SecuML: Machine Learning for Computer Security Experts

Anaël Bonneton

anael.bonneton@ssi.gouv.fr

ANSSI, ENS Paris, INRIA

PyParis 2017

Intrusion Detection System

Information System

Communications

Detection System

Alerts

Traditional Detection Methods

Signatures: Precise detection rules built by security experts

- ✓ Low false alert rate
- ✓ Alerts easy to interprete
- ✗ Not robust to attack variations, to new attacks

Traditional Detection Methods

Signatures: Precise detection rules built by security experts

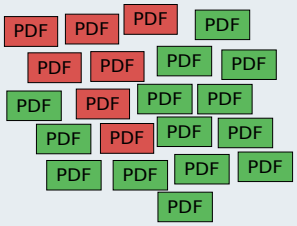
- ✓ Low false alert rate
- ✓ Alerts easy to interprete
- ✗ Not robust to attack variations, to new attacks

Machine Learning !

Supervised Detection Model

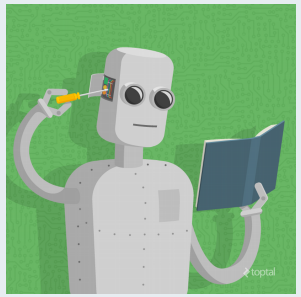
Training

Training Data



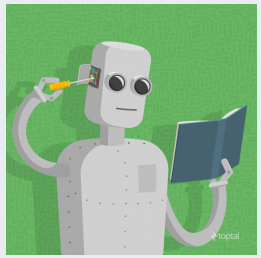
Learning Algorithm

Classifier



Predicting

PDF



Ok

Alert

Computer Security Specificities

Non Machine Learning Experts

- Machine Learning pipeline
- Machine Learning jargon

Computer Security Specificities

Non Machine Learning Experts

- Machine Learning pipeline
- Machine Learning jargon

Lack of training data

- Few public labelled datasets
- No crowdsourcing

Computer Security Specificities

Non Machine Learning Experts

- Machine Learning pipeline
- Machine Learning jargon

Lack of training data

- Few public labelled datasets
- No crowdsourcing

Need for interpretation

- How does the detection model work ?
- Why an alert has been raised ?

Applying Machine Learning

✗ Deep Learning Frameworks

- ✗ TensorFlow
- ✗ Microsoft Cognitive Toolkit
- ✗ Paddle

Applying Machine Learning

✗ Deep Learning Frameworks

- ✗ TensorFlow
- ✗ Microsoft Cognitive Toolkit
- ✗ Paddle

✗ Cloud Solutions

- ✗ Google Cloud ML
- ✗ Microsoft Azure
- ✗ Amazon Machine Learning

Applying Machine Learning

✗ Deep Learning Frameworks

- ✗ TensorFlow
- ✗ Microsoft Cognitive Toolkit
- ✗ Paddle

✗ Cloud Solutions

- ✗ Google Cloud ML
- ✗ Microsoft Azure
- ✗ Amazon Machine Learning

Machine Learning Libraries

- ✓ scikit-learn
- ✗ Mahout, Weka, Vowpal Wabbit

SecuML: Beyound scikit-learn

Scikit-learn

- Classification, clustering, dimension reduction, etc.
- Scaling, grid search, cross validation, etc.

SecuML: Beyound scikit-learn

Scikit-learn

- Classification, clustering, dimension reduction, etc.
- Scaling, grid search, cross validation, etc.

SecuML

- Automation of the Machine Learning pipeline
- Interactive labelling to acquire training data at low cost
- Graphical User Interface

SecuML

Machine Learning for Computer Security Experts

- Algorithms (`scikit-learn`, `metric-learn`, active learning)
- Web user interface (`flask` server)

