Détection d'intrusions par l'analyse Big Data des fichiers logs

Réalisé par : ALLAL Yahia , AIT ICHOU Mustapha , AMMOURI Yassire

Faculté des Sciences de Rabat

12 Juillet 2024

Plan

Problématique

Problématique

1. Les systèmes informatiques sont essentiels.
2. La sécurité des systèmes est très importante.(Alors, comment peut-on protéger... ?)
3. Les anciennes méthodes ne sont pas efficaces.
4. En utilisant les outils de big data pour résoudre ce problème.

Définition des Fichiers de Logs et leurs Types

Définition des Fichiers de Logs

Figure: Définition des Fichiers de Logs et leurs Types

Les fichiers de logs sont des enregistrements chronologiques des événements ou des transactions générés par les systèmes informatiques.

Types des Fichiers de Logs

Figure: Les Types de Fichiers de Logs

- **Log web** : Sont des événements web (les requêtes HTTP, les adresses IP des visiteurs, et les erreurs serveur).
- **Log de réseau** : Sont les activités et les erreurs liées aux services de réseau (les connexions entrantes et sortantes, les transferts de données).
- **Log système** : Sont des événements généraux du système (les démarrages et arrêts, les mises à jour du système).

DataSet

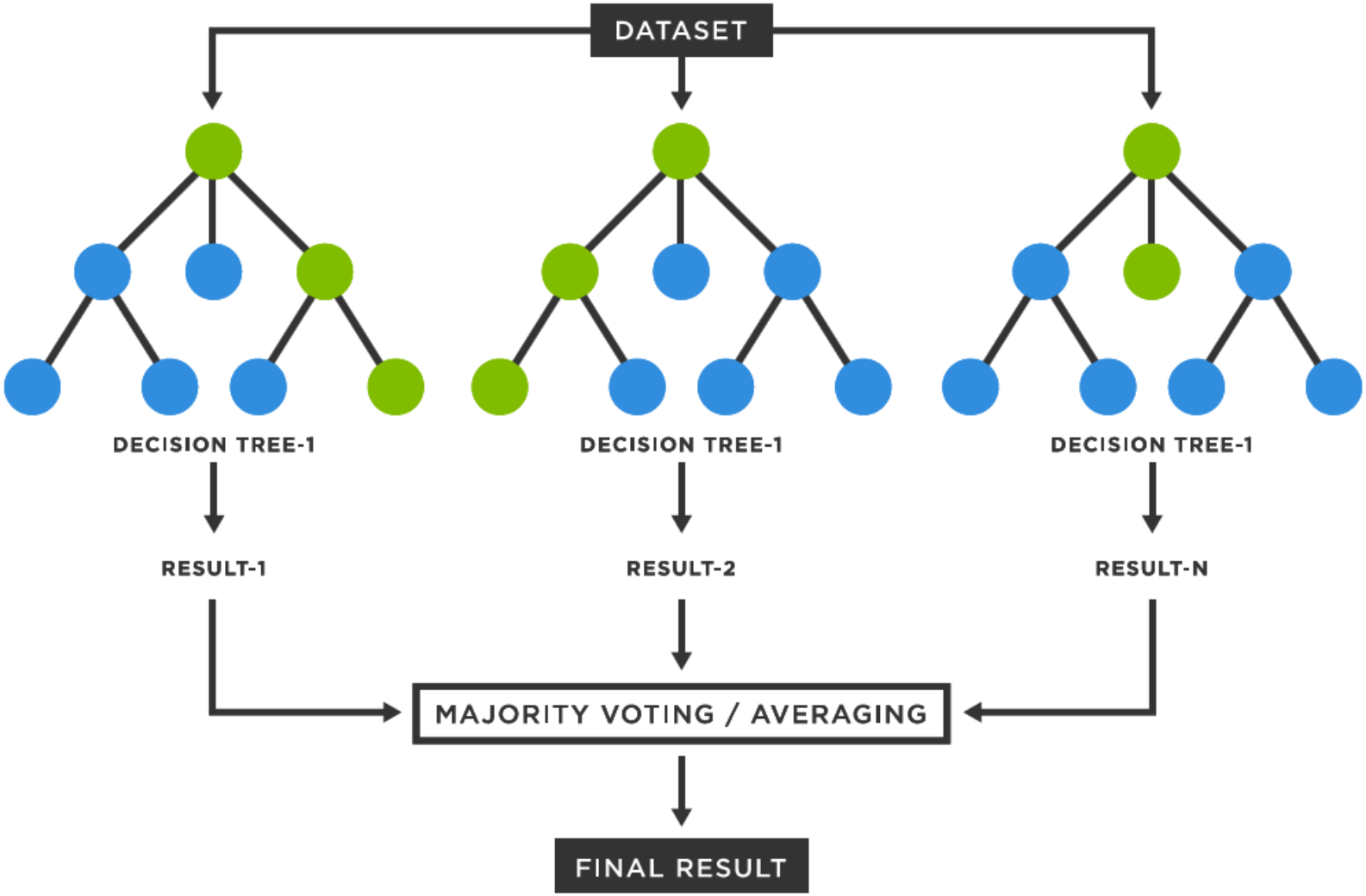
Data set : KDDCup99

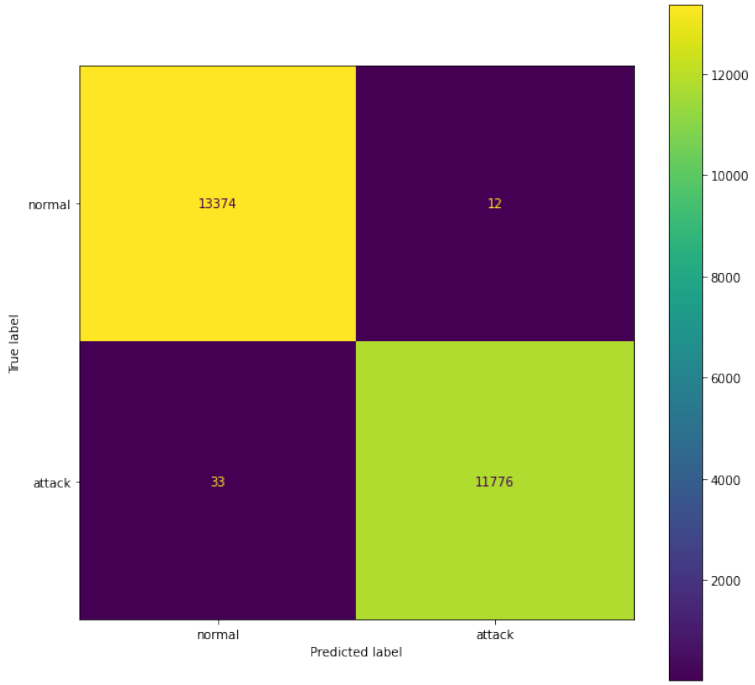
KDDCup99

- **Nombre d'Instances :** Environ 4,9 millions d'instances de connexions réseau.
- **Nombre de features :** 42 caractéristiques
- **Types d'Attaques :** 22 types d'attaques répartis en quatre catégories principales :
  - **Denial of Service (DoS)** : Déni de service
  - **Remote to Local (R2L)** : À distance vers local
  - **User to Root (U2R)** : Utilisateur vers superutilisateur
  - **Probing or port scanning** : Exploration ou balayage de port

