

Analisi del dataset

dataset: **kdd_10_percent_red_CFS_BFSsmall10_2classesSmote3**

In questo dataset le classi sono state diminuite a 2: *normal* e *attack*, come nel precedente dataset, tuttavia in questo caso il numero delle istanze tra *normal* e *attack* è stato bilanciato.

E' stato individuato un leggerissimo miglioramento generale, tenendo conto che anche lo scorso dataset aveva ottenuto ottimi risultati.

Qui di seguito riporto per comodità uno screenshot dei due file *Analisi_algoritmi*, dei rispettivi dataset.

*Dataset avente due classi **NON** bilanciate*

Algoritmo	Accuratezza %	K statistic	FP_rate	TP_rate	FN_rate (confidence 0.01)
J48	99.9	0.9968	0.001	0.999	0
Naive Bayes	99.64	0.9576	0.017	0.986	0.02
BayesNet	99.93	0.9981	0.001	0.999	0
Multilayer Perceptron	99.39	0.9809	0.007	0.994	0.003
IBk (1)	99.93	0.9979	0.001	0.999	0
IBk (3)	99.91	0.9974	0.002	0.999	0
IBk (5)	99.92	0.9976	0.001	0.999	0

*Dataset corrente (con istanze **bilanciate**)*

Algoritmo	Accuratezza %	K statistic	FP_rate	TP_rate	FN_rate (confidence 0.01)
J48	99.95	0.9989	0.001	0.999	0
Naive Bayes	98.34	0.9669	0.017	0.983	0.018
BayesNet	99.95	0.9989	0.001	0.999	0.001
Multilayer Perceptron	99.54	0.9908	0.005	0.995	0.002
IBk (1)	99.95	0.999	0	1	0.001
IBk (3)	99.95	0.999	0	1	0
IBk (5)	99.95	0.9989	0.001	0.999	0

normal_TP	normal_%	normal_istanze	attack_TP	attack_%	attack_istanze
0.999	99.7	9727	0.999	99.9	39667
0.982	98.26	9727	0.987	98.67	39667
0.999	99.85	9727	1	99.95	39667
0.993	98.21	9727	0.994	99.86	39667
0.999	99.85	9727	1	99.95	39667
0.998	99.82	9727	0.999	99.94	39667
0.999	99.78	9727	0.999	99.93	39667

normal_TP	normal_%	normal_istanze	attack_TP	attack_%	attack_istanze
1	99.92	38908	0.999	99.92	39667
0.982	98	38908	0.985	98.54	39667
0.999	99.94	38908	0.999	99.94	39667
0.998	99.25	38908	0.993	99.26	39667
0.999	99.95	38908	1	99.95	39667
1	99.95	38908	0.999	99.94	39667
1	99.95	38908	0.999	99.92	39667