Secureworks®

Threat Class Predictor

François Labrèche

Serge-Olivier Paquette

Agenda

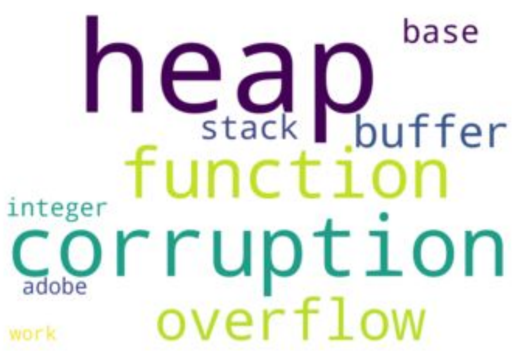
- Semantic representation of vulnerabilities
- Building an explainable threat score and trend score
- Dashboard
- Discussion
- Conclusion

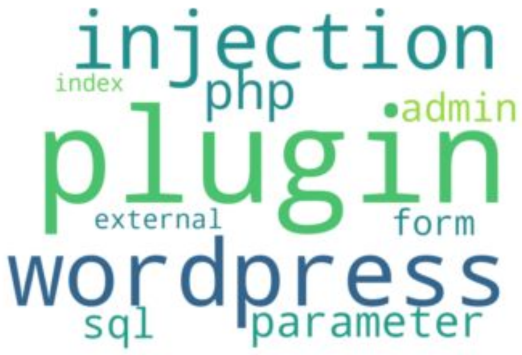
Goal

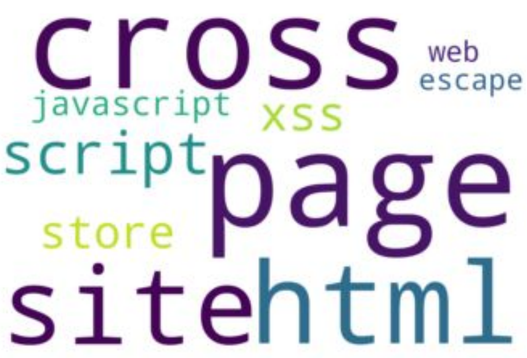
Build an explainable machine learning framework to predict threats associated with disclosed vulnerabilities

Semantic Representation

- Uses Topic Modeling, specifically Latent Dirichlet Allocation (LDA)
- Built on filtered vulnerability descriptions from the NIST
- 30 topics generated
- Vector of 30 weights for each vulnerability







Secureworks

Semantic Representation

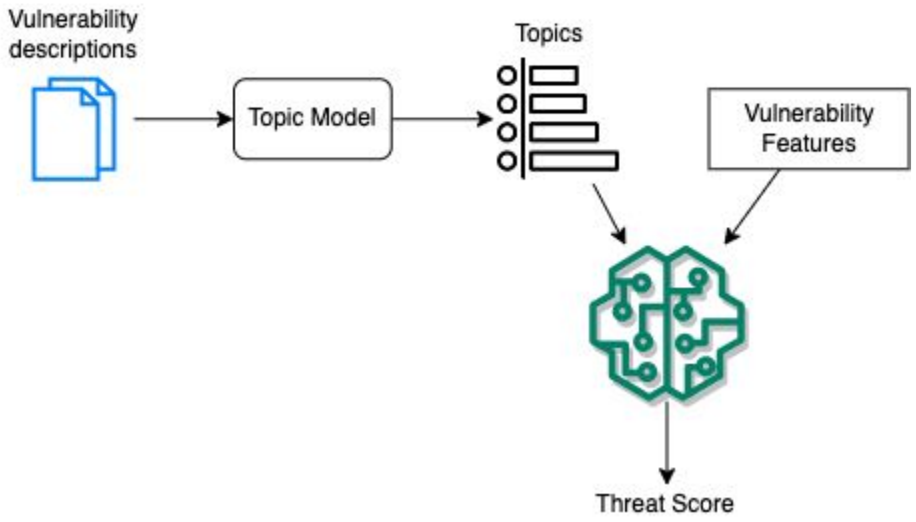
Topic vector example

Weight	Word
0.128	page
0.122	cross
0.121	html
0.118	site
0.109	script
0.070	xss
0.063	store
0.048	javascript
0.035	escape
0.034	web

