Analisi dei risultati ottenuti

Costruzione dei dataset e descrizione

Come prima cosa sono state riunite in unico file tutte le classi normal e anomaly appartenenti sia al file Train che Test di partenza dati dal dataset NSL.
Questo file è stato dato in "pasto" al software WEKA e tramite la selezione degli attributi (CFS, con ricerca bidirezionale) è stato creato un nuovo dataset con meno attributi, presente nella cartella dataset con il nome di `KDD_AllNormal_AllAnomaly_Filtered.arff`.
A partire da questo dataset sono state estratte l' 80% delle istanze *normal*, per poter creare il file di train, con il nome di `KDD_Train_80%Normal_Filtered.arff`.
Per creare il file di test, sono state prese il 20% delle istanze scartate precedentemente, insieme a tutte le istanze *anomaly* presenti nel file `KDD_AllNormal_AllAnomaly_Filtered.arff,` questo file è presente nella cartella dataset con il nome di `KDD_Test_NormalAnomaly`.

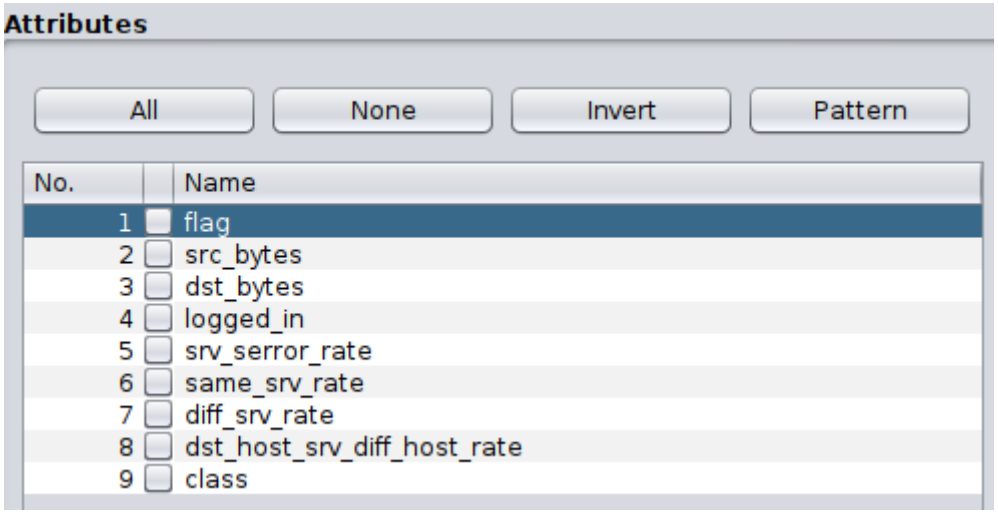
Risultati ottenuti

Il file `KDD_AllNormal_AllAnomaly_Filtered.arff` contiene:

- 77054 istanze normal
- 71463 istanze anomaly

di conseguenza la selezione degli attributi ha preso in esame questo numero di istanze per le due classi.

Attributi selezionati



Train_set	**Test_set**
KDD_Train_80%Normal_Filtered	KDD_Test_NormalAnomaly

I risultati ottenuti sono presenti nel file `Analisi.csv`, qui di seguito vi è riportato uno screenshot della tabella:

File	%Correttezza	%Anomaly	%FN	Istanze ? nel test	Istanze normal nel test	Tot istanze nel test	%FP
Train_NormalFiltered_Test_NormalAnomalyFiltered_csv	85.84	88.6	11.4	71463	15410	86873	26.98

Considerazioni

Sembrerebbe che l'esperimento abbia dato risultato leggermente peggiori rispetto a quelli della volta scorsa (*forse la selezione degli attributi non ha funzionato come si sperava?*).
Per comodità riporto di seguito la tabella dei risutati degli esperimenti scorsi:

File	%Correttezza	%Anomaly	%FN	Istanze ? nel test	Istanze normal nel test	Tot istanze nel test	%FP
FullTrain_FullTest_csv	-	98.66	1.34	71463	0	71463	-
TrainNormal20_TestAnomaly20_csv	-	98.94	1.06	14294	0	14294	-
TrainNormal20_TestNormal20Anomaly80_csv	89	98.93	1.07	11389	2860	14249	50
FullTrainNormal_TestNormal20Anomaly80_csv	88.85	98.98	1.02	11389	2860	14249	51

Sembra esserci tuttavia un miglioramento però per i falsi positivi (FP), ovvero tutte quelle istanze normal riconosciute erroneamente come anomalie.