SecuML : le Machine Learning pour la détection d'intrusion

Anaël Bonneton[1,2] - Antoine Husson[1]

prenom.nom@ssi.gouv.fr

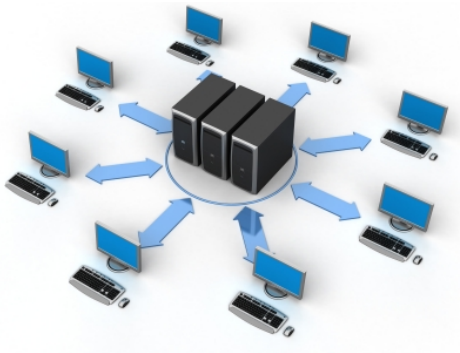
[1] ANSSI

[2] ENS INRIA

SSTIC 2017

Plan

Système de détection d'intrusion



Système d'information

Communications

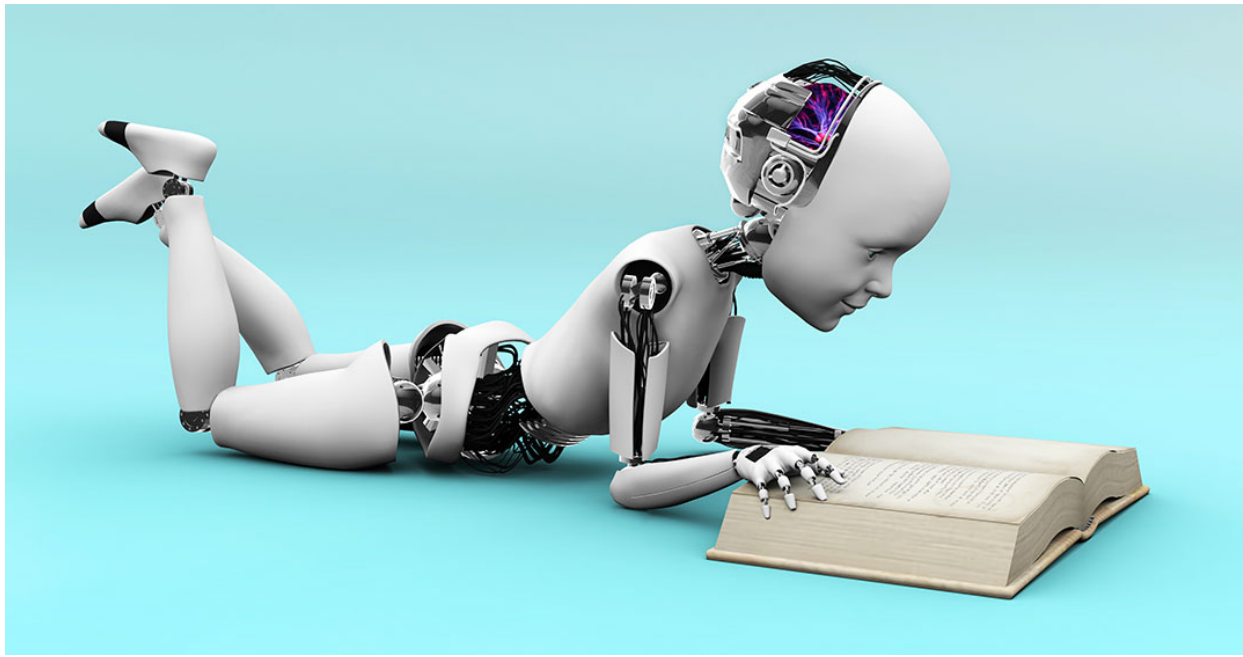
Système de détection

Alertes

Signatures

Règles de détection précises construites par des experts

- ✓ Minimise le taux de faux positifs
- ✓ Alertes faciles à interpréter
- ✗ Peu robustes aux variations des attaques et aux nouvelles attaques



Intelligence Artificielle

Deep Learning

Data Science

Big Data

Machine Learning

Machine Learning et détection d'intrusion

Les plaquettes marketing !

- ✓ Analyse comportementale
- ✓ Attaques inconnues
- ✓ 0-day

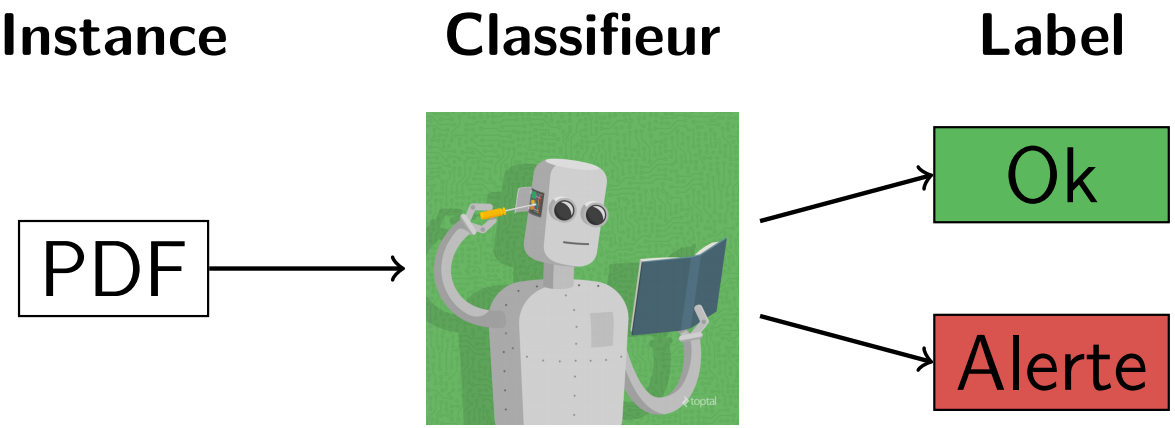
Les sceptiques ...