Any Type of Data

PDF

DOC

PCAP

JavaScript

EXE

Netflows

SecuML - Input Data

features.csv

```
id,f0,f1,f2,f3,f4,...
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

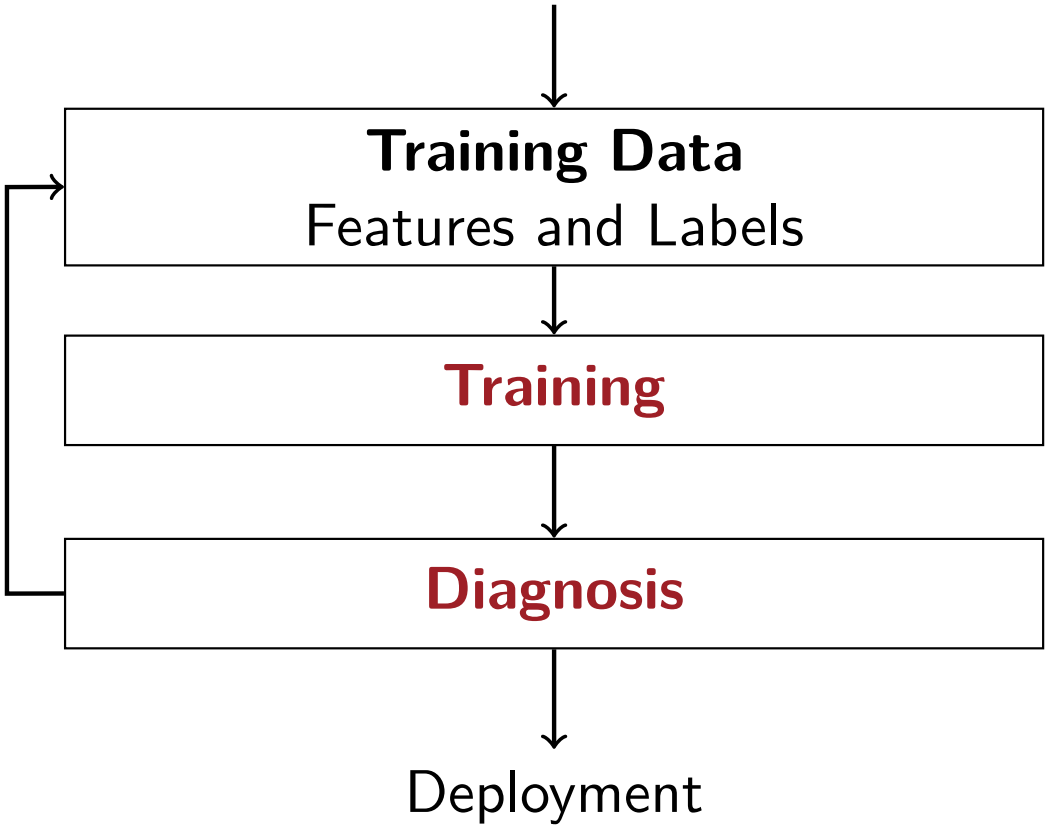
```
id,label,family
0,M,CVE-2017-30-10
1,M,CVE-2017-30-10
2,B,slides
3,M,CVE-2016-0945
4,M,CVE-2016-0945
5,B,user_manual
6,B,user_manual
7,B,technical_report
8,B,technical_report
...
```

SecuML

1. Set up a Detection Model
2. Acquire a Representative Training Dataset at Low Cost

Set up
a Detection Model

Set up a Detection Model



Automation of the Machine Learning Pipeline

Training Pipeline

- Scaling
- Cross validation to select the hyperparameters

Automation of the Machine Learning Pipeline

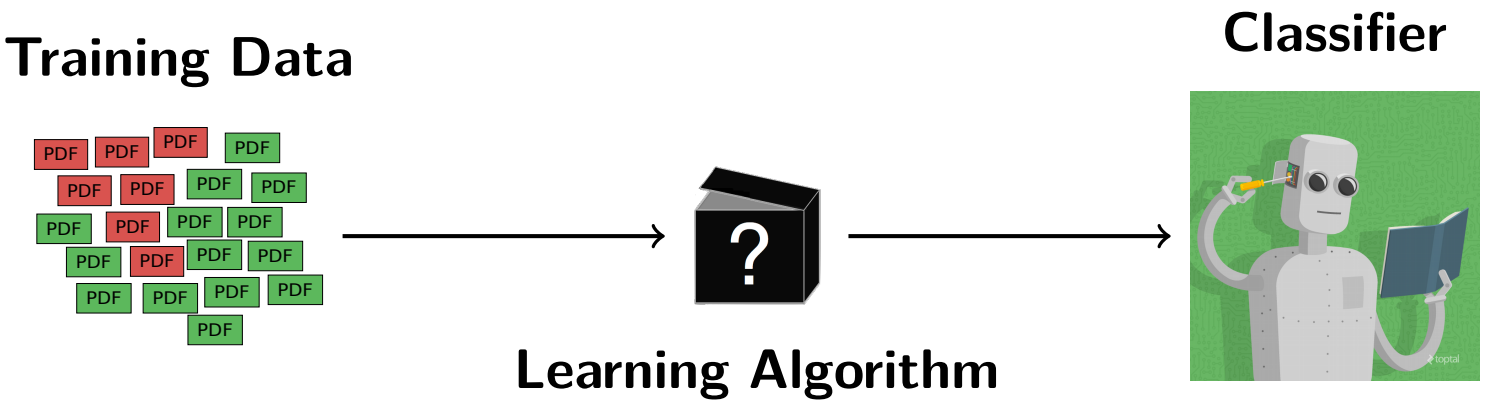
Training Pipeline

- Scaling
- Cross validation to select the hyperparameters

Validation of a Detection Model

Training	Validation
90% $Data$	10% $Data$

Trust in the Detection Model

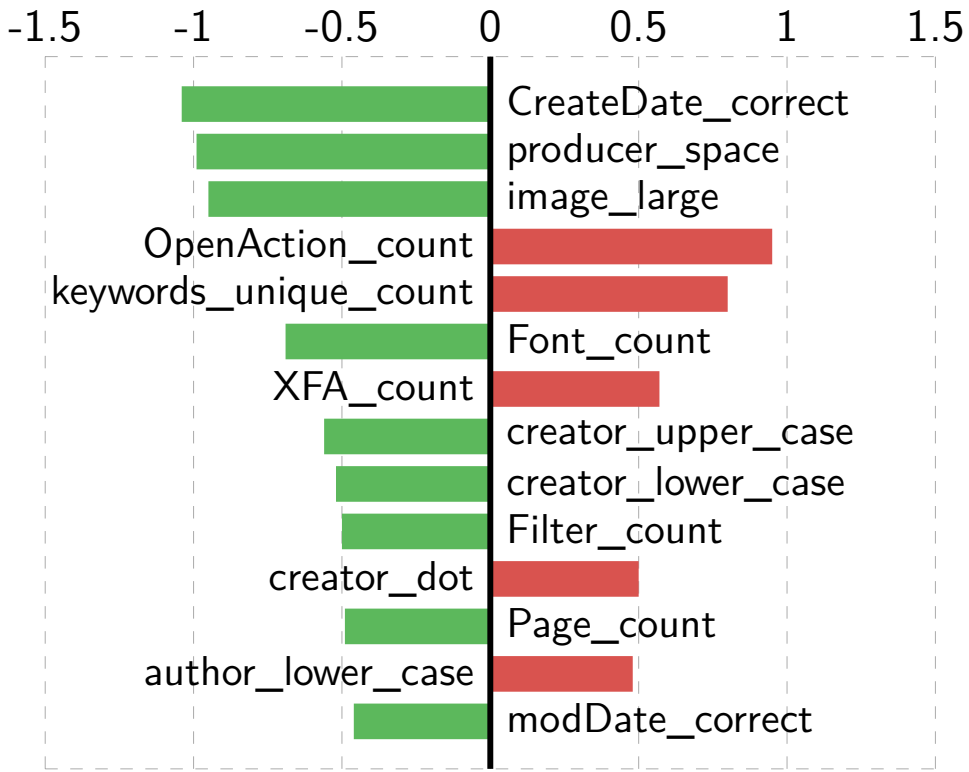


Understanding the Classifier

- How does the detection model work ?
- Why an alert has been raised ?

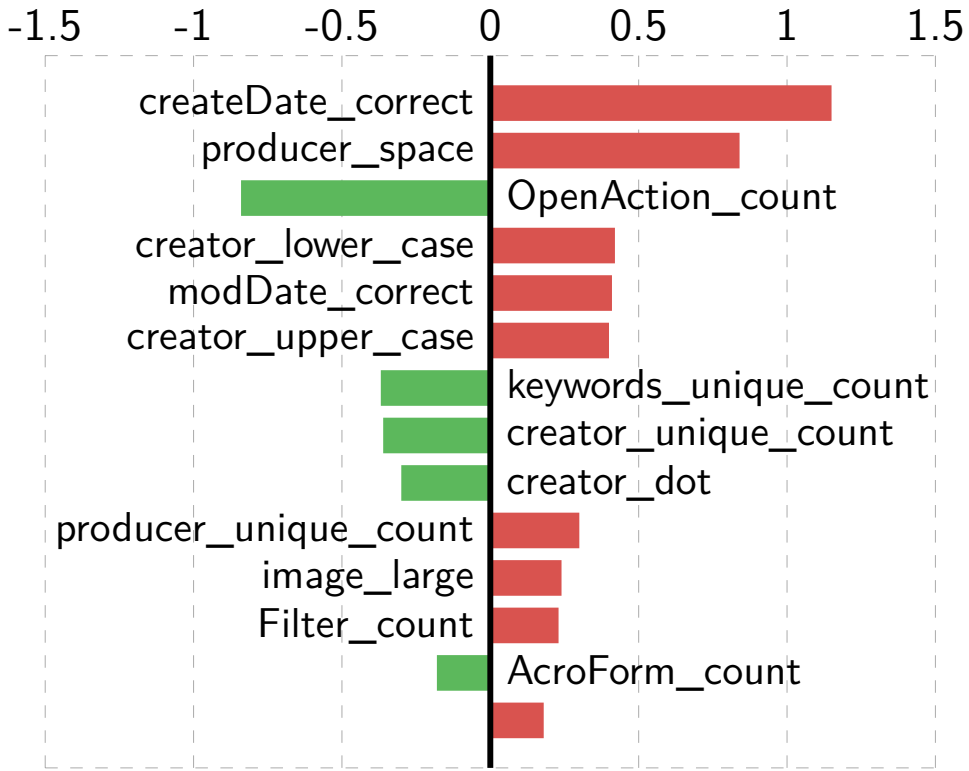
Trust in the Detection Model

How does the detection model work ?



Trust in the Detection Model

Why an alert has been raised ?



Train and Diagnose a Detection Model

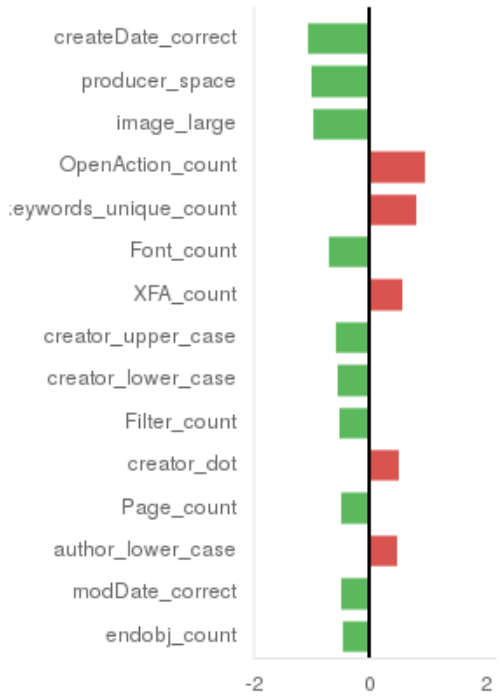
Demo