Threat Score

Supervised Machine Learning Model

- One model trained per threat class
- Uses vulnerability data and the topics as features
- Predicts the likelihood of an attack



Secureworks®

Threat Score

Additional Features Used

- The length of the description
- The number of references available for the vulnerability at the time of publication
- The number of affected software configurations by this vulnerability
- The CVSSv2 score
- The CVSSv2 metrics
- Vulnerabilities from 2008 and up, from the NIST National Vulnerabilities Database

Dataset

Labels

Exploits

- ExploitDB
- Packetstorm
- Github POCs

Malware

- ClamAV signatures
- CTU malware reports

Dataset	N samples
CVE Database	152,585
ExploitDB	22,441
Packetstorm	5,471
Github POCs	3,219
ClamAV	2,956
CTU	184

Evaluation

Two approaches used for managing the unbalanced data

- Class weighting
- Threshold-moving on f2 score

Setup

- 10-fold cross validation
- Random Forest Classifier
- Gridsearch

Gridsearch Parameter	Exploits	Malware
max depth	30	50
min samples leaf	8	6
min samples split	22	16
n trees	300	200

Results

Model Performance

Metrics of interest

- Accuracy
- Recall
- F2 score

Priority goal

- Identify attacks, i.e., true positives
- Misrepresented False Positives

Results

Model Performance

Exploit Publication		Malware Inclusion	
Metric	**Value**	**Metric**	**Value**
Accuracy	**88.81%**	Accuracy	**98.01%**
Recall	**79.92%**	Recall	**87.96%**
Precision	36.92%	Precision	47.77%
F1-Score	50.51%	F1-Score	61.90%
F2-Score	64.82%	F2-Score	75.27%
Threshold	0.34	Threshold	0.46