- ✗ Trop de faux positifs
- ✗ Boîte noire incompréhensible

Comment adapter le Machine Learning à la détection d'intrusion ?

Plan

Machine Learning - Classifieur

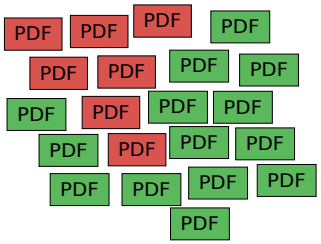


Deux étapes

1. Apprentissage du classifieur
2. Détection

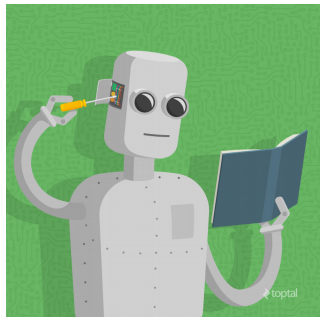
1- Apprentissage d'un classifieur

Jeu de données labélisées

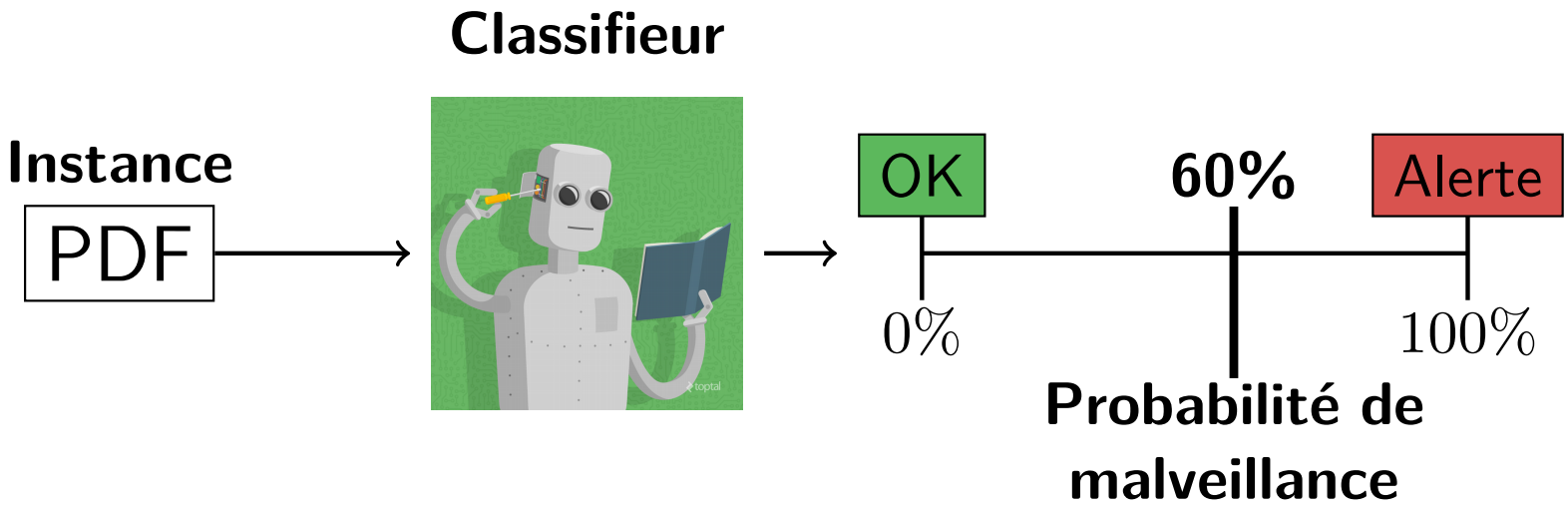


Algorithme d'apprentissage automatique

Classifieur



2- Détection grâce au classifieur



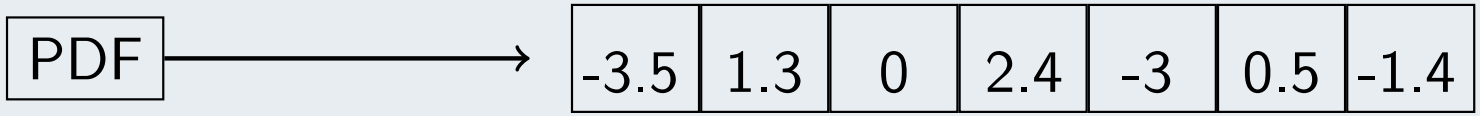
Probabilité de malveillance

- Prioritisation des alertes
- Compromis entre taux de détection et taux de faux positifs