```
./SecuML_classification LogisticRegression PDF contagio
```

Train and Diagnose a Detection Model

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

Train and Diagnose a Detection Model

Train and Diagnose a Detection Model

Experiment

Model Coefficients

Train

Performance | Predictions

Indicators | ROC | Confusion Matrix

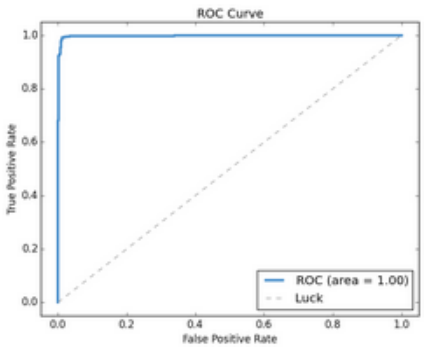
Detection threshold: 42%

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

Test

Performance | Predictions

Indicators | ROC | Confusion Matrix



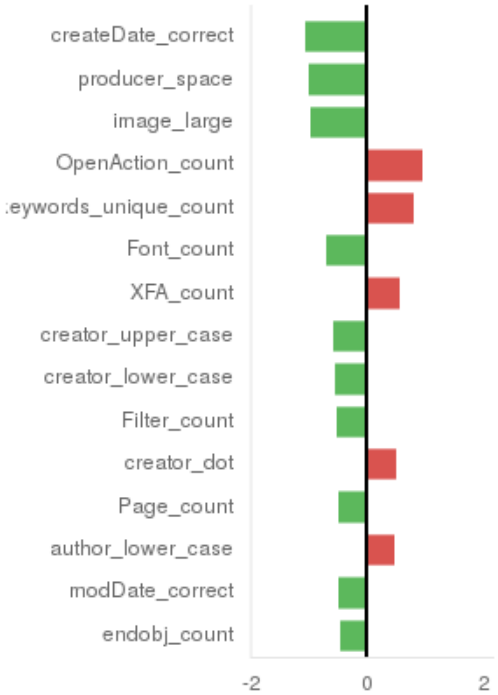
Alerts Analysis

Top N | Random

Train and Diagnose a Detection Model

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

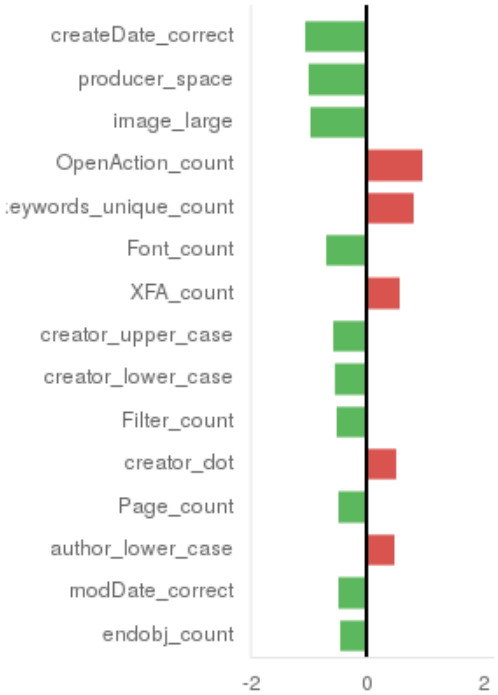
Alerts Analysis

Top N | Random

Train and Diagnose a Detection Model

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

Train and Diagnose a Detection Model

False Positives

Undetected

False Positives

Error 1 / 11 Prev Next

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

Description

Pdf | Features | Weighted Features

