AD4IDS Sous-projet 2
Flow Classification

P-F. Marteau, ENSIBS Cyber Sécurité du Logicile 3A

October-November 2023

1 Preparing the classification tasks

We will use a cross validation approach. In this framework, the ISCX dataset will be first subdivided according to the appName field (protocol name of the application layer) attached to the flows (at minima HTTP and SSH):

- HTTPWeb
- HTTPImageTransfer
- POP
- IMAP
- DNS
- SSH
- SMTP
- FTP
- ICMP

For each above listed application protocol, the ISCX subset of flows is partitioned in 5 subsets of equal size S_1, S_2, S_3, S_4 and S_5.

Task$_i$, $i \in \{1, .., 5\}$ corresponds to the situation where S_i is used for testing and $T_i = \cup_{j \neq i} S_j$ is used for the training.

Finally we will evaluate each classifier on the 5 previous tasks and will average the evaluation measures.

2 k-NN classification

Implement a k-NN classifier, binary version, implemented in the sklearn Python library [1], and run it over the 5 Task$_i$. Get the evaluation measures and comment your results.

3 Naive Bayes classification

Implement a Multinomial Naive Bayes classifier, binary version, implemented in the sklearn Python library [2], and run it over the 5 Task$_i$. Get the evaluation measures and comment your results.

4 Random Forest classification (optional)

Implement a random forest classifie, binary version, implemented in the sklearn Python library [3], and run it over the 5 Task$_i$. Get the evaluation measures and comment your results.

5 Multilayer Perceptron classification (otional)

Implement a Multilaye Perceptron classifier, binary version, implemented in the sklearn Python library [4], and run it over the 5 Task$_i$. Get the evaluation measures and comment your results.

6 Model comparison

Rank the previously tested model according to the various evaluation metrics you have got. What is your recommendation if you were to implement an IDS with supervised learning capabilities?

[1] `https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html`

[2] `https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html`

[3] `https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html`

[4] `https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html`

7 Deliverable

Put your well commented code and a mini documentation (user install/manual + testing code) in a zip file and deposit it into the "Rendu" folder on the ENT/Moodle.

Do not forget the list of licensed code you have reused if any!