Data Mining e Scoperta di Conoscenza

Progetto 8

Realizzazione di uno Intrusion Detector

Si realizzi un modello predittivo (classificatore) capace di distinguere tra “cattive” o “buone” connessioni verso un computer.
Una connessione è una sequenza di pacchetti TCP che iniziano e finiscono in fissati istanti di tempo, realizzando uno scambio di dati tra un indirizzo IP sorgente ed un indirizzo IP destinatario sulla base di un preciso protocollo. Ogni connessione è etichettata come normale o come attacco (in tal caso specificando un preciso tipo di attacco).
Gli attacchi ricadono in 4 principali categorie:

- DOS: denial-of-service;
- R2L: accesso non autorizzato da una macchina remota, per esempio indovinando la password;
- U2R: accesso non autorizzato ai privilegi di amministratore;
- probing: port scanning.

In osservanza allo scenario descritto, si chiede di:

1. di preprocessare opportunamente i dati di training forniti;
2. dopo aver costruito il classificatore, testarlo sui dati di test forniti.
3. effettuare una ROC analysis e valutare il classificatore scelto rispetto ad altri modelli di classificazione presenti in letteratura.

I datasets completi, oltre ad un file contenente una spiegazione dettagliata degli attributi ivi contenuti, vengono forniti contestualmente al progetto.

NOTE PER L’ESECUZIONE DEL PROGETTO

1. Scrivi un rapporto di circa 10 pagine in cui
    a. Descrivi analiticamente l’algoritmo che hai implementato.
    b. Commenti le parti essenziali del codice Java che hai scritto, e metti in un’appendice l’intero codice
    c. commenti e illustri graficamente e quantitativamente gli esperimenti effettuati.
2. Prepare delle slides Powerpoint (non più di 10 slides) in cui riassumi gli esiti del progetto