.D

Test options

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

More options...

(Nom) class

Start Stop

Result list (right-click for options)

10:09:31 - mi.TLC

Classifier output

Class 4 (class=external\_storage)

Decision Stump

Classifications

partition\_153 <= 0.5 : 0.12987323536987894  
partition\_153 > 0.5 : -1.4308803877308735  
partition\_153 is missing : -0.01895950886820279

Number of performed iterations: 10

Time taken to build model: 0.99 seconds

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

Correctly Classified Instances	576	84.4575 %
Incorrectly Classified Instances	106	15.5425 %
Kappa statistic	0.792	
Mean absolute error	0.091	
Root mean squared error	0.2394	
Relative absolute error	24.3515 %	
Root relative squared error	55.3875 %	
Total Number of Instances	682	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.990	0.008	0.980	0.990	0.985	0.979	0.996	0.984	trusted
	0.790	0.058	0.810	0.790	0.800	0.739	0.960	0.902	broadcast_intent
	0.725	0.103	0.682	0.725	0.703	0.609	0.926	0.828	shared_preferences
	0.838	0.034	0.882	0.838	0.859	0.818	0.973	0.926	external_storage
Weighted Avg.	0.845	0.048	0.847	0.845	0.845	0.797	0.966	0.914	

=== Confusion Matrix ===

a	b	c	d	<-- classified as
198	0	2	0	a = trusted
0	128	31	3	b = broadcast_intent
2	27	116	15	c = shared_preferences
2	3	21	134	d = external_storage

Split 1046secondi  
Circa 16 istanze per bag

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize Auto-WEKA DI4J Inference CPython Scripting

Classifier

Choose **MISMO** -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.ml.supportVector.MIPolyKernel" -E 1.0 -C 250007"

Test options

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

More options...

(Nom) class

Start Stop

Result list (right-click for options)

09:47:58 - mi.CitationKNN  
09:49:22 - mi.TLC  
09:51:23 - mi.SimpleML  
09:51:26 - mi.MIWwrapper  
09:51:31 - mi.MISMO

Classifier output

Class 4 (class=external\_storage)

Decision Stump

Classifications

partition\_416 <= 0.5 : 0.17315703397504506  
partition\_416 > 0.5 : -1.3762688266565373  
partition\_416 is missing : -0.006754871740958355

Number of performed iterations: 10

Time taken to build model: 1.93 seconds

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

Correctly Classified Instances	576	84.4575 %
Incorrectly Classified Instances	106	15.5425 %
Kappa statistic	0.7921	
Mean absolute error	0.089	
Root mean squared error	0.2362	
Relative absolute error	23.8153 %	
Root relative squared error	54.6276 %	
Total Number of Instances	682	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.995	0.002	0.995	0.995	0.995	0.993	0.998	0.990	trusted
	0.840	0.063	0.805	0.840	0.822	0.765	0.968	0.915	broadcast_intent
	0.694	0.067	0.760	0.694	0.725	0.647	0.942	0.860	shared_preferences
	0.813	0.071	0.778	0.813	0.795	0.731	0.955	0.881	external_storage
Weighted Avg.	0.845	0.048	0.844	0.845	0.844	0.796	0.968	0.916	