```
%PDF-1.3
%
4 0 obj
<< /Length 5 0 R /Filter /FlateDecode >>
stream
x
4
nJt
k
endstream
endobj
5 0 obj
```

Train and Diagnose a Detection Model

False Positives

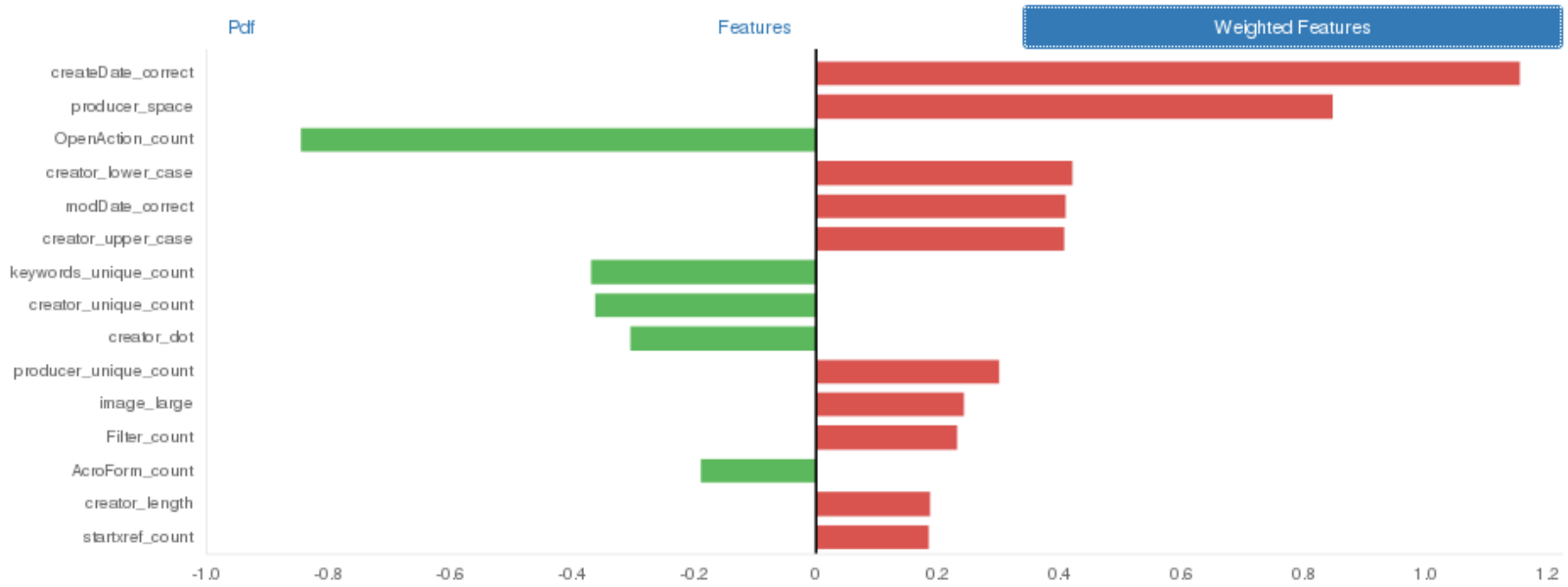
Undetected

False Positives

Error 1 / 11 Prev Next

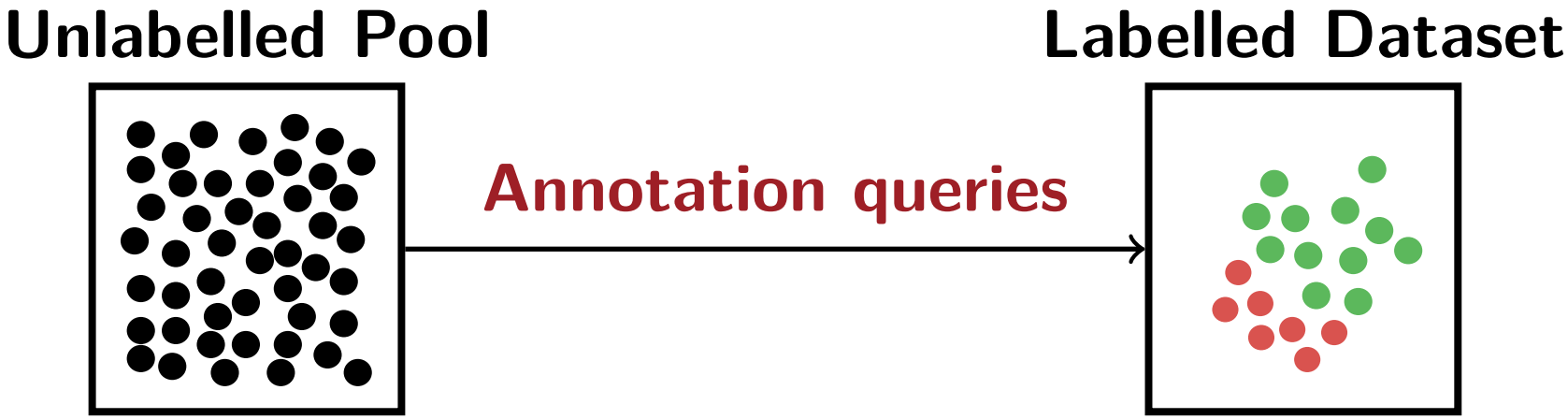
Instance 11530: Contagio/CLEAN/CLEAN_PDF_9000_files/Barbara_Harmon.pdf

Description



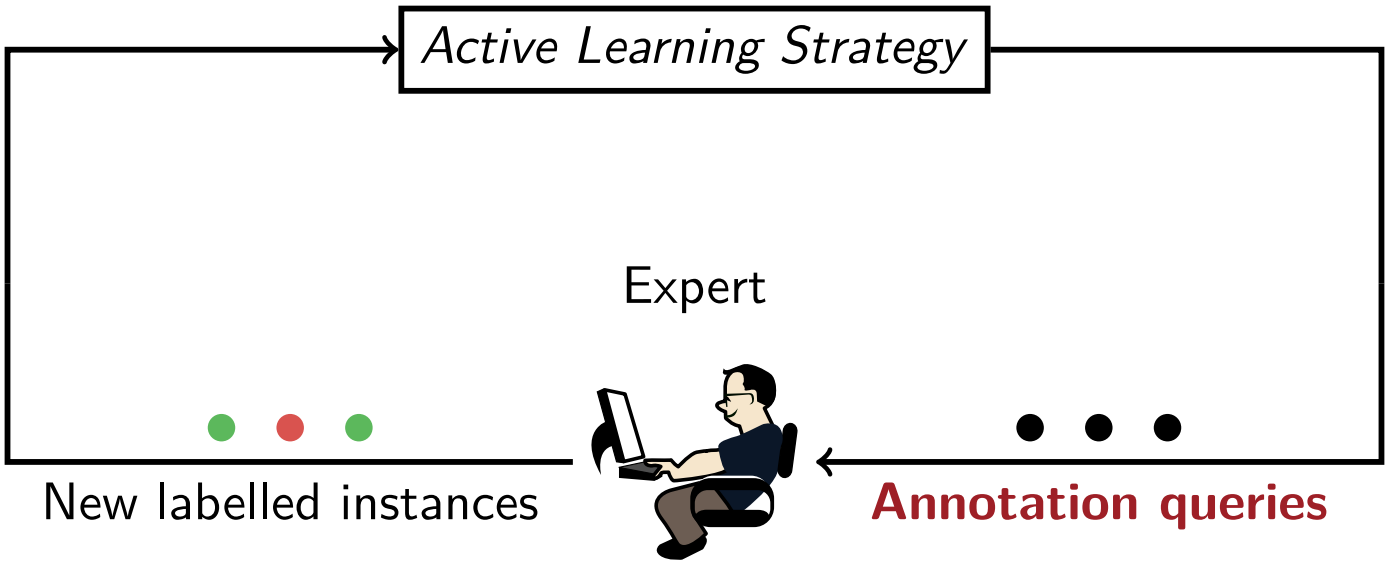
Acquire a Representative Training Dataset at Low Cost

Annotation System



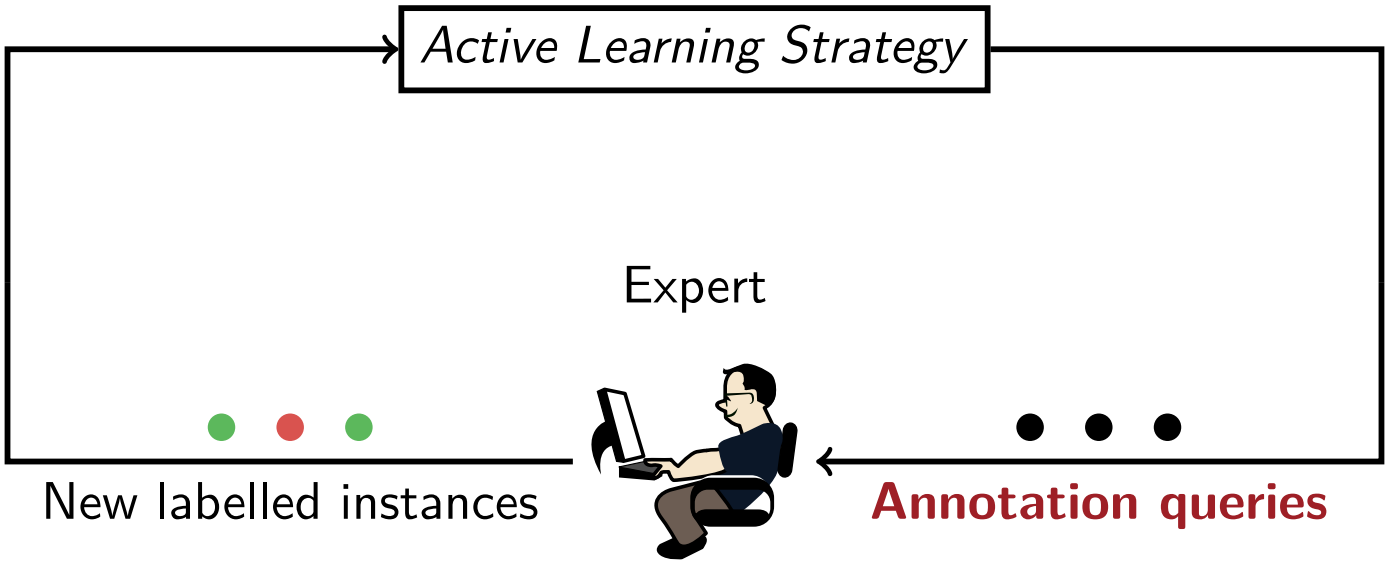
Security experts = expensive resources

Active Learning Strategy



Which instances should be annotated ?
Which instances are the most informative ?

