Classifying Exploits in the Wild

Mateus Nogueira
(UFRJ)

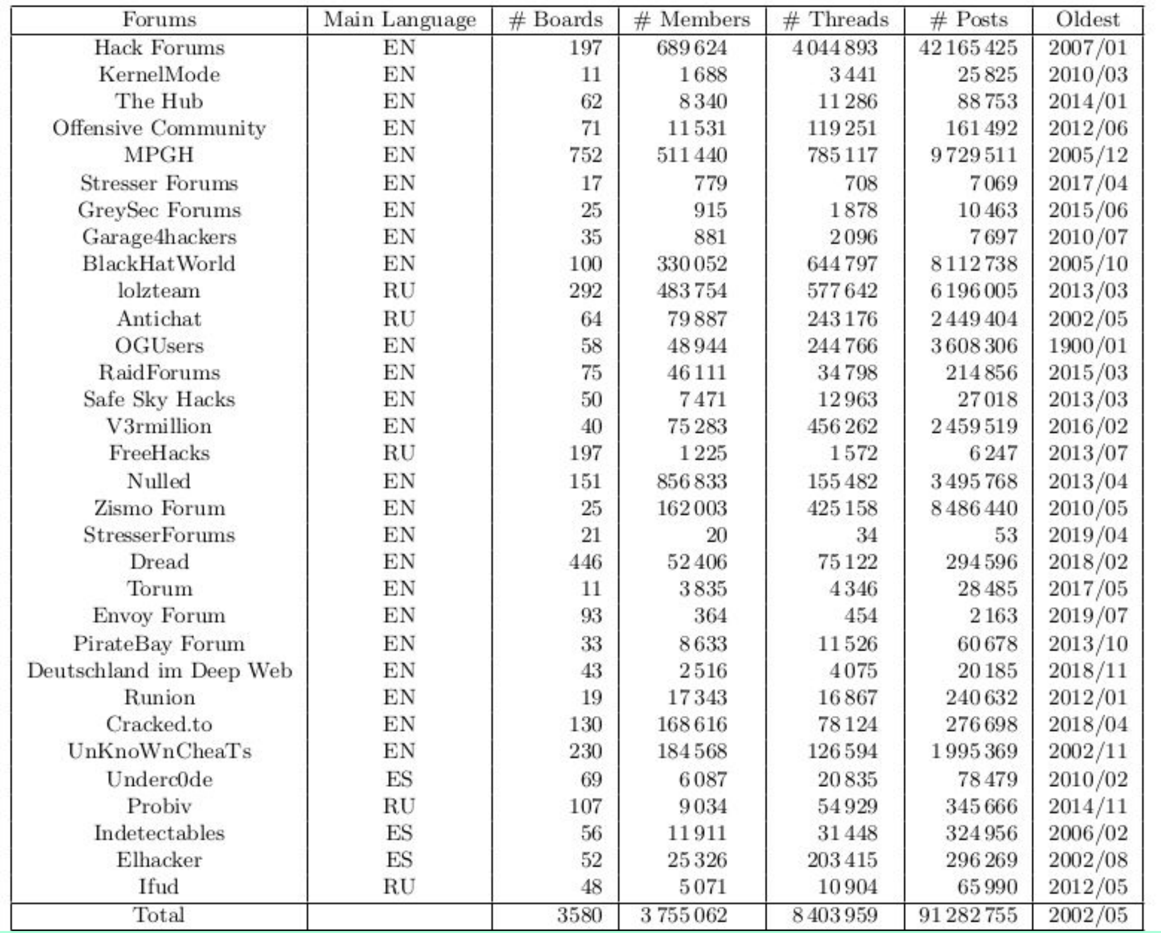
Research on CrimeBB and TI Feeds

- **Theme**: Alternative cybersecurity vulnerabilities information sources
- **Problem**: Is it possible to automatically classify vulnerabilities?
    - **Class 1:** vulnerabilities for which exploitation activity is available in crimeBB and TI Feeds
    - **Class 2:** others
- **Goal**:
    - Develop automatic classifier for **events (in TI Feeds)** and **posts or threads (in CrimeBB)**
        - Design heuristics
        - Train ML algorithm with text features and stuctured features
        - Evaluate algorithm