Vecteurs d'attributs numériques

1- Extraction d'attributs

Donnée brute → **Vecteur numérique**



Vecteurs d'attributs numériques

1- Extraction d'attributs

Donnée brute

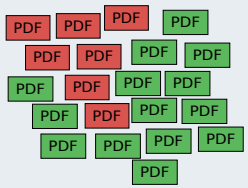
PDF

Vecteur numérique

-3.5	1.3	0	2.4	-3	0.5	-1.4

2- Apprentissage d'un classifieur

Données brutes

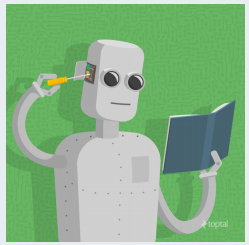


Vecteurs numériques

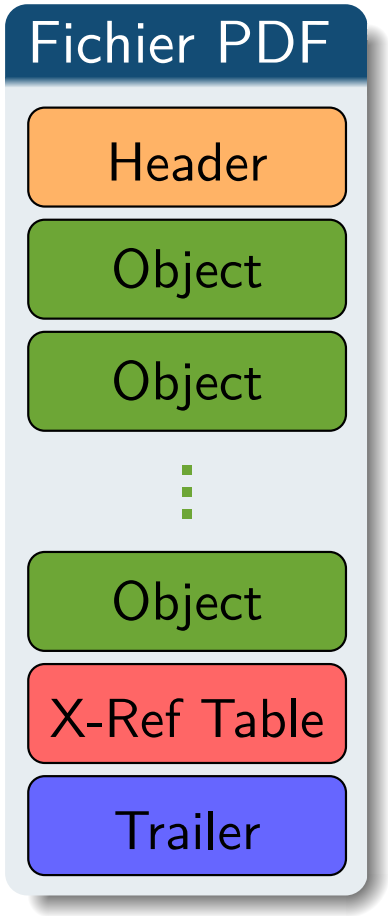
-3.5	1.3	0	2.4	-3	0.5	-1.4
0	3.4	-1	0.3	2.3	-0.5	1.1
...						
-1.3	0	3.2	1.3	-0.7	0	-2.9

Apprentissage automatique

Classifieur



Application aux fichiers PDF



Caractéristiques d'un fichier PDF

- Propres au fichier (taille, timestamp)
- Meta-données (auteur, créateur)
- Objets (type, taille)