Reducing the number of annotations is not enough !



- Low expert waiting time
- Feedback: "Your annotations are useful !"
- User interface for annotating

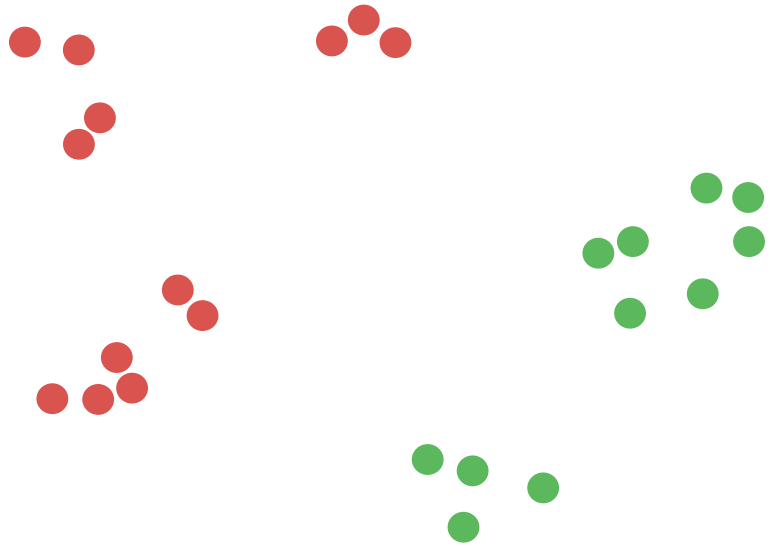
ILAB: Interactive LABelling

A whole annotation system

- Active learning strategy
- Feedback to the expert
- User interface for annotating

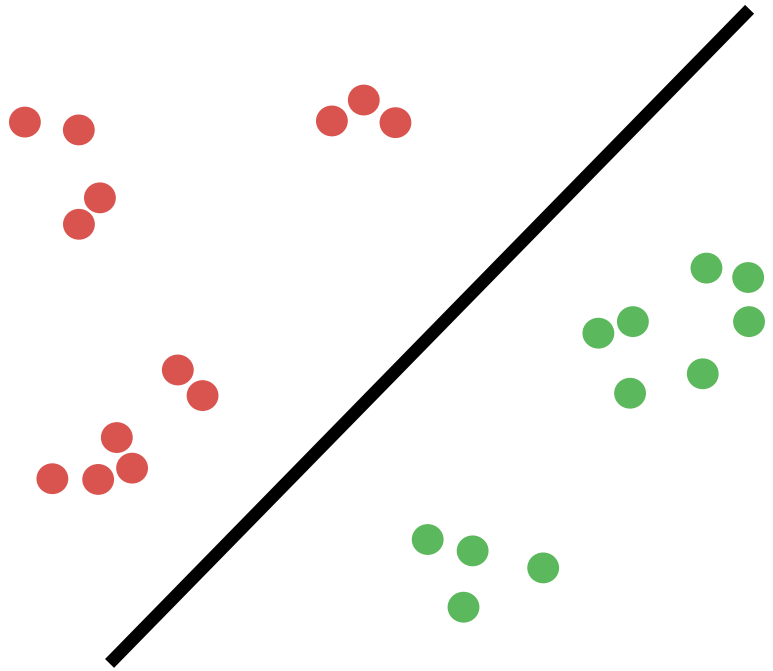
ILAB's Active Learning Strategy

Annotations Queries



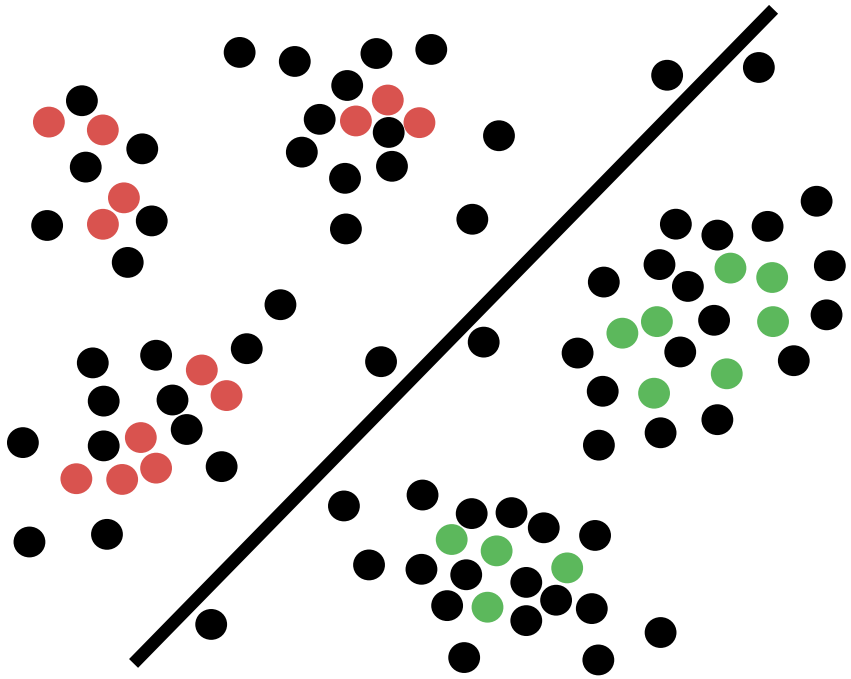
ILAB's Active Learning Strategy

Annotations Queries



ILAB's Active Learning Strategy

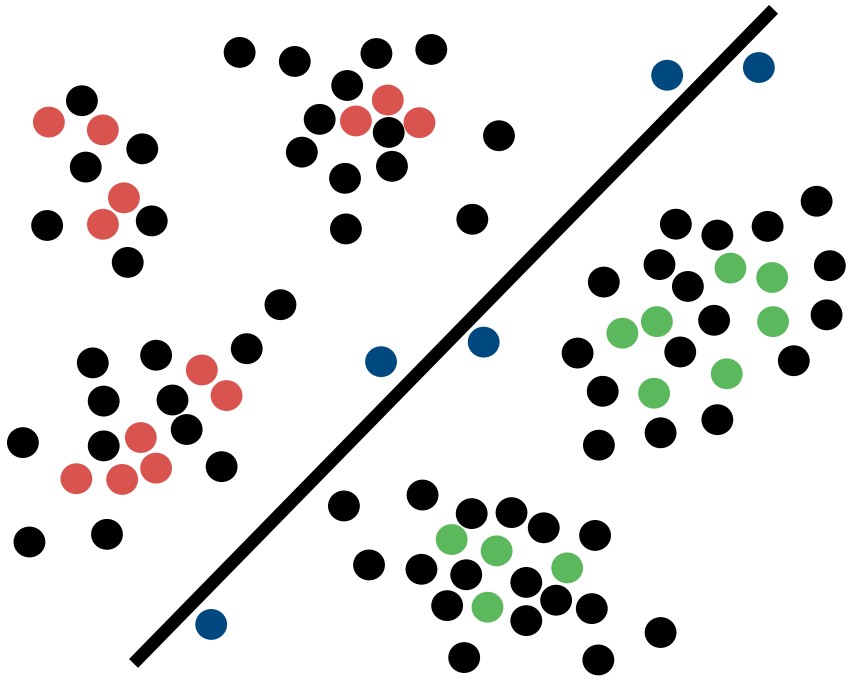
Annotations Queries



ILAB's Active Learning Strategy