	CrimeBB	TIFeeds
Basic Elements	Post	Threat feed event
Structured features for machine learning	Number of posts in threads, **but no hint on groundtruth.**	Indicators of Compromise (IoCs) of threat feed basic elements, tags and **threat level**
Unstructured features for machine learning	Words of posts	Free-text description of event
Groups of basic elements	Threads	N/A
Element to be classified, i.e., instances for ML training and classification	Thread	Threat feed event
Source	Forum	Org. Contributor Id. [note: Org. Id. is always SIEMENS in our dataset, and Org. Id. is different from Org. Contributor Id.]
Expert signing the element	Author	N/A
Labeling information **(*target classes for ML*)**	poc, weaponization, exploitation	Exploitation vs no exploitation, assigned to incidents

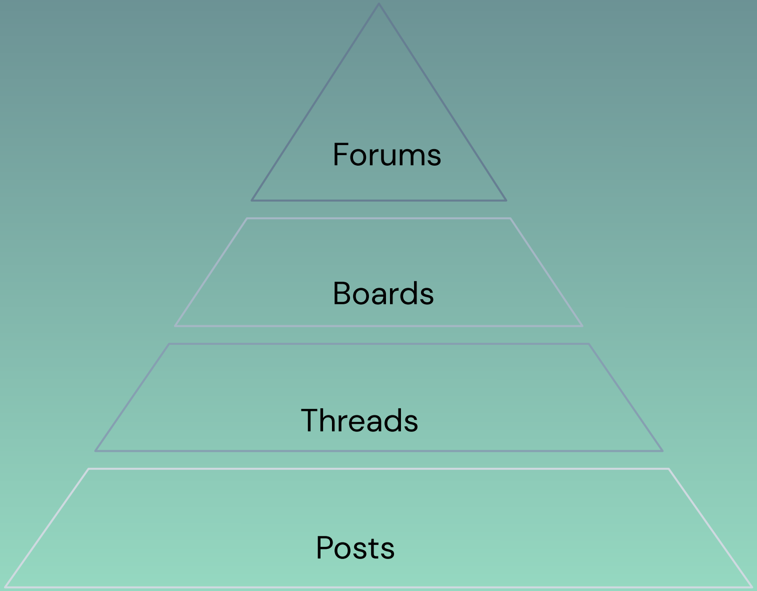
What is CrimeBB?

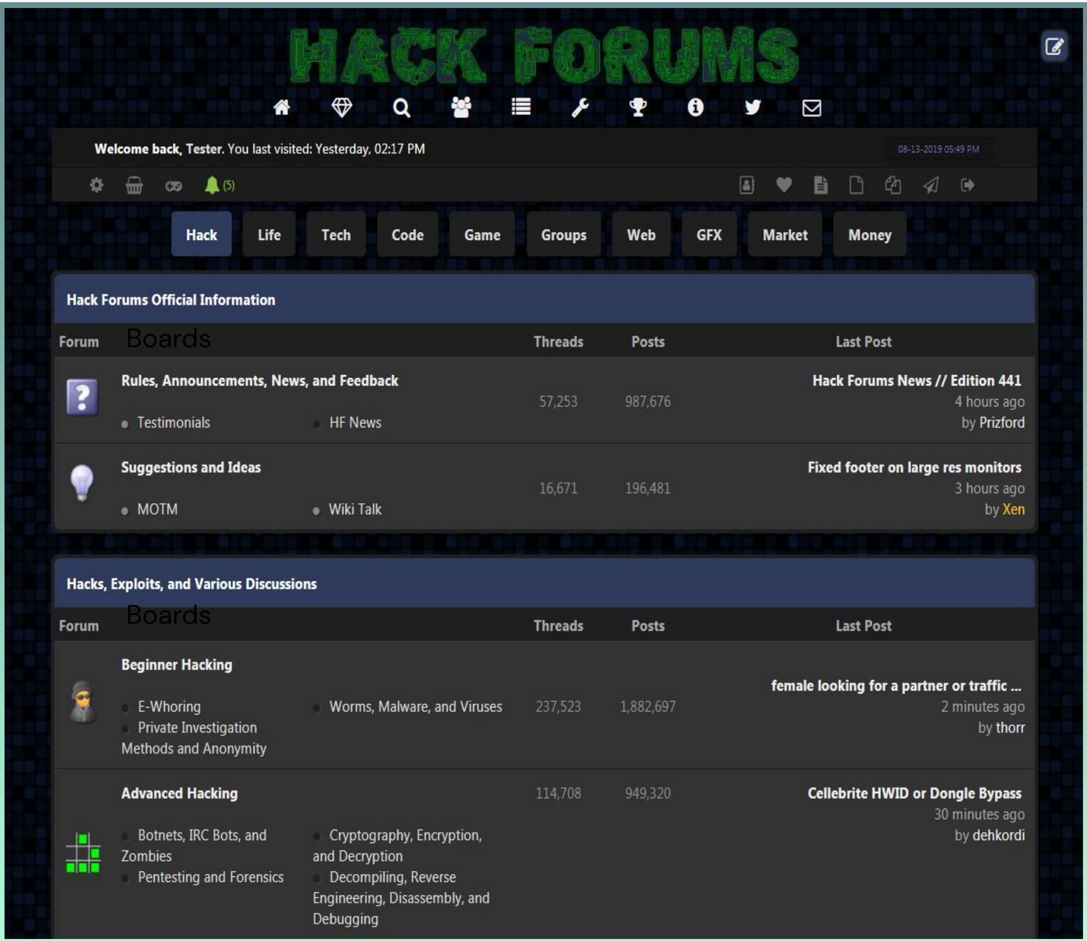
- 32 forums
- Mostly in English
- Millions of posts