Results

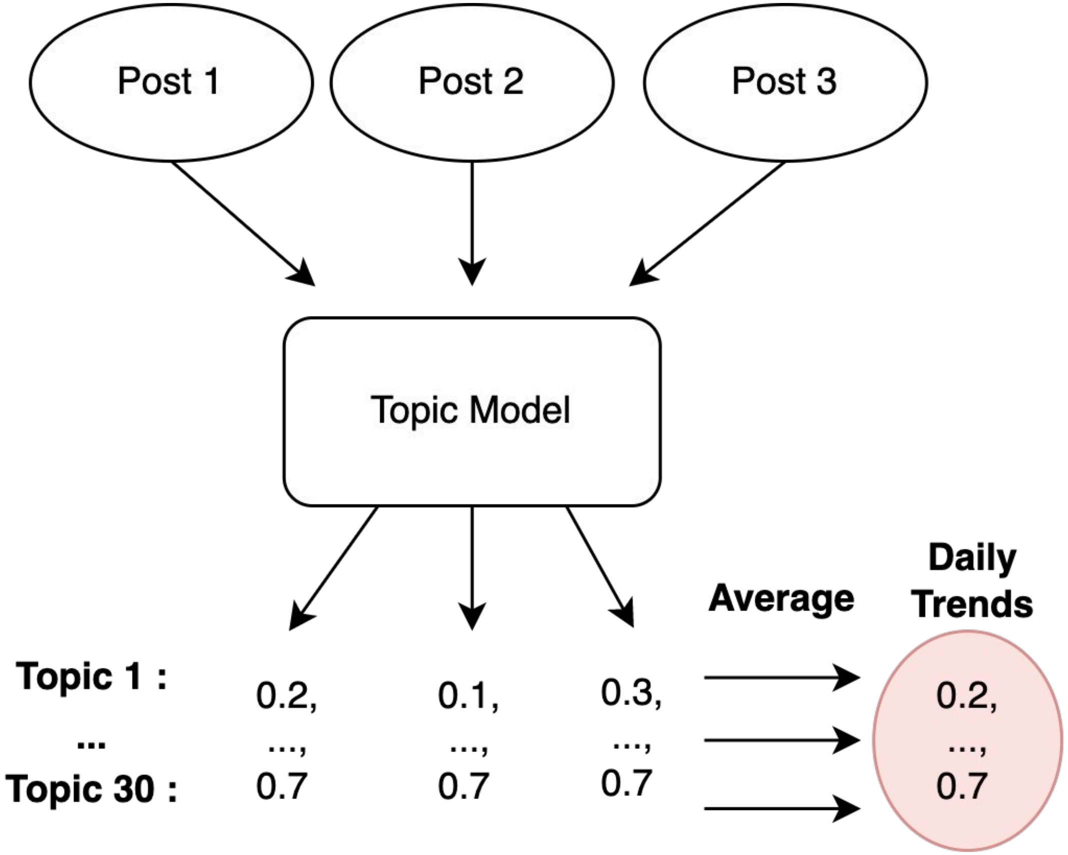
Model Performance

Exploit Publication		Malware Inclusion	
Metric	**Value**	**Metric**	**Value**
Accuracy	**88.81%**	Accuracy	**98.01%**
Recall	**79.92%**	Recall	**87.96%**
Precision*	36.92%	Precision*	47.77%
F1-Score	50.51%	F1-Score	61.90%
F2-Score	64.82%	F2-Score	75.27%
Threshold	0.34	Threshold	0.46

* False positives are misrepresented: they are higher due to our incomplete dataset

Trend Score

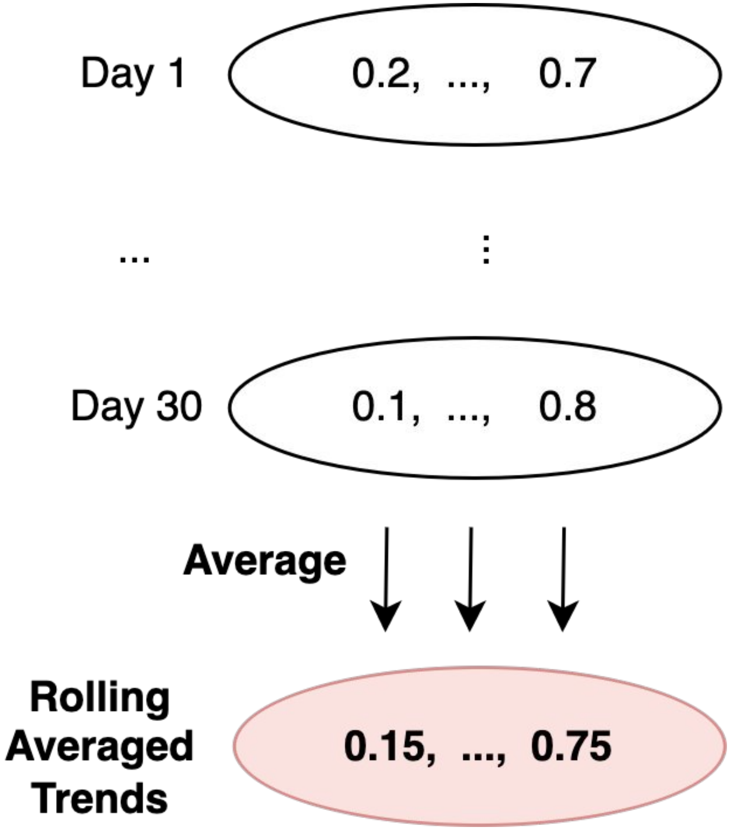
Previously trained topic model applied to social network data and dark web forum posts