Annotations Queries

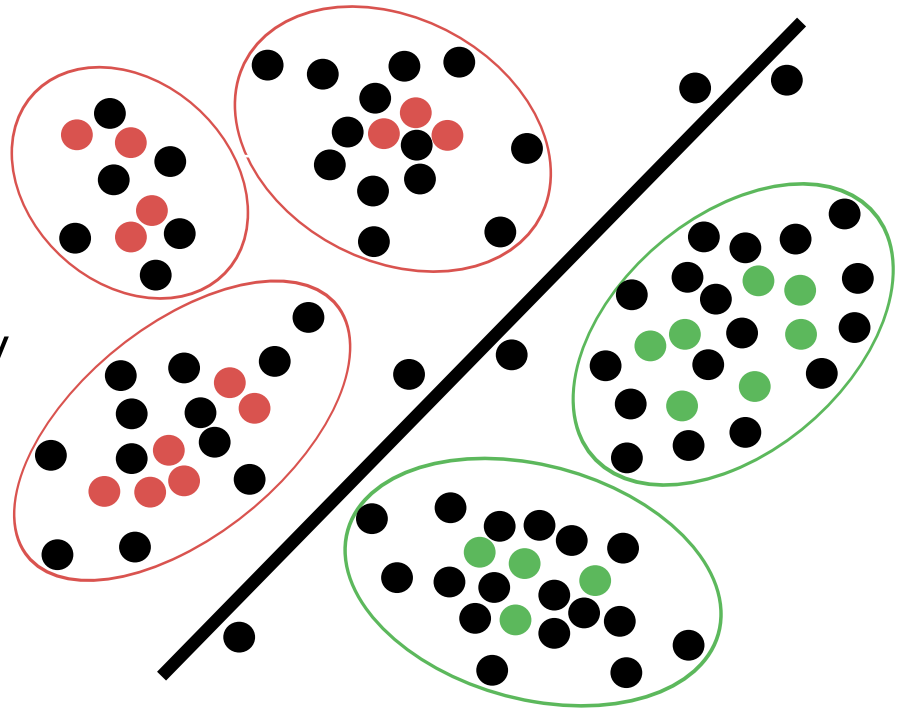
- Close to the decision boundary



ILAB's Active Learning Strategy

Annotations Queries

- Close to the decision boundary

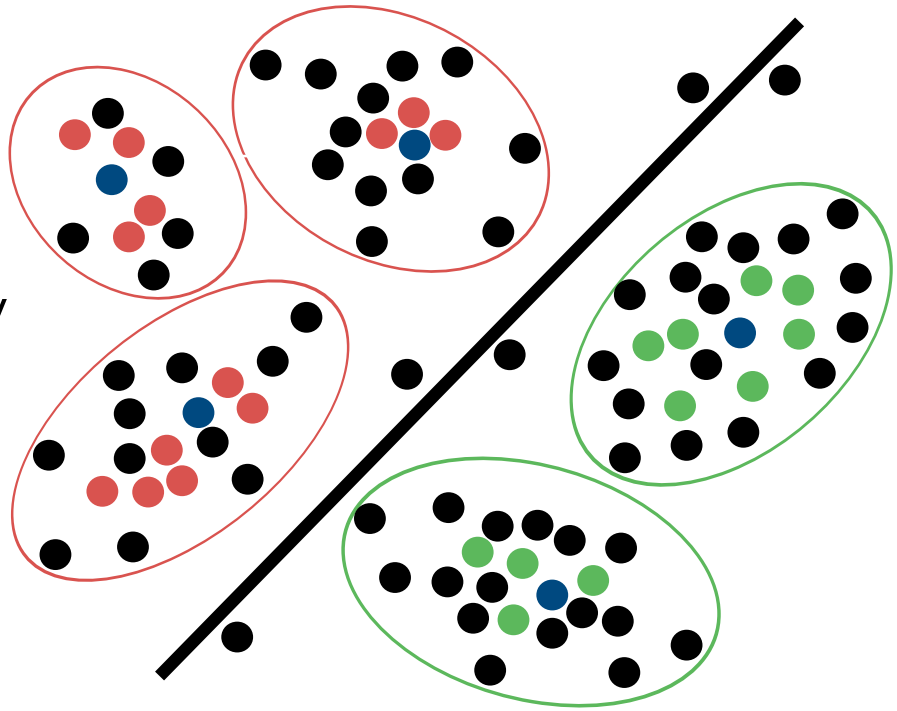


Clusters = User-defined Families

ILAB's Active Learning Strategy

Annotations Queries

- Close to the decision boundary
- Center of the clusters

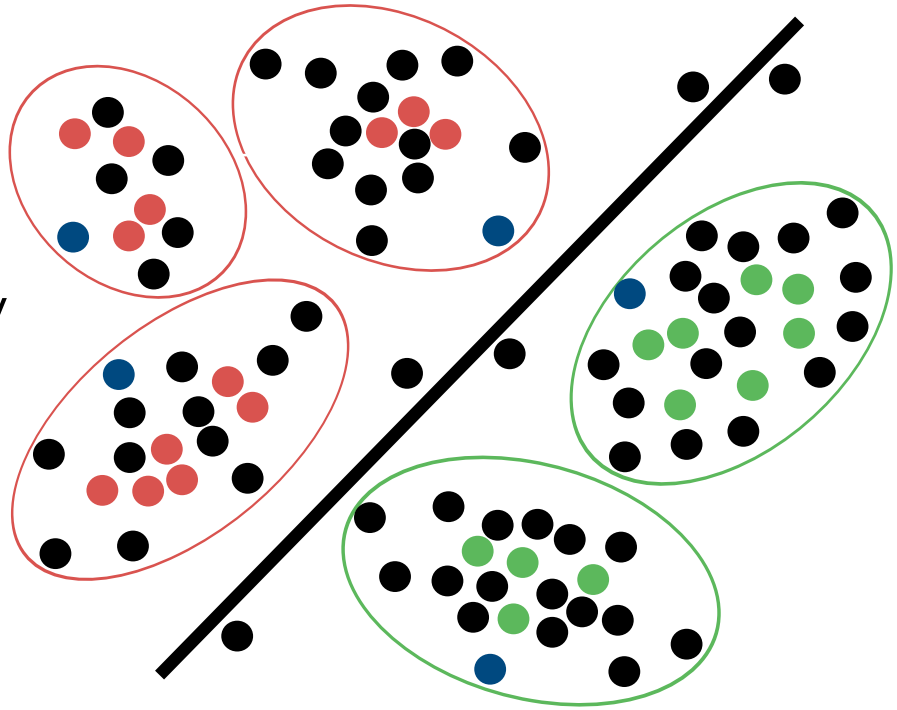


Clusters = User-defined Families

ILAB's Active Learning Strategy

Annotations Queries

- Close to the decision boundary
- Center of the clusters
- Edge of the clusters



Clusters = User-defined Families

Will be published at RAID 2017.

ILAB

Demo