=== Confusion Matrix ===

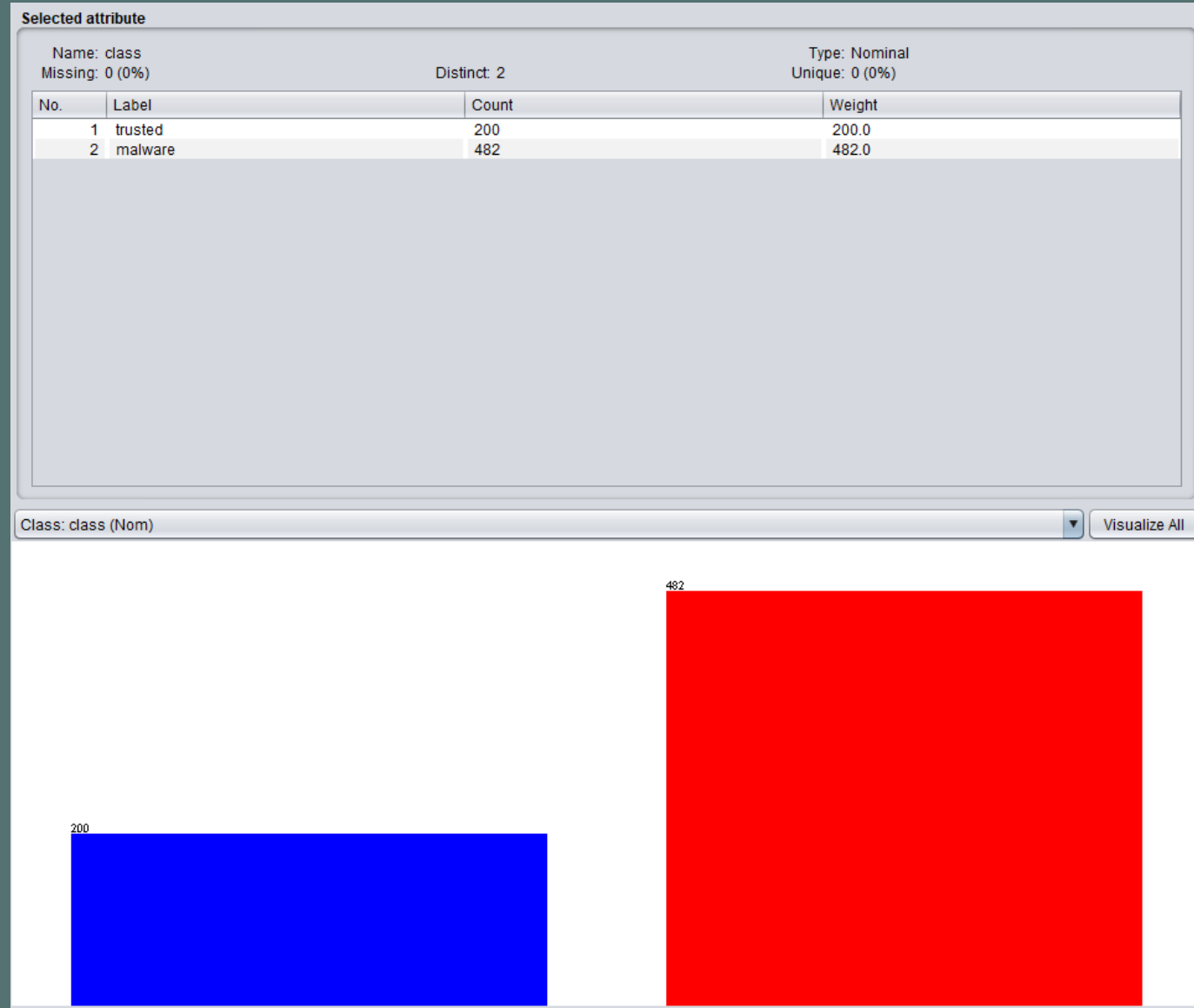
a	b	c	d	<-- classified as
199	0	0	1	a = trusted
0	136	16	10	b = broadcast_intent
0	23	111	26	c = shared_preferences
1	10	19	130	d = external_storage

# OSSERVAZIONI DATASET CON LABEL CLASS TRUSTED / MALWARE

*I due dataset contengono le features degli  
audio splittati.*

*Il primo dataset ha split di 2093 secondi  
il secondo dataset ha split di 1046  
secondi.*

*Le istanze delle bag sono quindi il doppio  
del primo dataset*



- Le istanze non sono bilanciate (200 **Trusted** e 482 **Malware**).
- I risultati migliori hanno valore di precision e recall coincidente.

