



**Schéma De Traitement KDD99**

### **Pre-Processing :**

D'une manière générale, le prétraitement des données est une étape importante du processus d'exploration de données, permet d'éliminer les mauvaises données qui risquent de causer des problèmes sur l'ensemble du data.

Pour éviter de traiter des données erronées et améliorer la performance du modèle, il faut impérativement analyser les données, détecter les anomalies le plus tôt possible et déterminer les étapes de prétraitement et de nettoyage appropriées. Ceci est effectué sur KDD99 comme suivant :

Premièrement, on a collecté les données à traiter des deux fichiers (Training and Testing) format csv. Ces derniers sont fiables et correctement structurés afin que les données importées soient en meilleure qualité possible. On les nettoyant ; on supprime les champs nuls avec les valeurs duplicates.

Ensuite, on a essayé d'appliquer des fonctions qui font :

- Nettoyage des données
- Classification des champs selon les catégories.
  - a. Classification des services
  - b. Classification des types de protocole
  - c. Classification des flags
- Définir et trier ses catégories selon les types.
- Conversion des valeurs en numériques.
- Numérisation des données dans un intervalle de [0,1] afin d'éliminer toute sorte de valeurs.
- Trier les attaques selon les types de protocoles
- Classifier ses attaques selon les 4 grandes catégories : [DOS, Probe, R2L, U2R, Normal]
- Numériser la partie des attaques.

Après le traitement de ces données, elles sont stockées pour une utilisation ultérieure, avec un schéma de statistique qui représente le pourcentage de la présence des différentes attaques tirées du Training Data pendant l'utilisation des 3 types de protocoles [ICMP, UDP, TCP].

**Par : KHAWLA OUGRIRANE**

**MOHAMED HAMDAOUI**