```
./SecuML_activeLearning ILAB PDF contagio
```

ILAB

Experiment

Select an Iteration

Iterations

1
2
3
4
5
6

Annotating

Annotation Progress

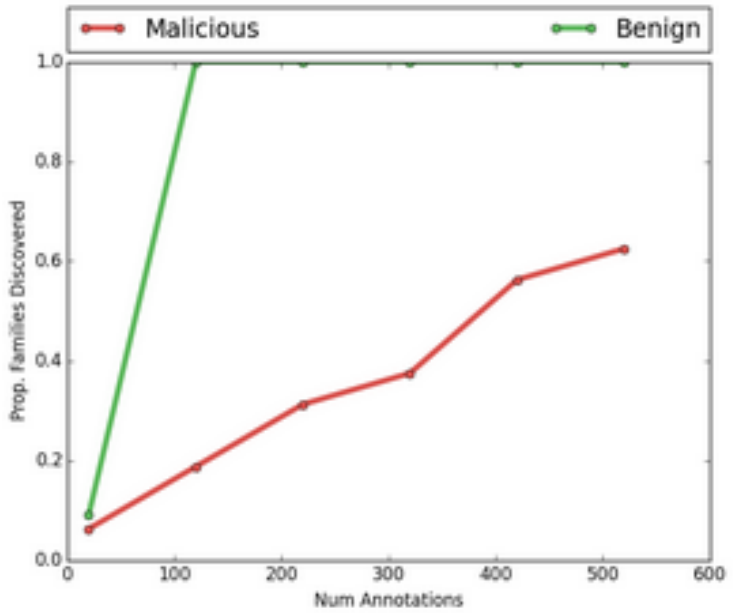
- 520 annotations
- 8510 unlabeled instances

Annotated Instances

Family Editor

Evolution Monitoring

Families | Suggestions



ILAB

Uncertain ➤➤➤ Malicious ➤➤➤ Benign Next Iteration

Annotation Queries

Annotation Query 1 / 10 Prev Next

Instance 14949: Contagio/EVIL/MALWARE_PDF_PRE_04-2011_10981_files/c3f2835c93e0f2b4d2244c3befda5856cd6d292f

Annotation

Suggestion	Annotation
None	CVE-2014-1453

Malicious Families

- CVE-2013-9453
- CVE-2014-1453
- CVE-2014-7239
- CVE-2014-9658
- CVE-2015-0346
- CVE-2015-1236

Add

Benign Families

- CV
- academic_paper
- administrative_forms
- e_ticket
- image
- invoice

Add

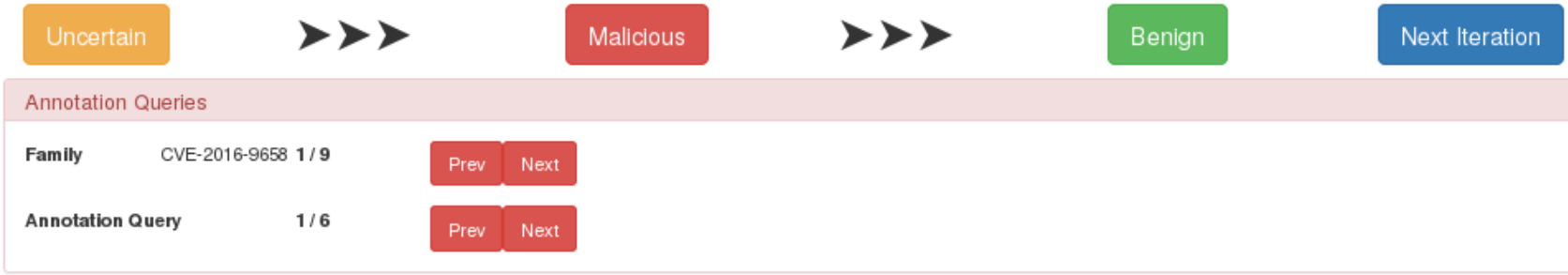
Ok Remove

Description

Pdf Features