Les modèles de machine learning choisis

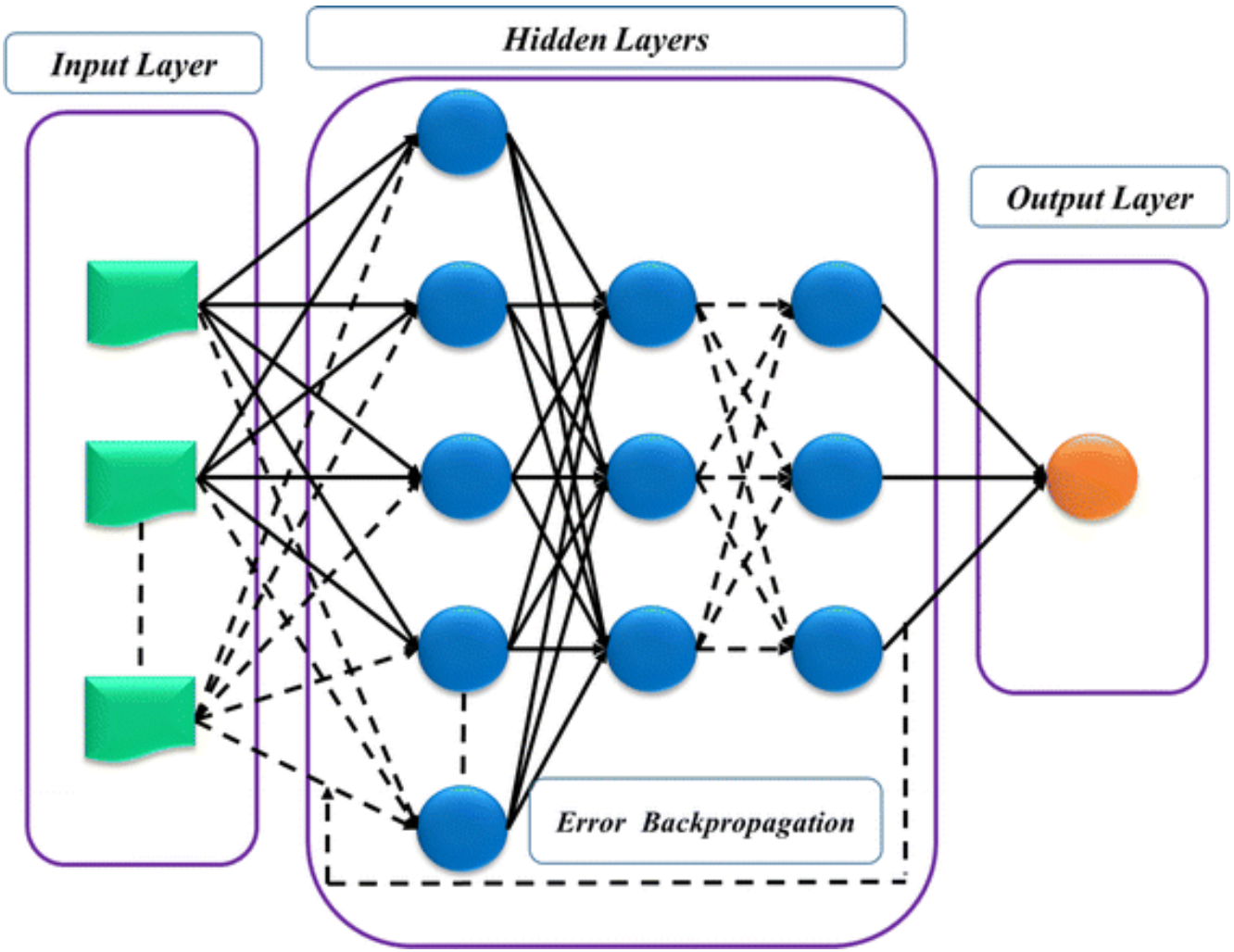
Random Forest





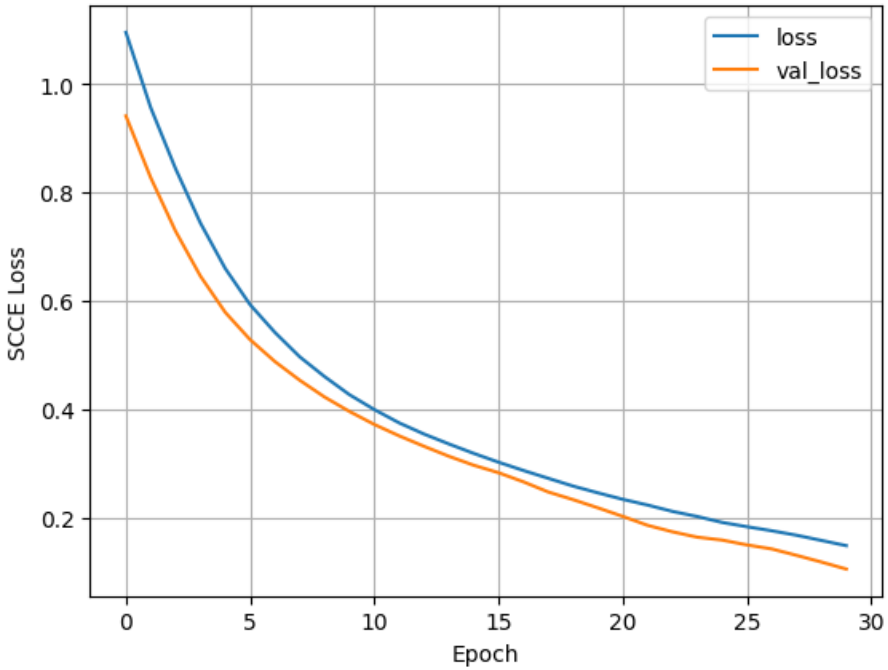

```
Testing Set:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00     13386
           1       1.00      1.00      1.00     11809

    accuracy                           1.00     25195
   macro avg       1.00      1.00      1.00     25195
weighted avg       1.00      1.00      1.00     25195

Confusion Matrix:
[[13379     7]
 [   25 11784]]
Accuracy: 0.9987299067275253
```