Secureworks®

Trend Score

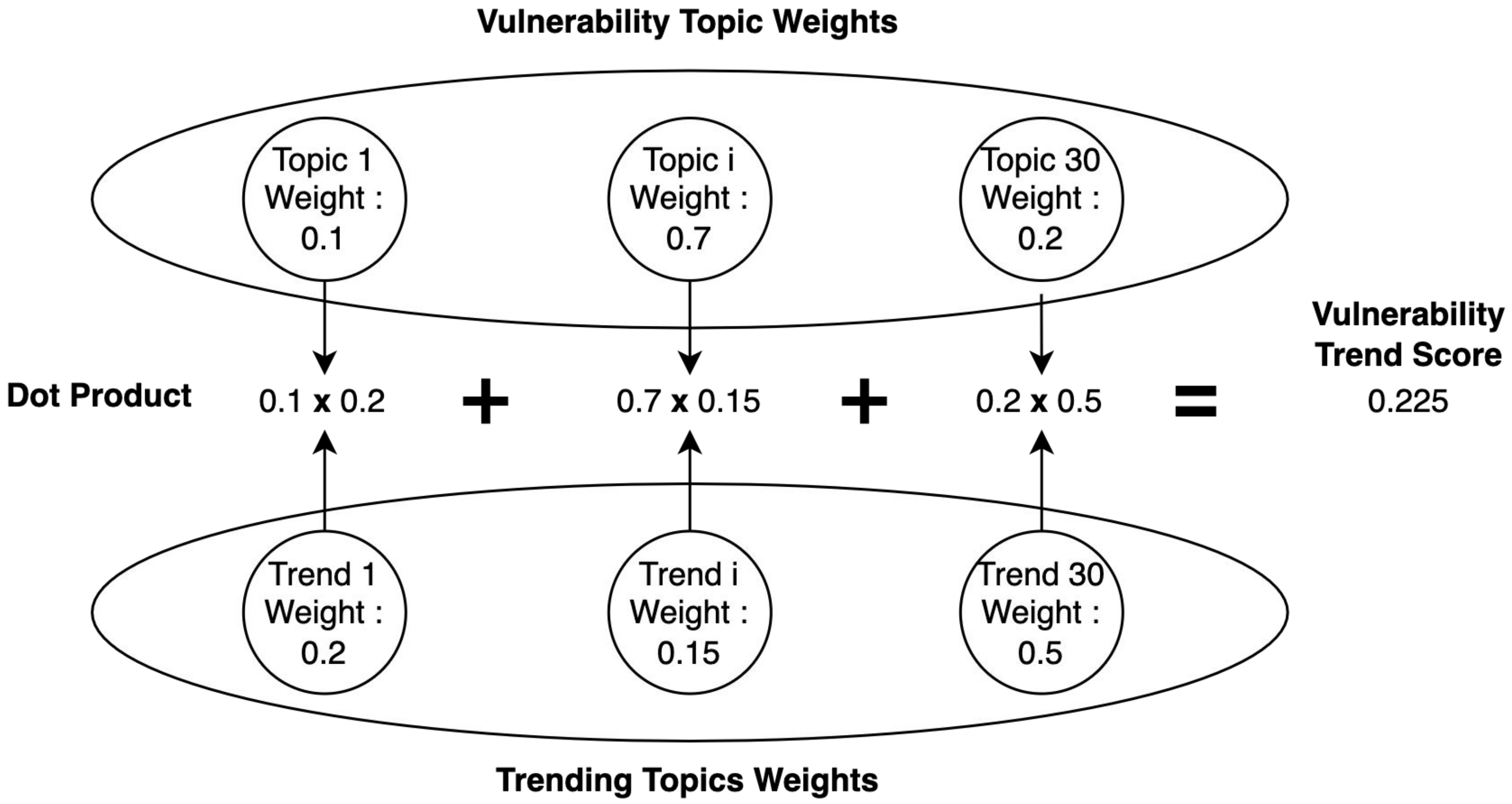
Trends are averaged between sources and over the past 30 days