Construction d'un vecteur numérique de taille fixe

Listes numériques

$[3,\ 1,\ 2, \ldots,\ 4,\ 5] \longrightarrow$

mean	max	min	var	med
2.4	7	1	1.3	3

Construction d'un vecteur numérique de taille fixe

Listes numériques

$[3,\ 1,\ 2, \ldots,\ 4,\ 5] \longrightarrow$

mean	max	min	var	med
2.4	7	1	1.3	3

Catégories

/Type=JavaScript $\longrightarrow$

XFA	Media	JS	Text	Stream	Font
0	0	1	0	0	0

Construction d'un vecteur numérique de taille fixe

Listes numériques

$[3,\ 1,\ 2, \ldots,\ 4,\ 5] \longrightarrow$

mean	max	min	var	med
2.4	7	1	1.3	3

Catégories

/Type=JavaScript $\longrightarrow$

XFA	Media	JS	Text	Stream	Font
0	0	1	0	0	0

Chaînes de caractères

recrutement@ssi.gouv.fr $\longrightarrow$

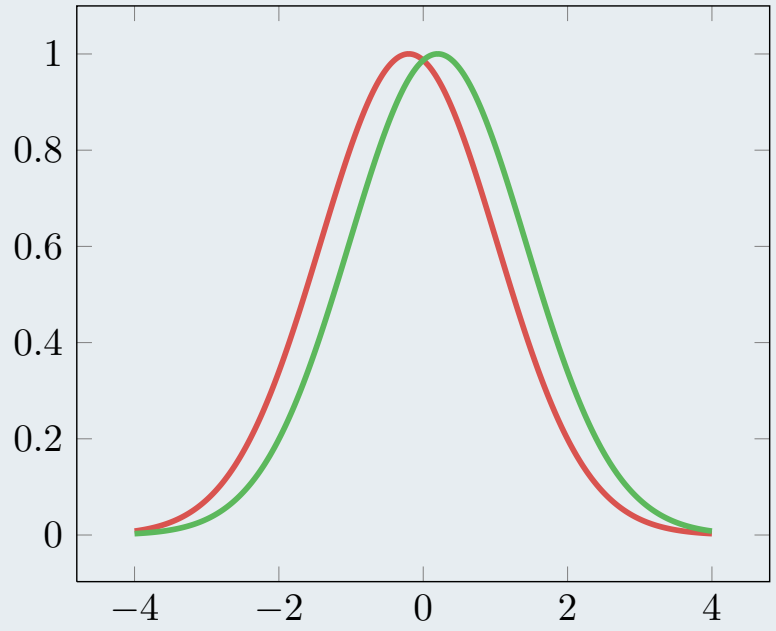
chars	a-z	A-Z	0-9	.	@-£
25	22	0	0	2	1

Plan

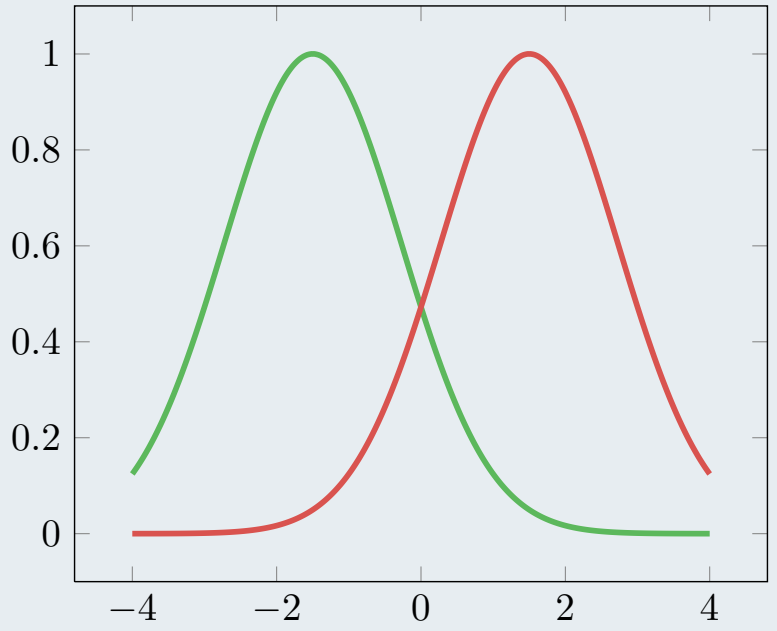
1- Attributs discriminants

- Spécifiques à chaque problème de détection
- Connaissances expert

Non discriminant



Discriminant



2- Choix du modèle

Contraintes

- Prédiction rapide
- Mise à jour périodique du modèle
- Interprétable

2- Choix du modèle

Contraintes

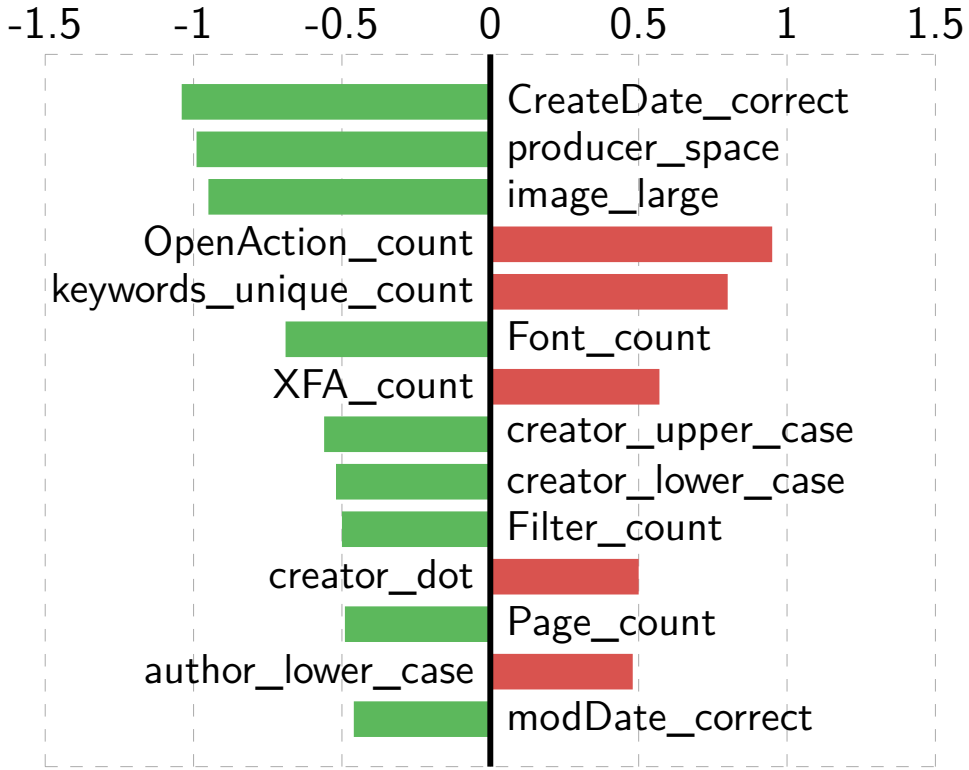
- Prédiction rapide
- Mise à jour périodique du modèle
- Interprétable

✗ Réseaux de neurones

✓ Modèle linéaire

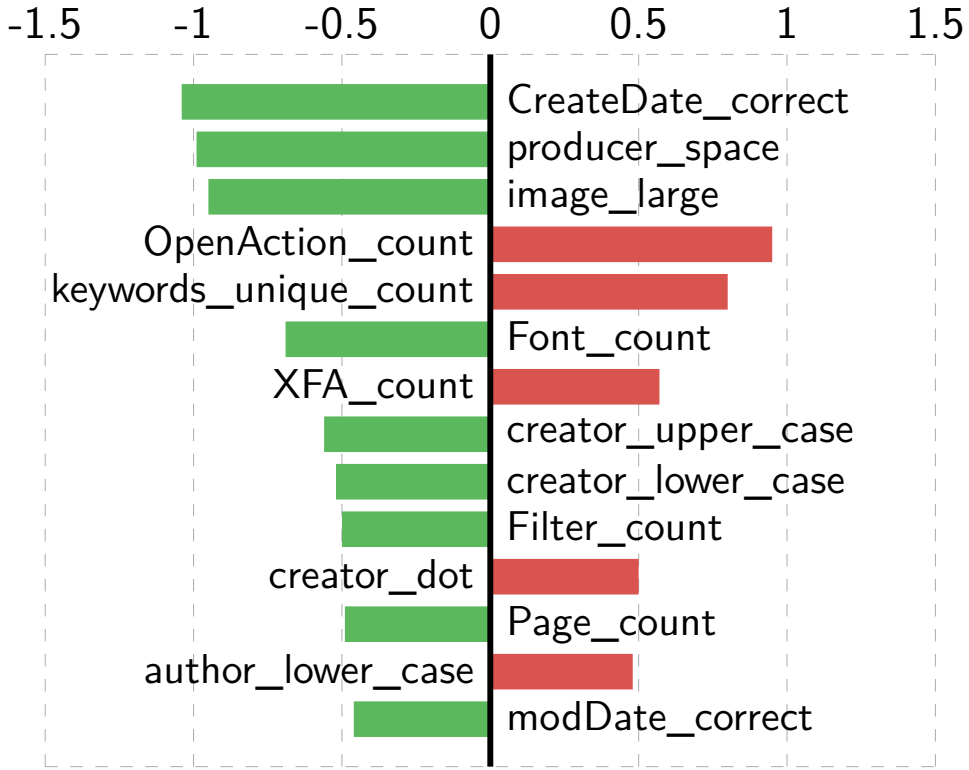
2- Choix du modèle

Les modèles linéaires sont interprétables.



2- Choix du modèle

Les modèles linéaires sont interprétables.