Split 2093 secondi  
Circa 8 istanze per bag

Split 1046 secondi  
Circa 16 istanze per bag

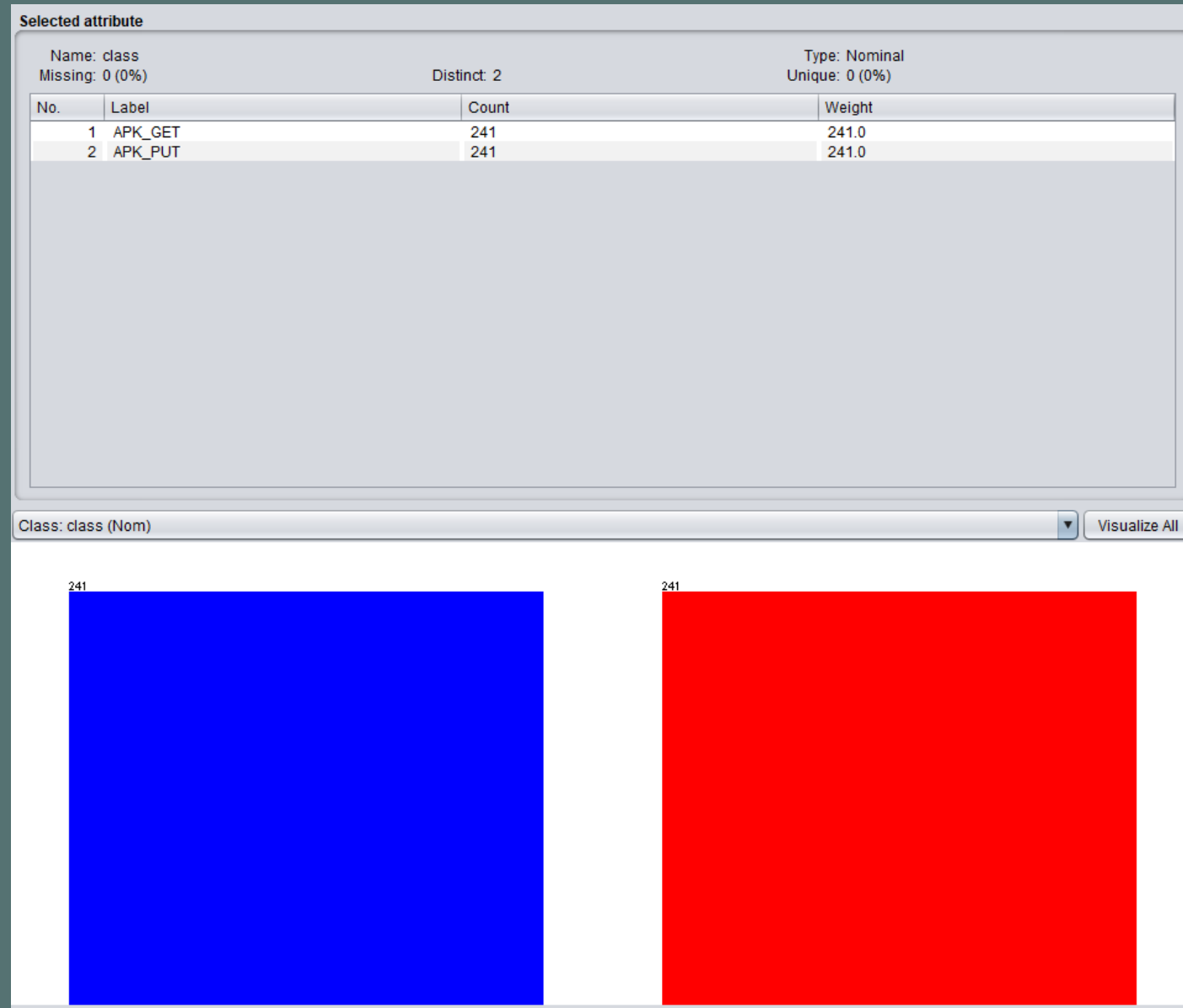
//	2093 SECONDI	SPLIT	1046 SECONDI	SPLIT2
ALGORITM	PRECISION	RECALL	PRECISION	RECALL
MIEMDD	0.999	0.999	0.997	0.997
MIDD	0.997	0.997	-	-
QUICK DD	0.996	0.996	0.991	0.991
TLC	0.993	0.993	0.996	0.996
MITI	0.981	0.981	0.913	0.902
MIRI	0.98	0.979	0.913	0.902
MILR	0.949	0.946	0.981	0.981
KNN	0.924	0.915	0.916	0.905
MDD	0.852	0.818	0.915	0.905
MISVM	0.806	0.733	0.802	0.724
TLD	0.79	0.299	-	-

# OSSERVAZIONI DATASET CON LABEL CLASS GET / PUT

*I due dataset contengono le features degli  
audio splittati.*

*Il primo dataset ha split di 2093 secondi  
il secondo dataset ha split di 1046  
secondi.*

*Le istanze delle bag sono quindi il doppio  
del primo dataset*



- Le istanze sono bilanciate (241 apk\_get e 241 apk\_put) dato che ogni set ha un apk di get ed uno di put.
- I primi tre risultati miglior hanno valore di precision e recall coincidente.

Split 2093 secondi Circa 8 istanze per bag			Split 1046 secondi Circa 16 istanze per bag	
//	2093 SECONDI	SPLIT	1046 SECONDI	SPLIT2
ALGORITMO	PRECISION	RECALL	PRECISION	RECALL
TLC	0.979	0.979	0.992	0.992
MITI	0.971	0.971	0.979	0.979
MIRI	0.965	0.965	0.975	0.975
MIDD	0.878	0.861	0.913	0.913
MDD	0.804	0.712	0.805	0.703
TLDSIMPLE	0.755	0.519	0.754	0.515
MIEMDD	0.751	0.502	0.7	0.618
QUICKDD	0.729	0.726	0.869	0.838
MILR	0.599	0.562	0.574	0.573
MISVM	0.523	0.523	-	-
BOOST	0.498	0.498	0.498	0.498
WRAPPER	0.498	0.498	0.498	0.498
KNN	0.497	0.498	0.497	0.498
SIMPLEMI	0.494	0.498	0.494	0.498