Forums	Main Language	# Boards	# Members	# Threads	# Posts	Oldest
Hack Forums	EN	197	689 624	4 044 893	42 165 425	2007/01
KernelMode	EN	11	1 688	3 441	25 825	2010/03
The Hub	EN	62	8 340	11 286	88 753	2014/01
Offensive Community	EN	71	11 531	119 251	161 492	2012/06
MPGH	EN	752	511 440	785 117	9 729 511	2005/12
Stresser Forums	EN	17	779	708	7 069	2017/04
GreySec Forums	EN	25	915	1 878	10 463	2015/06
Garage4hackers	EN	35	881	2 096	7 697	2010/07
BlackHatWorld	EN	100	330 052	644 797	8 112 738	2005/10
lolzteam	RU	292	483 754	577 642	6 196 005	2013/03
Antichat	RU	64	79 887	243 176	2 449 404	2002/05
OGUsers	EN	58	48 944	244 766	3 608 306	1900/01
RaidForums	EN	75	46 111	34 798	214 856	2015/03
Safe Sky Hacks	EN	50	7 471	12 963	27 018	2013/03
V3rmillion	EN	40	75 283	456 262	2 459 519	2016/02
FreeHacks	RU	197	1 225	1 572	6 247	2013/07
Nulled	EN	151	856 833	155 482	3 495 768	2013/04
Zismo Forum	EN	25	162 003	425 158	8 486 440	2010/05
StresserForums	EN	21	20	34	53	2019/04
Dread	EN	446	52 406	75 122	294 596	2018/02
Torum	EN	11	3 835	4 346	28 485	2017/05
Envoy Forum	EN	93	364	454	2 163	2019/07
PirateBay Forum	EN	33	8 633	11 526	60 678	2013/10
Deutschland im Deep Web	EN	43	2 516	4 075	20 185	2018/11
Runion	EN	19	17 343	16 867	240 632	2012/01
Cracked.to	EN	130	168 616	78 124	276 698	2018/04
UnKnoWnCheaTs	EN	230	184 568	126 594	1 995 369	2002/11
Underc0de	ES	69	6 087	20 835	78 479	2010/02
Probiv	RU	107	9 034	54 929	345 666	2014/11
Indetectables	ES	56	11 911	31 448	324 956	2006/02
Elhacker	ES	52	25 326	203 415	296 269	2002/08
Ifud	RU	48	5 071	10 904	65 990	2012/05
Total		3580	3 755 062	8 403 959	91 282 755	2002/05

Structure





HACK FORUMS

Welcome back, Tester. You last visited: Yesterday, 02:17 PM

08-13-2019 05:49 PM

(5)