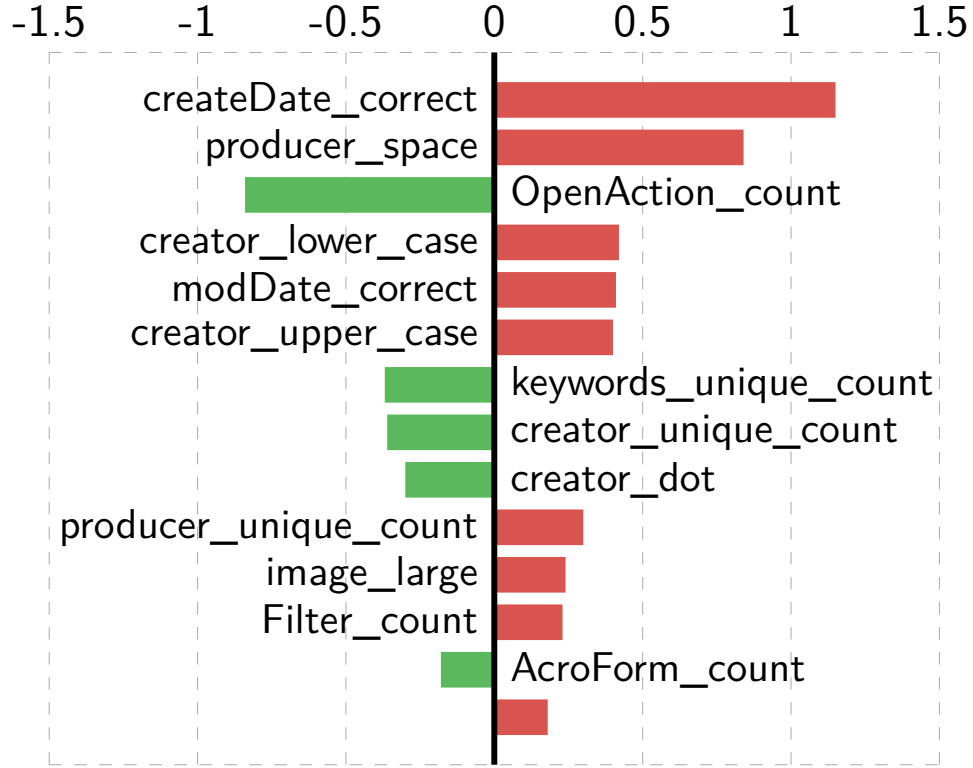
Méthode de scoring

Coefficients optimaux appris automatiquement à partir des données labélisées

2- Choix du modèle

Les prédictions sont aussi interprétables !

Pourquoi une alerte a été générée ?



3- Valider le modèle

Avant la mise en production !

Jeu de données de validation

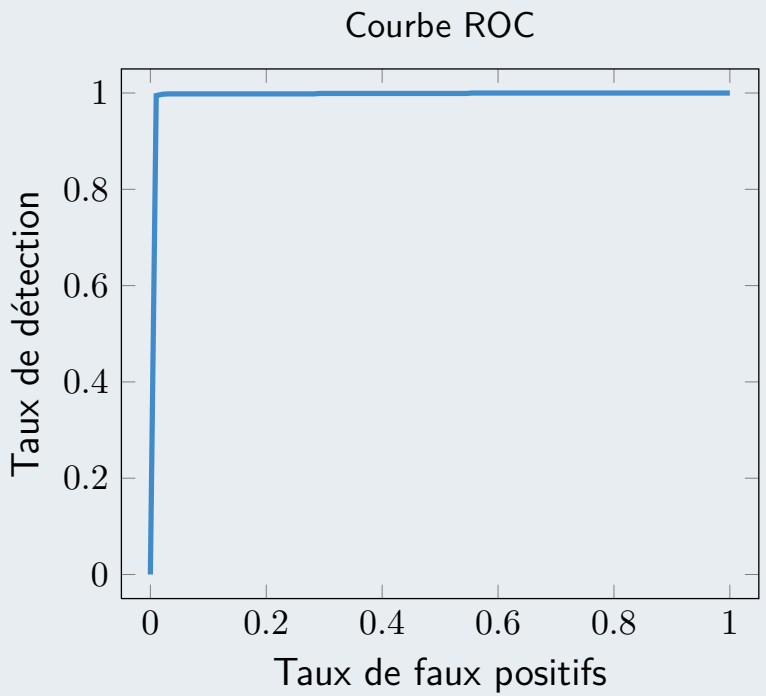
- Données labélisées
- Validation sur des données non utilisées pour l'apprentissage

Méthode de validation

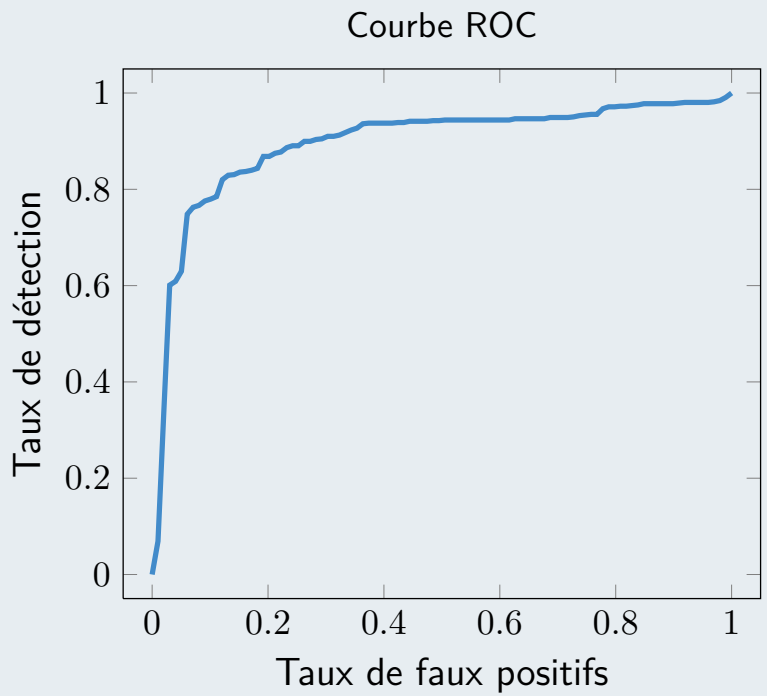
Apprentissage	Validation
90% données	**10% données**

4- Attention aux biais d'apprentissage !