```
%PDF-.3
%ÐÐí
1 0 obj<</Type/Catalog/Outlines 2 0 R/Pages 3 0 R/OpenAction 6 0 R>>endobj
2 0 obj<</Type/Outlines/Count 0>>endobj
3 0 obj<</Type/Pages/Kids[4 0 R]/Count 1>>endobj
4 0 obj<</Type/Page /Annots[ 5 0 R ]/Parent 3 0 R/MediaBox [0 0 612 792]>>endobj
5 0 obj<</Type/Annot /Subtype /Text /Name /Comment/Rect[25 100 60 115] /Subj 8 0 R
6 0 obj<</Type/Action/S/JavaScript/#4AS 7 0 R>>endobj
7 0 obj<</Length 216/Filter/#46#6C>>
stream
x-ÐÐnÐ DÐÐWZÐ#žÐÐ?ÐÐ#ÐÐÐieÊÐÐÐÐÐÐÐfÐÐÐÐxÐÐÐGÐ(Ð»GGZÐc:QÐØ]Ð
#ÐZGÐÐÐÐÐÐÐJÐGÐ□ÐÐOg=Ðc ÐÐÐCÐÐÐ4SGÐÐÐÐDFÐÐ;Ð<$Ð*Ð`ÐtÐ~Ð忻-ÐÐt4Ð\ÐÐ d_&jRÐNIÐ
endstream
objend
8 0 obj<</Length 10823/Filter/#46#6C>>
stream
xÐÐ[ÐÐHÐ+Ð)eÐÐ5Ðzt ÐÐ.ÐÐVJJI]□FzÐÐpxÐÐ(ÐÐÐÐÐsÐÐOoÐ?Ð}Ð□OÐÐ>xÐ$Ð<ÐÐÐÐxÐ>ÐÐÐÐÐÇÐ{ÐÐ
>8x>8_ÐÐUÐÐÐUÐGsÐÐÐÐÐDÐÐÐÐ5Lk_ÐÐÐÐÐ.uÐÐÐÐÐ?ÐÐGÐÐÐÐÐ<ÐÐ8Ðo~pÐ□ÐÐÐ/ÐÐ·ÐÐ_XÐÐÐÐ/ÐÐÐs
ÐÐÐCÐ#[ÐD`ÐeÐW9ÐÐ'Ð?ÐÐhÐ9ÐÐ□mÐÐGÐ!ÐÐÐsÐZ%Ð(*r#ÐÐhÐÐØv]x□ÐC+ÐÐÐQ ÐÐÐGÐSÐÐk
ÐÐÐÐDF>ÐTÐ.ÐHÐRwÐ;ÐlujÐIÐÐDÐyÐÐÐÐzÐÐÐ7Ð(ÐÐMÐÐXGGÐÐÐGÐÐ1LÐÐ)ÐÐzÐvÐ*ÐQÐÐÐÐ-Ð"hv'ÐÐÐ
```

ILAB


Uncertain
Malicious
Benign
Next Iteration
Annotation Queries
Family CVE-2016-9658 1 / 9
Prev
Next
Annotation Query 1 / 6
Prev
Next
Instance 5877: Contagio/EVIL/MALWARE_PDF_PRE_04-2011_10981_files/acc67da843edfcf33bdc8766d83846a6156f98d5
Annotation
Suggestion
CVE-2016-9658
Malicious Families
CVE-2013-9453
CVE-2014-1453
CVE-2014-7239
CVE-2014-9658
CVE-2015-0346
CVE-2015-1236
Benign Families
CV
academic_paper
administrative_forms
e_ticket
image
invoice
Add
Add
Ok
Remove
Description
Pdf
Features
%PDF-1.6
%
8 0 obj<</Length 2799 /Filter/FlateDecode>>
stream
endstreamendobj16 0 obj<</MediaBox [0 0 595 842] /Contents 8 0 R /Parent 11 0 R /
stream

ILAB

Experiment

Select an Iteration

Iterations

1
2
3
4
5
6

Annotating

Annotation Progress

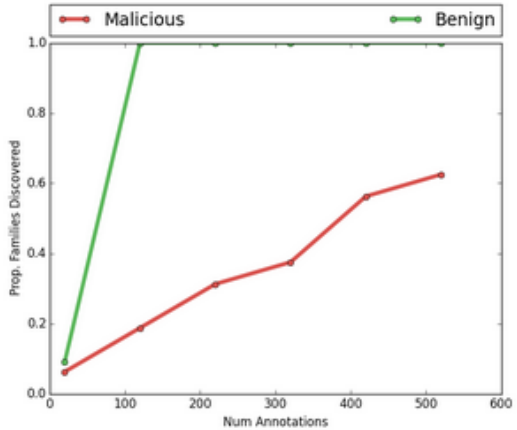
- 520 annotations
- 8510 unlabeled instances

Annotated Instances

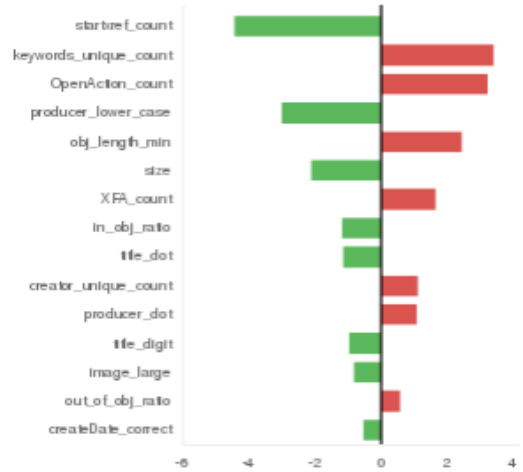
Family Editor

Evolution Monitoring

Families | Models | Suggestions



Model Coefficients



Train

Performance | Predictions

Indicators | ROC | Confusion Matrix

Detection threshold: 50%

Detection	100.0%
False alarms	0.0%
F-score	100.0%
AUC	100.0%

Cv

Performance | Predictions

Indicators | ROC | Confusion Matrix

Detection threshold: 50%

Indicator	Mean	Std
Detection	98.36%	0.0208
False alarms	1.78%	0.0068
F-score	97.56%	0.0163
AUC	99.43%	0.0045

Test

Performance | Predictions

Indicators | ROC | Confusion Matrix

Detection threshold: 50%

Detection	97.27%
False alarms	1.26%
F-score	92.36%
AUC	98.93%

SecuML

SecuML

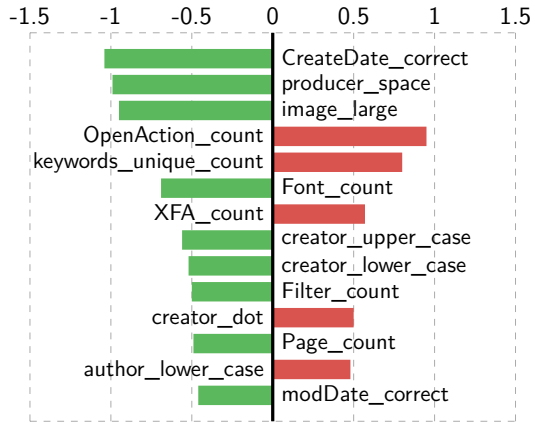
1. Acquire a Representative Training Dataset at Low Cost
2. Set up a Detection Model

What's next ?

- GUI to launch the experiments

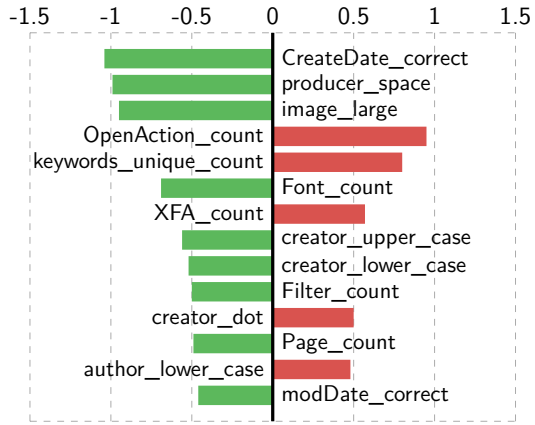
What's next ?

- GUI to launch the experiments
- Interpretation of more complex models

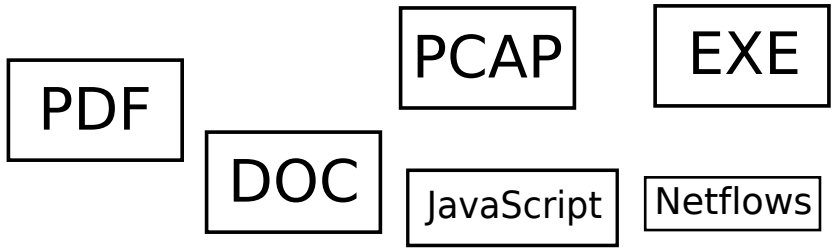


What's next ?

- GUI to launch the experiments
- Interpretation of more complex models



- Automatic feature extraction



SecuML

Algorithms and Corresponding Interfaces !

Detection Models

- Logistic regression, SVM, Naive Bayes, ...
- Performance, interpretation

Interactive Machine Learning

- Active learning, Rare category detection
- Annotations, feedback

SecuML

Algorithms and Corresponding Interfaces !

Clustering

- K-means, Gaussian Mixtures, ...
- Instances in each cluster

Projection

- PCA, RCA, LDA, LMNN, ..
- Projection on two components

SecuML is available online !

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

Only for Computer Security experts ?

- Model interpretation
- Interactive labelling
- Data visualization with projections
- Clustering display

SecuML is available online !

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

Only for Computer Security experts ?

- Model interpretation
- Interactive labelling
- Data visualization with projections
- Clustering display