Hack Life Tech Code Game Groups Web GFX Market Money

Hack Forums Official Information

Boards

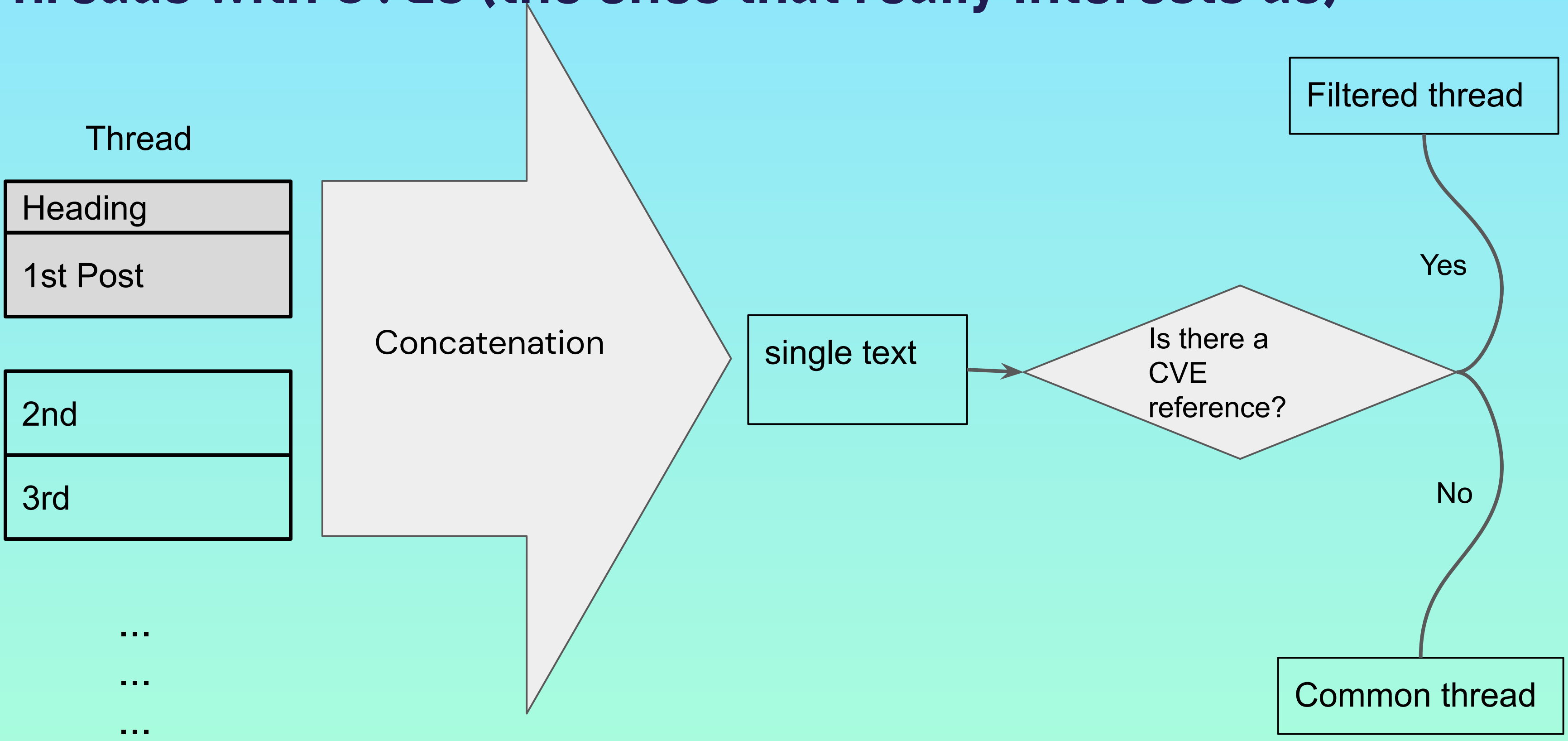
Forum	Threads	Posts	Last Post
Rules, Announcements, News, and Feedback Testimonials HF News	57,253	987,676	**Hack Forums News // Edition 441** 4 hours ago by Prizford
Suggestions and Ideas MOTM Wiki Talk	16,671	196,481	**Fixed footer on large res monitors** 3 hours ago by Xen

Hacks, Exploits, and Various Discussions

Boards

Forum	Threads	Posts	Last Post
Beginner Hacking E-Whoring Worms, Malware, and Viruses Private Investigation Methods and Anonymity	237,523	1,882,697	**female looking for a partner or traffic ...** 2 minutes ago by thorr
Advanced Hacking Botnets, IRC Bots, and Zombies Cryptography, Encryption, and Decryption Pentesting and Forensics Decompiling, Reverse Engineering, Disassembly, and Debugging	114,708	949,320	**Cellebrite HWID or Dongle Bypass** 30 minutes ago by dehkordi

Threads with CVEs (the ones that really interests us)



Thread

Heading

1st Post

2nd

3rd

...