90% Contagio/ 10% Contagio



100% Contagio/ 100% WebPdf



Plan

SecuML

Interface de diagnostic d'un classifieur

- Python, bibliothèque de Machine Learning `scikit-learn`
- Mise en place du modèle avant sa mise en production

Outil générique

PDF

DOC

PCAP

JavaScript

EXE

Netflows

SecuML - Données en entrée

features.csv

```
instance_id,version,num_objects,...
0,1,3,0,0,0
1,1,4,5,4,1
2,1,4,32,13,0
3,1,3,0,0,0
4,1,3,0,0,0
5,1,3,0,0,0
6,1,6,7,6,0
7,1,3,0,0,0
8,1,3,0,0,0
```

true_labels.csv

```
instance_id,label
0,malicious
1,malicious
2,benign
3,malicious
4,malicious
5,benign
6,benign
7,benign
8,benign
```

```
./SecuML_classification LogisticRegression Pdf contagio
```

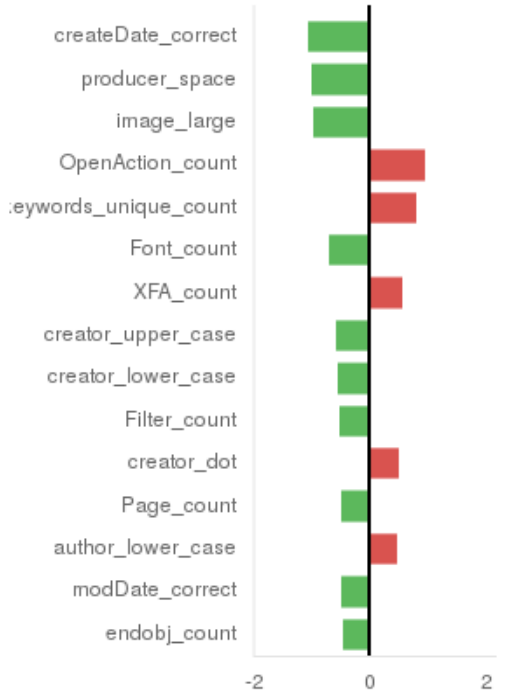
Diagnostic d'un modèle avec SecuML

Démo : Interface de diagnostic du modèle

Diagnostic d'un modèle avec SecuML

Experiment

Model Coefficients



Train

Performance | Predictions

Indicators | ROC | Confusion Matrix

Detection threshold: 42%

Detection	99.35%
False alarms	0.88%
F-score	99.31%
AUC	99.9%

Test

Performance | Predictions

Indicators | ROC | Confusion Matrix

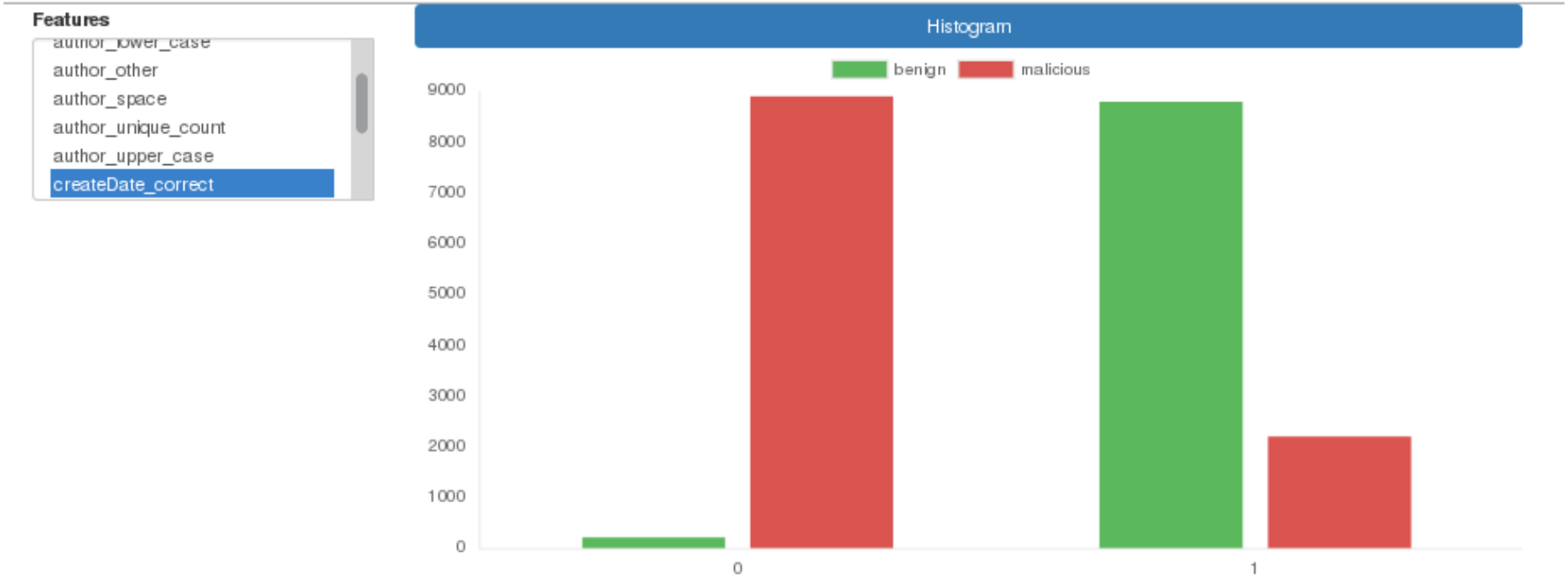
Detection threshold: 50%

Detection	99.28%
False alarms	1.3%