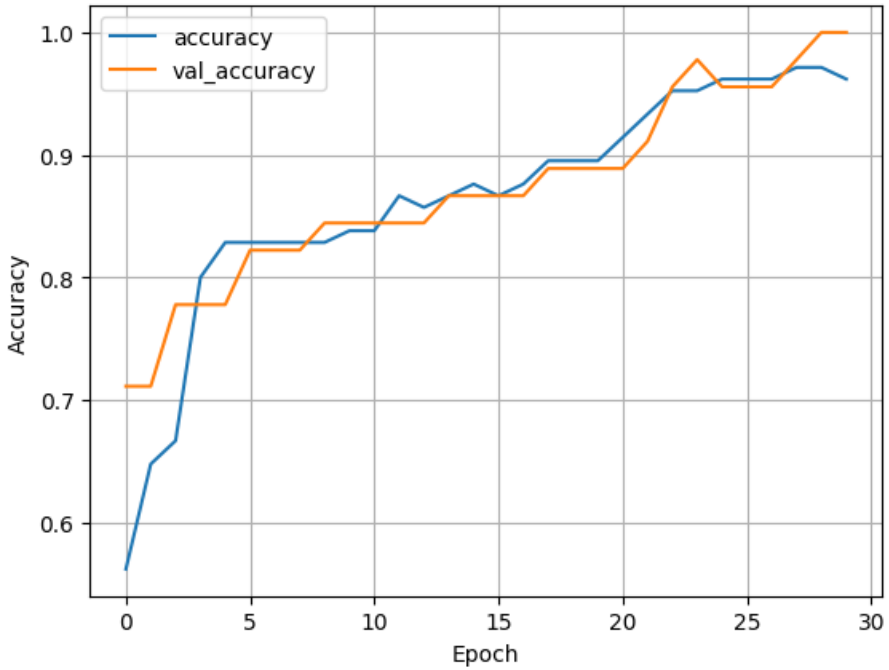
Artificial Neural Network (ANN)



```python
model = tf.keras.Sequential([
    tf.keras.layers.Dense(units=64, activation='relu', input_shape=(x_train.shape[1:]),
                          kernel_regularizer=regularizers.L1L2(l1=1e-5, l2=1e-4),
                          bias_regularizer=regularizers.L2(1e-4),
                          activity_regularizer=regularizers.L2(1e-5)),
    tf.keras.layers.Dropout(0.4),
    tf.keras.layers.Dense(units=128, activation='relu',
                          kernel_regularizer=regularizers.L1L2(l1=1e-5, l2=1e-4),
                          bias_regularizer=regularizers.L2(1e-4),
                          activity_regularizer=regularizers.L2(1e-5)),
    tf.keras.layers.Dropout(0.4),
    tf.keras.layers.Dense(units=512, activation='relu',
                          kernel_regularizer=regularizers.L1L2(l1=1e-5, l2=1e-4),
                          bias_regularizer=regularizers.L2(1e-4),
                          activity_regularizer=regularizers.L2(1e-5)),
    tf.keras.layers.Dropout(0.4),
    tf.keras.layers.Dense(units=128, activation='relu',
                          kernel_regularizer=regularizers.L1L2(l1=1e-5, l2=1e-4),
                          bias_regularizer=regularizers.L2(1e-4),
                          activity_regularizer=regularizers.L2(1e-5)),
    tf.keras.layers.Dropout(0.4),
    tf.keras.layers.Dense(units=1, activation='sigmoid'),
])
```

```
4/4 [==============================] - 0s 15ms/step - loss: 0.1675 - accuracy: 0.9714 - val_loss: 0.1304 - val_accuracy: 0.9778
Epoch 29/30
4/4 [==============================] - 0s 18ms/step - loss: 0.1579 - accuracy: 0.9714 - val_loss: 0.1180 - val_accuracy: 1.0000
Epoch 30/30
4/4 [==============================] - 0s 14ms/step - loss: 0.1483 - accuracy: 0.9619 - val_loss: 0.1051 - val_accuracy: 1.0000
Model: "sequential_8"
```

