

## Analisi del dataset

*dataset:* **kdd\_10\_percent\_red\_CFS\_BFSsmall10\_2classes**

In questo dataset le classi sono state diminuite a 2: *normal* e *attack*.

Questo ha influito positivamente:

- **sull' accuratezza:** ogni algoritmo utilizzato ha restituito una % di accuratezza maggiore del 99% (la minima è stata 99.39% data dalla rete neurale), mentre nel dataset precedente (con più classi) la minima è stata del 93.18% (Naive bayes). Inoltre ogni algoritmo ha ottenuto una precisione migliore su questo dataset rispetto all' altro
- **normal true positive rate:** anche qua si sono ottenuti risultati uguali o migliori (anche se di poco), tranne per il NaiveBayes che ha ottenuto un miglioramento notevole su questo dataset
- **probabilità media di riconoscere un' attacco:** nel dataset precedente risultava più difficile classificare e calcolare correttamente la probabilità media di riconoscere un qualsiasi attacco (dato che dipendeva da molte classi), ora invece la probabilità media di riconoscere un attacco informatico è molto più semplice e facile da leggere, e anche qua la probabilità si è mostrata essere sempre superiore del 99%
- **sullo sbilanciamento fra classi:** alcuni attacchi nel dataset precedente avevano qualche centinaia di istanze, mentre altri addirittura una sola soltanto. In questo modo, suddividendo il dataset solamente in due classi, si è ottenuto un bilanciamento molto più equo ed omogeneo
- **miglior comprensione generale dei dati** (*come il tasso di falsi negativi*)

Generalmente mi ritengo soddisfatto dei risultati ottenuti, malgrado tutto, in entrambi i dataset i risultati sono stati molto buoni.

L'unico problema di questo dataset è lo sbilanciamento del numero di istanze delle classi, tuttavia questo verrà migliorato nel successivo dataset.