Secureworks®

Trend Score

Vulnerabilities are linked to trends using a dot product

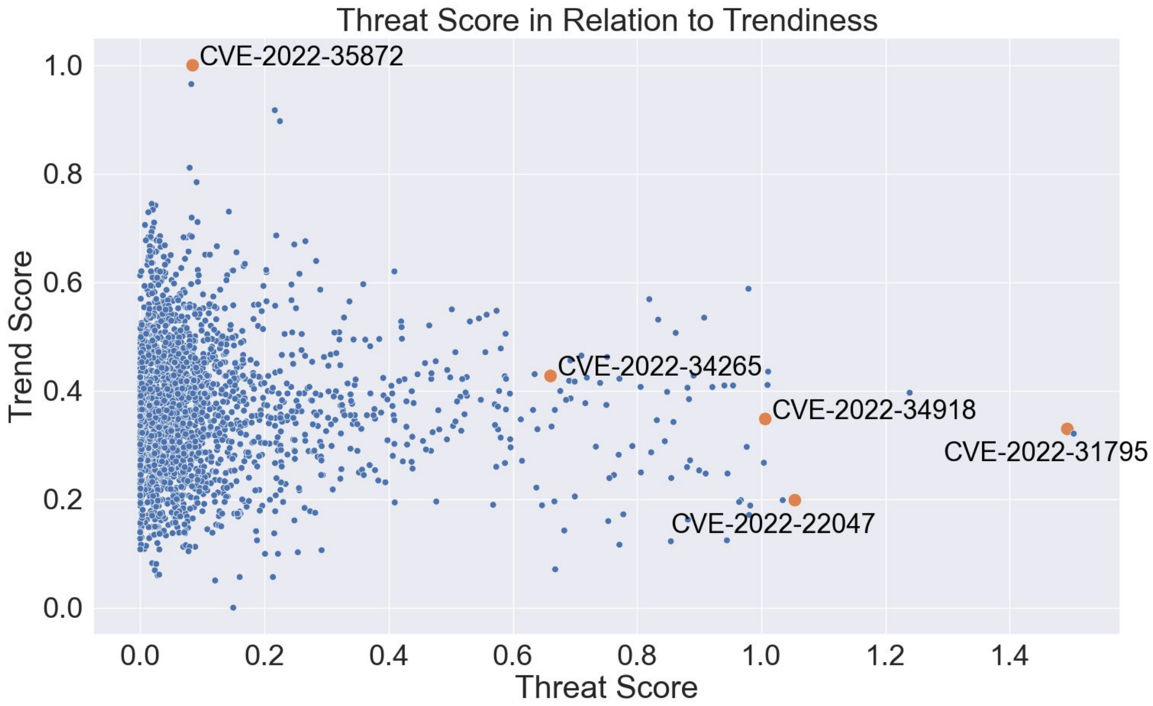


Visual Dashboard

Combination of both threat score and trend score

- X axis: Threat score
- Y axis: Trend score

With this dashboard, an analyst can identify vulnerabilities that go under the radar of what's trending.



Visual Dashboard

Vulnerability with predicted exploits:

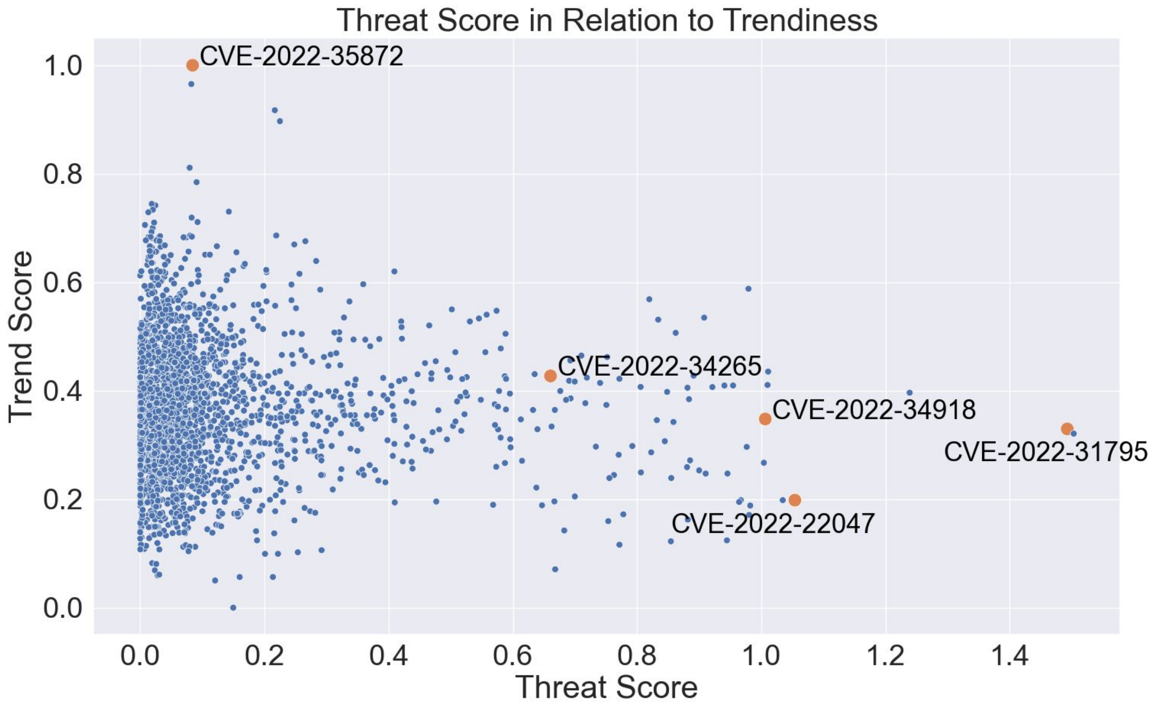
- CVE-2022-34265

Vulnerability with predicted malware:

- CVE-2022-22047

Vulnerability matching trends:

- CVE-2022-35872

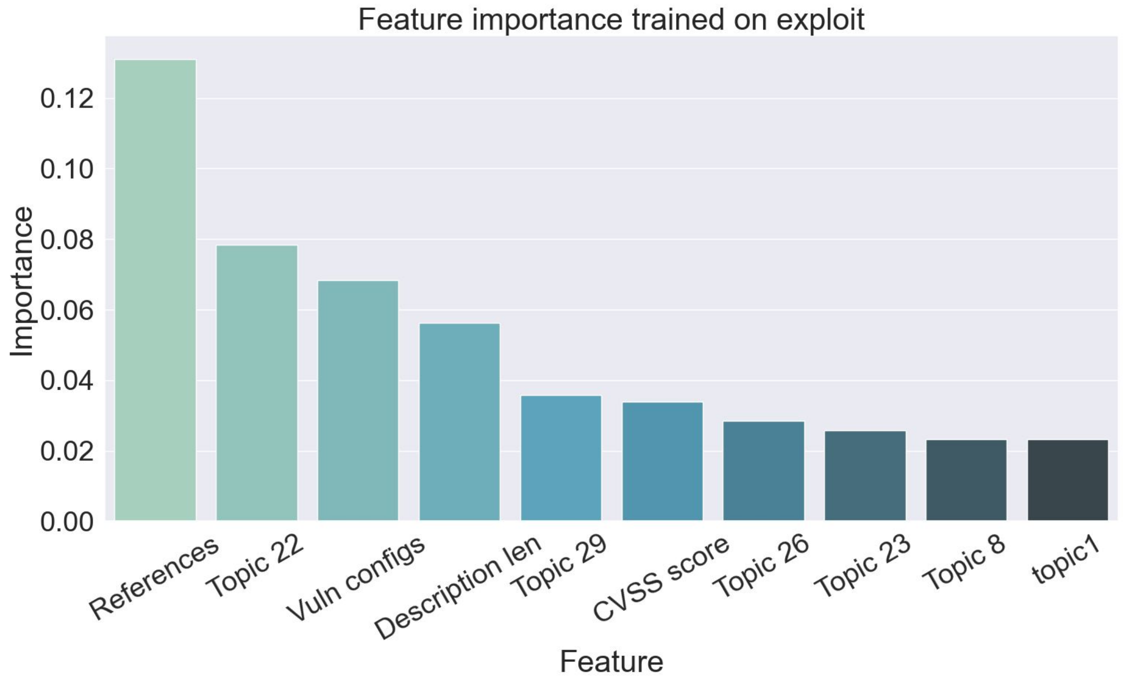


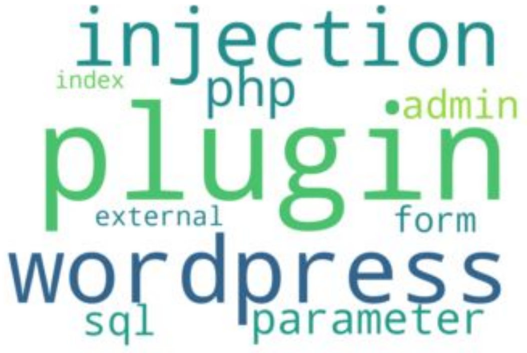
Discussion

An explainable framework

Top topics when predicting the publication of exploits

- Topic 22 - Parameter and SQL injections
- Topic 29 - Google and OAuth
- Topic 26 - Cross-Site Scripting (XSS)
- Topic 23 - Denial of Service (DOS)