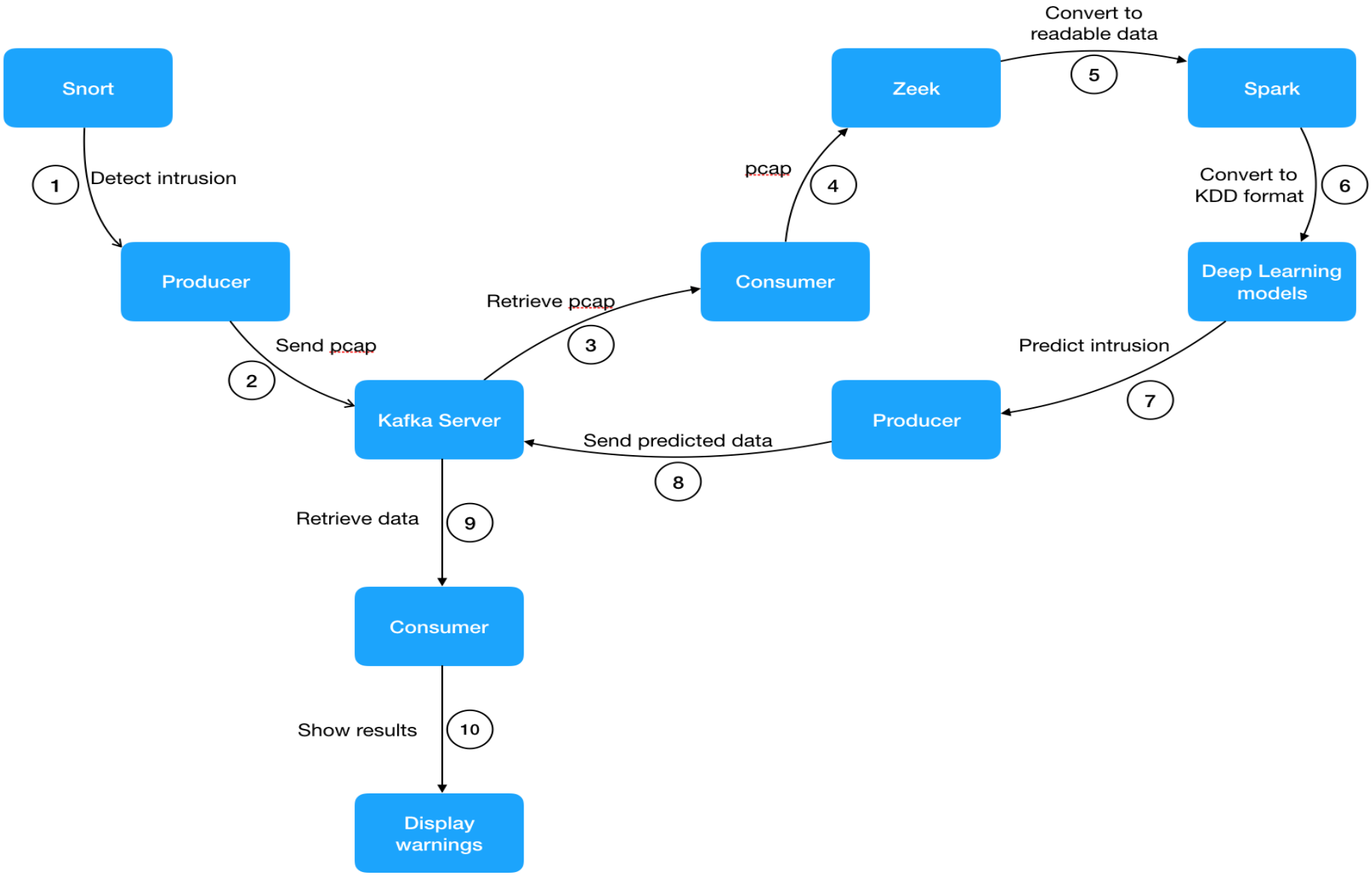

loss
val_loss
SCCE Loss
Epoch
1.0
0.8
0.6
0.4
0.2
0
5
10
15
20
25
30


accuracy
val_accuracy
Accuracy
Epoch
1.0
0.9
0.8
0.7
0.6
0
5
10
15
20
25
30

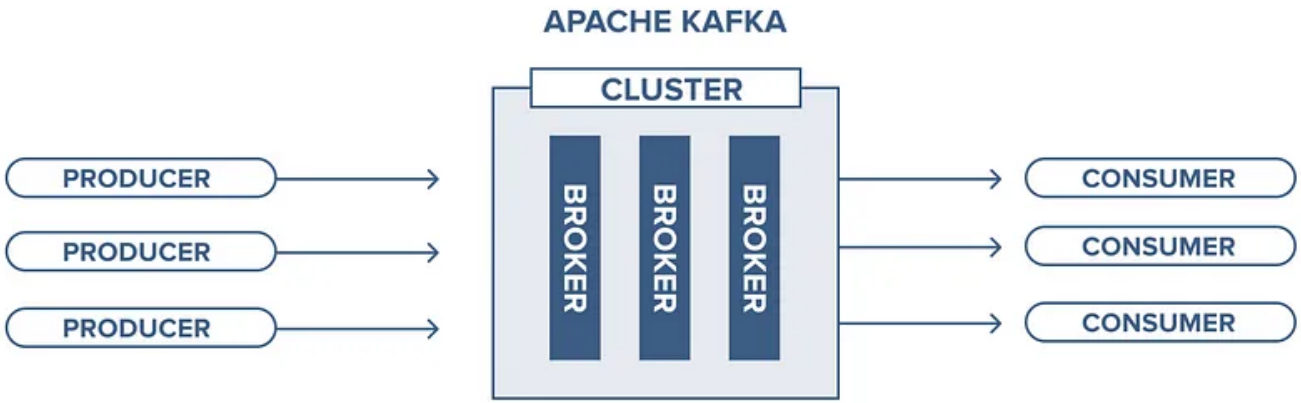
Les outils & Implémentation

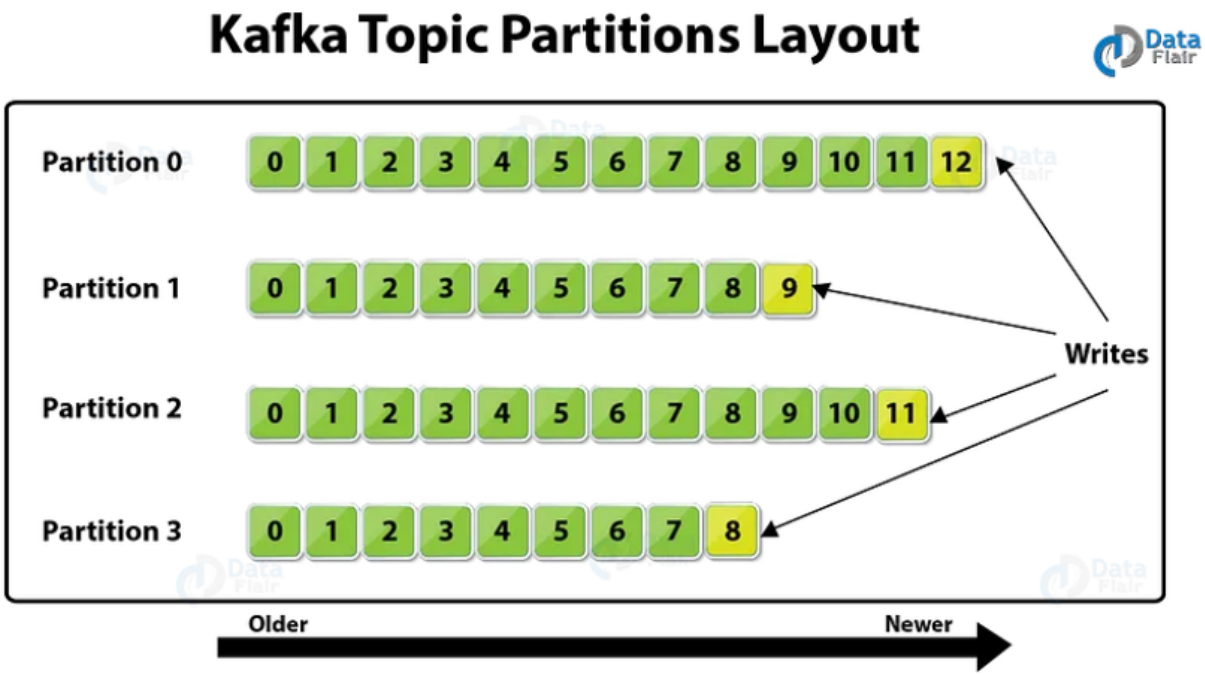
Les outils

Implmentation



Comment fonctionne Kafka





Démonstration

Conclusion

Conslusion

La détection d'intrusions par l'analyse Big Data des fichiers logs présente plusieurs défis :

1. **Attaques zero-day** : Difficiles à détecter car elles exploitent des vulnérabilités inconnues.
2. **Complexité des environnements réseau :** Adaptation aux réseaux modernes complexes et distribués.
3. **Évolution des techniques d'attaque :** Les systèmes doivent s'adapter rapidement aux nouvelles menaces.

Merci pour votre attention !