...

...

Concatenation

single text

Is there a CVE reference?

Yes

No

Filtered thread

Common thread

Filtered Threads Labeling

PoC

Thread

[TUT] How to run Exploit Scripts! [TUT]

In this **TuT** im going to show you how to run this example Python Exploit and how you would run other types of scripts.
Code:
CODE
# Reference: CVE-2007-1531
# Description: Microsoft Windows Vista (SP0) dumps interfaces when it receives this ARP packet. This DoS is useful for an Hoagland
# Vulnerable: Microsoft Windows Vista (SP0)
# **Tested**: # * victim == Windows Vista Enterprise (SP0) [English]
# * **attacker** == Ubuntu Feisty (7.04)

Very nice tutorial, i always like to learn something new.

PoC or Exploitation?

Thread

Noob DDoS Question (Attack Methods)

What would be the best method to take down a website like; cambornescience.co.uk with no ddos protection?
layer 7 attack (which method)
thanks!

it looks like running apache on your target. Reference to **CVE-2011-3192** with **POC**.
one http request is increased to hundred times ever 2 thousand times by vul on apache server.
after a few minutes server will be freeze causing in full of memory and cpu. **This is fully tested on serveral targets.** You have some question then send PM to me.

Exploitation

Thread

Bleeding Life v2: RELOADED **Exploit Pack** (SICK for hacking! Very dangerous tool!)

[...] If you want a low cost, high rate and great quality pack... Purchase BleedingLife v2 Reloaded!

Very nice tool. Good luck with your sales mate,i heard it's very good :D

Very interested, I PM you

PMing you now for further info. I have been looking a long time for a program like this, looking forward to buy this!

Filtered Threads Labeling

PoC

Thread

[TUT] How to run Exploit Scripts! [TUT]

In this **TuT** im going to show you how to run this example Python Exploit and how you would run other types of scripts.
Code:
```
***CODE***
# Reference: CVE-2007-1531
# Description: Microsoft Windows Vista (SP0) dumps interfaces when  it receives this ARP packet.  This DoS is useful for an Hoagland
# Vulnerable: Microsoft Windows Vista (SP0)
# Tested: # * victim == Windows Vista Enterprise (SP0) [English]
# * attacker == Ubuntu Feisty (7.04)
```

Very nice tutorial, i always like to learn something new.

PoC or Exploitation?

Thread

Noob DDoS Question (Attack Methods)

What would be the best metho
to take down a website like;
cambornescience.co.uk with
ddos protection?
layer 7 attack (which metho
thanks!

it looks like running apache on your target. Reference to **CVE-2011-3192** with **POC**.
one http request is increased to hundred times ever 2 thousand times by vul on apache server. after a few minutes server will be freeze causing in full of memory and cpu. **This is fully tested on serveral targets.** You have some question then send PM to me.

Challenge: how to classify it? refers to PoC in text but looks like exploitation in the wild!

Exploitation

Thread

Bleeding Life v2: RELOADED **Exploit Pack** (SICK for hacking! Very dangerous tool!)

[...] If you want a low cost, high rate and great quality pack... Purchase BleedingLife v2 Reloaded!

Very nice tool. Good luck with your sales mate,i heard it's very good :D

Very interested, I PM you

PMing you now for further info. I have been looking a long time for a program like this, looking forward to buy this!

Dataset

- Hackforums (train)
  - 40+ million posts
  - 4 million threads
  - 1194 threads with CVEs
  - 764 threads with CVEs used
- OffensiveCommunity (test)
  - 161,492 posts
  - 119,251 threads
  - 29 threads with CVEs
- Antichat (test)
  - Russian
  - 2,449,396 posts
  - 243,171 threads
  - 219 threads with CVEs

The model is trained only with Hackforums, because of its way larger size and because we want to evaluate it on other forums and check the generalization power we get. I think it's important to check the model on data from different sources but of the same nature. Antichat is not even in english, so it had to be translated. If the model gets a good score on Antichat, it will be a good argument to support the created model.

Pipeline

Text Preprocessing

Tokenization, strip punctuation, lower/upper case, stopwords removal, etc

Doc2Vec / Bag of Words

Get the embeddings/tf-idf weights associated to each document (thread or TI feed event).

Classifier

Train model with the embeddings or tf-idf weights.