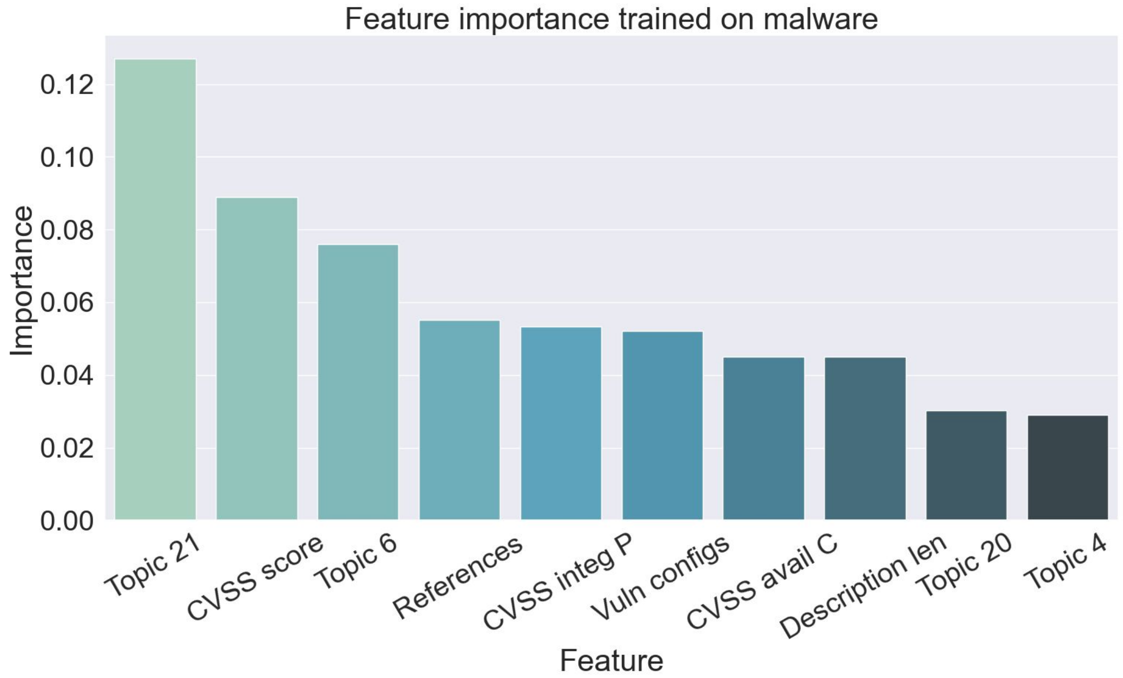
Secureworks®

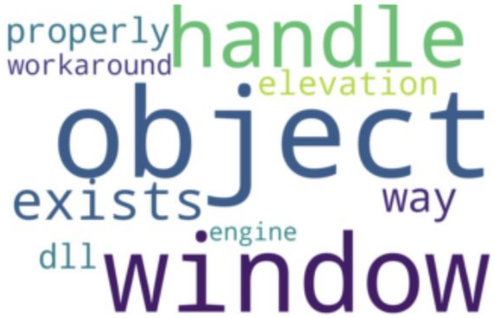
Discussion

An explainable framework

Top topics when predicting the inclusion of malware

- Topic 21 - Use of Windows handles
- Topic 6 - PDF vulnerabilities
- Topic 4 - Heap and buffer overflows





Conclusion

01 We presented a coherent and explainable framework to predict the threat associated with a vulnerability

02 Our results showcase vulnerabilities with a high likelihood of being included in real attacks that may be overlooked by the cybersecurity community

Secureworks®

Thank you

Questions?