F-score	99.15%
AUC	99.74%

Alerts Analysis

Top N | Random

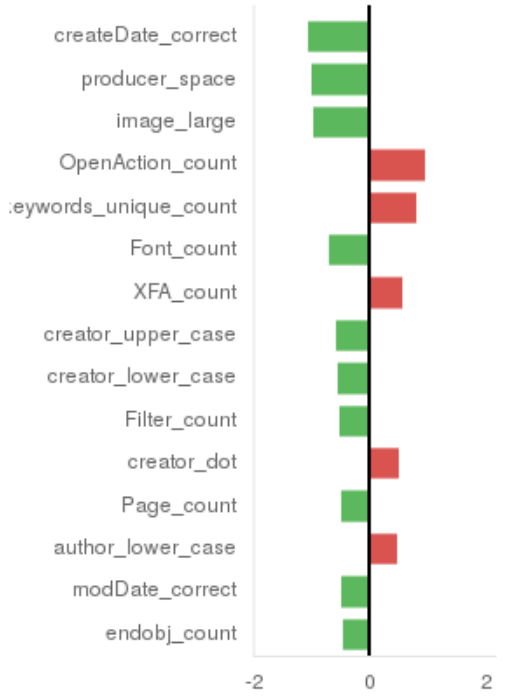
Diagnostic d'un modèle avec SecuML



Diagnostic d'un modèle avec SecuML

Experiment

Model Coefficients



Train

Performance | Predictions

Indicators | ROC | Confusion Matrix

Detection threshold: 42%

Detection	99.35%
False alarms	0.88%
F-score	99.31%
AUC	99.9%

Test

Performance | Predictions

Indicators | ROC | Confusion Matrix

Detection threshold: 50%

Detection	99.28%
False alarms	1.3%
F-score	99.15%
AUC	99.74%

Alerts Analysis

Top N | Random

Diagnostic d'un modèle avec SecuML

Experiment

Model Coefficients

Train

Performance | Predictions

Indicators | ROC | Confusion Matrix

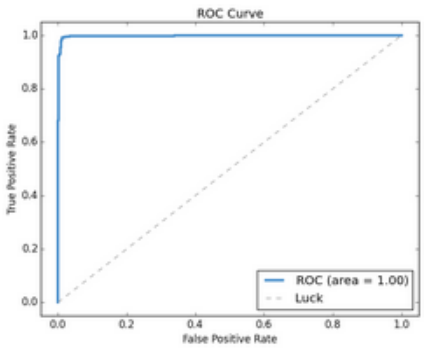
Detection threshold: 42%

Detection	99.35%
False alarms	0.88%
F-score	99.31%
AUC	99.9%

Test

Performance | Predictions

Indicators | ROC | Confusion Matrix



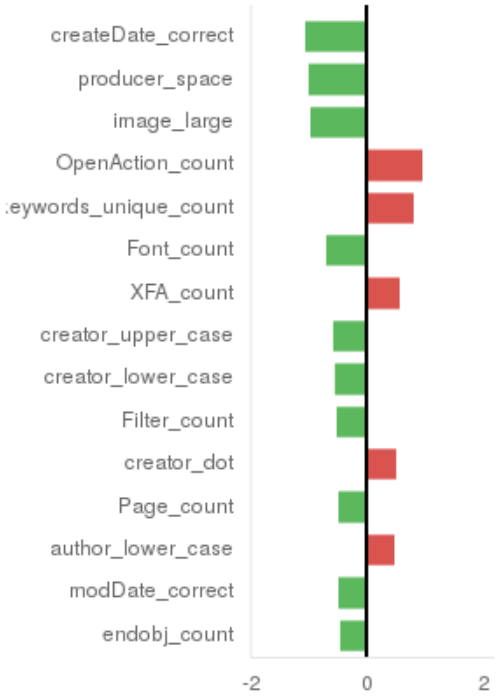
Alerts Analysis

Top N | Random

Diagnostic d'un modèle avec SecuML

Experiment

Model Coefficients



Train

Performance | Predictions

Indicators | ROC | Confusion Matrix

Detection threshold: 42%

Detection	99.35%
False alarms	0.88%
F-score	99.31%
AUC	99.9%

Test