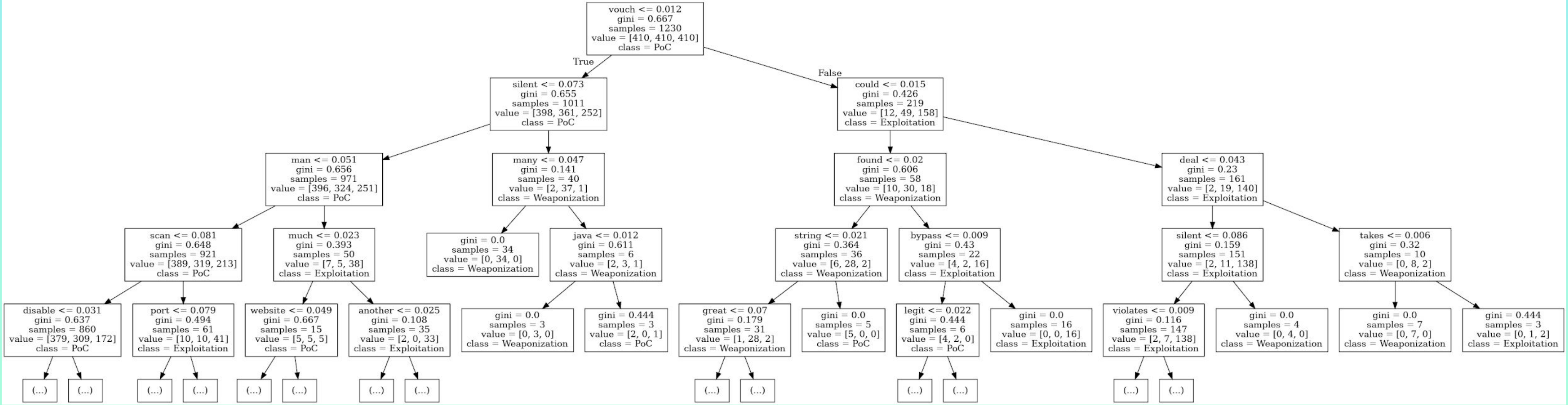
Evaluation

1. Because there are 3 classes, there are a couple of possible classifiers to choose as the final (PoC vs Weap vs Expl, Exploitation vs All, PoC vs all, etc). The idea is to evaluate each one of them and choose the one with the best scores.
2. To train the model, the techniques of Undersampling and Oversampling are applied to the dataset, so the model doesn't get affected by the imbalance. In the decision trees of the following slides you'll note that the number of samples per class is the same. I only show the trees with oversampling for simplicity and because they have similarities.

3 classes

- 82% accuracy with doc2vec
- 60% accuracy with tfidf

Decision tree



2 classes might be better: discriminator to decide which labels to use

PoC vs Weap + Expl	PoC + Weap vs Expl	PoC + Expl vs Weap
80% doc2vec	97% doc2vec	63% doc2vec
68% tfidf	87% tfidf	62% tfidf