Performance | Predictions

Indicators | ROC | Confusion Matrix

		Prediction	
		alert	ok
Truth	alert	1116	8
	ok	11	832

False Positives

Alerts Analysis

Top N | Random

Diagnostic d'un modèle avec SecuML

Diagnostic d'un modèle avec SecuML

False Positives

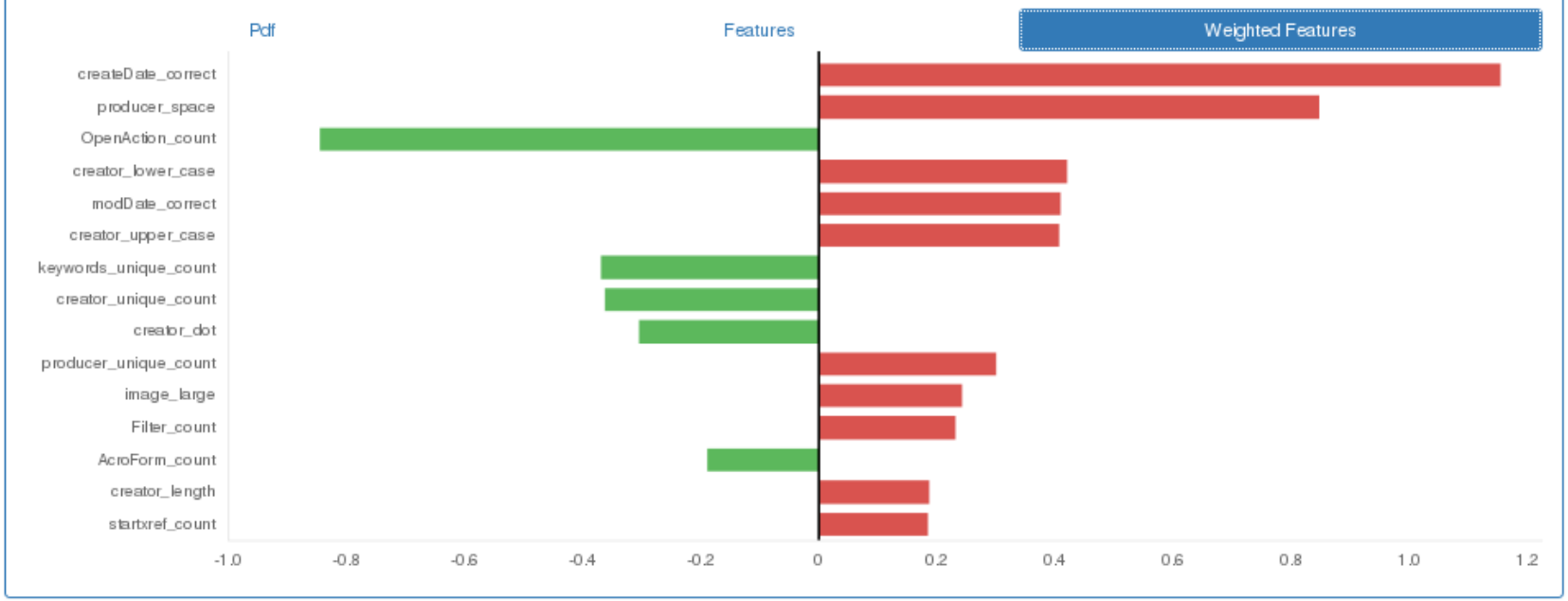
Undetected

False Positives

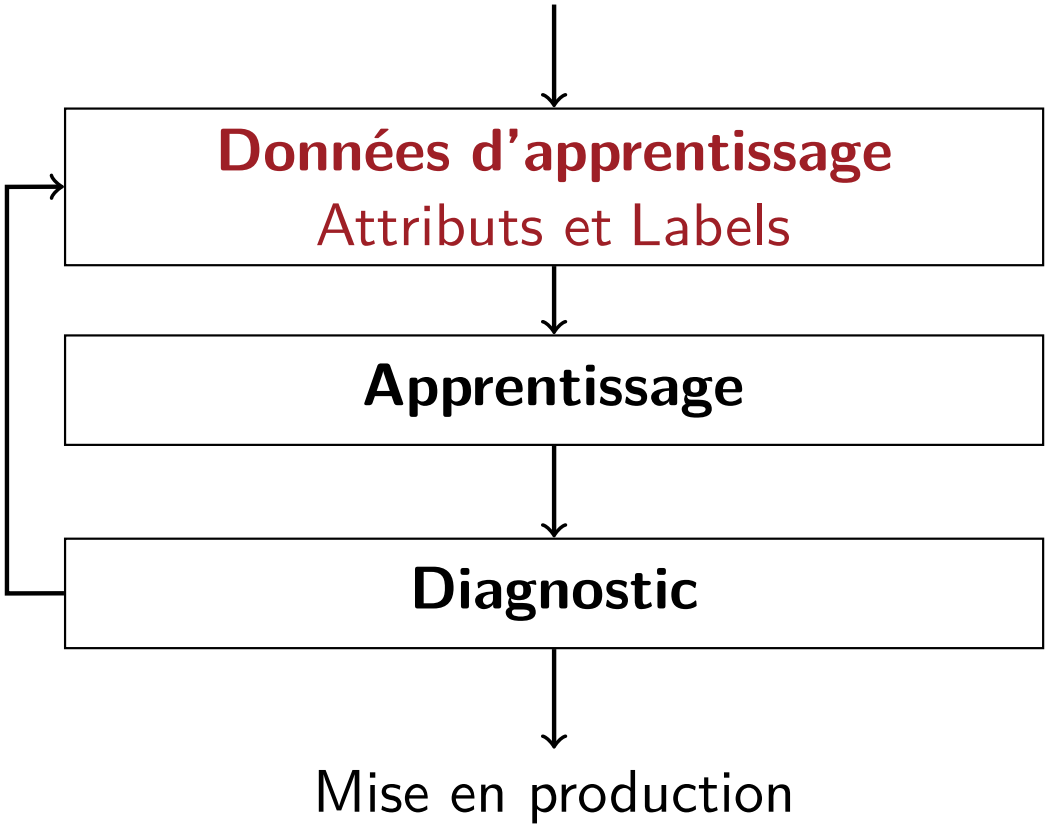
Error 1 / 11 Prev Next

Instance 11530: Contagio/CLEAN/CLEAN_PDF_9000_files/Barbara_Harmon.pdf

Description



Mise en place d'un classifieur



Mise en place d'un classifieur avec SecuML

Spécificités de chaque problème de détection

- Extraire des attributs discriminants
- Recueillir des données labélisées représentatives
- Visualisation spécifique d'une instance

Apports de SecuML

- Méthode générique
- Apprentissage / Validation
- Interface de diagnostic

Construisez vos modèles de détection avec SecuML

Testez SecuML !

https://github.com/ANSSI-FR/SecuML

- Jeu de données disponible pour la détection de spams
- Vos propres jeux de données

Construisez vos modèles de détection avec SecuML

Testez SecuML !

https://github.com/ANSSI-FR/SecuML

- Jeu de données disponible pour la détection de spams
- Vos propres jeux de données