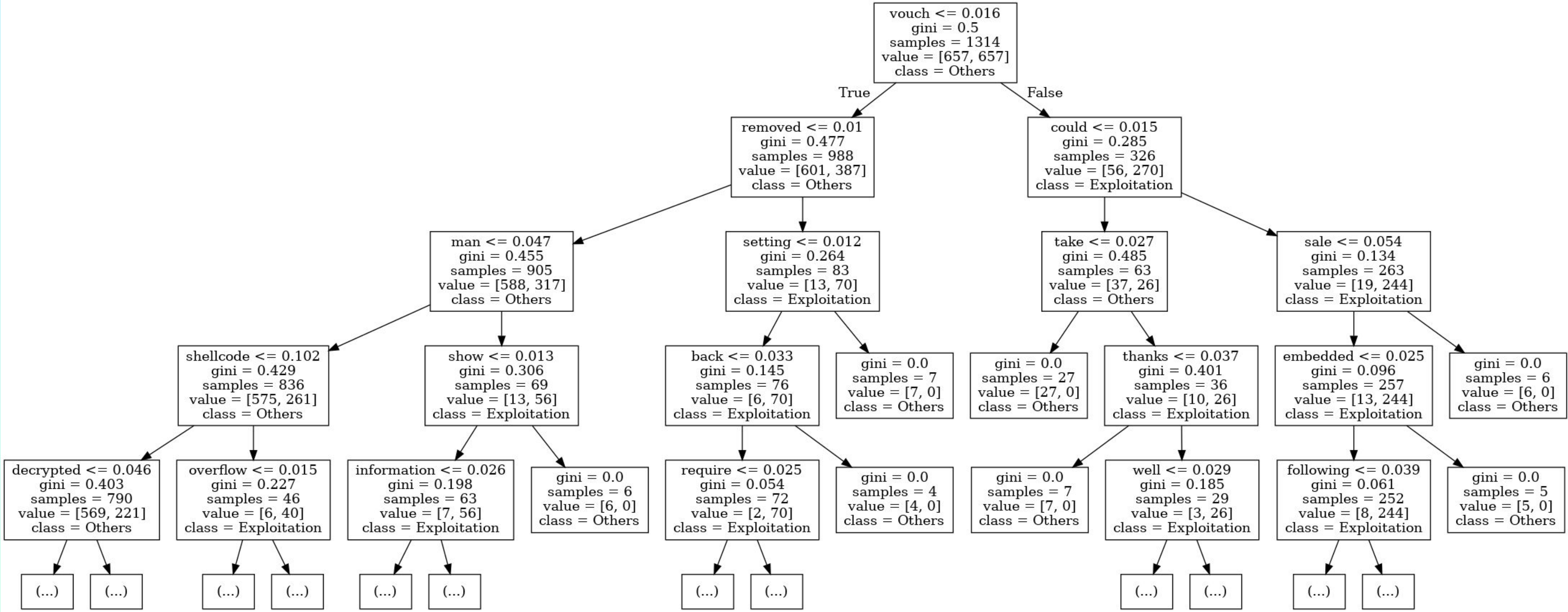
Discriminator

PoC vs Weap + Expl	**PoC + Weap vs Expl**	PoC + Expl vs Weap
80% doc2vec	**97% doc2vec**	63% doc2vec
68% tfidf	**87% tfidf**	62% tfidf

Preliminary Evaluation (labeling must be improved and assigned to more samples)

	Doc2Vec	TF-IDF
Hackforums	97%	87%
OffensiveCommunity	88%	-
Antichat (only 10 threads labeled for now)	83%	-

Decision Tree for 2 classes (PoC + Weap vs Exploitation)



vouch <= 0.016
gini = 0.5
samples = 1314
value = [657, 657]
class = Others

True

False

removed <= 0.01
gini = 0.477
samples = 988
value = [601, 387]
class = Others

could <= 0.015
gini = 0.285
samples = 326
value = [56, 270]
class = Exploitation

man <= 0.047
gini = 0.455
samples = 905
value = [588, 317]
class = Others

setting <= 0.012
gini = 0.264
samples = 83
value = [13, 70]
class = Exploitation

take <= 0.027
gini = 0.485
samples = 63
value = [37, 26]
class = Others

sale <= 0.054
gini = 0.134
samples = 263
value = [19, 244]
class = Exploitation

shellcode <= 0.102
gini = 0.429
samples = 836
value = [575, 261]
class = Others

show <= 0.013
gini = 0.306
samples = 69
value = [13, 56]
class = Exploitation

back <= 0.033
gini = 0.145
samples = 76
value = [6, 70]
class = Exploitation

gini = 0.0
samples = 7
value = [7, 0]
class = Others

gini = 0.0
samples = 27
value = [27, 0]
class = Others

thanks <= 0.037
gini = 0.401
samples = 36
value = [10, 26]
class = Exploitation

embedded <= 0.025
gini = 0.096
samples = 257
value = [13, 244]
class = Exploitation

gini = 0.0
samples = 6
value = [6, 0]
class = Others

decrypted <= 0.046
gini = 0.403
samples = 790
value = [569, 221]
class = Others

overflow <= 0.015
gini = 0.227
samples = 46
value = [6, 40]
class = Exploitation

information <= 0.026
gini = 0.198
samples = 63
value = [7, 56]
class = Exploitation

gini = 0.0
samples = 6
value = [6, 0]
class = Others

require <= 0.025
gini = 0.054
samples = 72
value = [2, 70]
class = Exploitation

gini = 0.0
samples = 4
value = [4, 0]
class = Others

gini = 0.0
samples = 7
value = [7, 0]
class = Others

well <= 0.029
gini = 0.185
samples = 29
value = [3, 26]
class = Exploitation

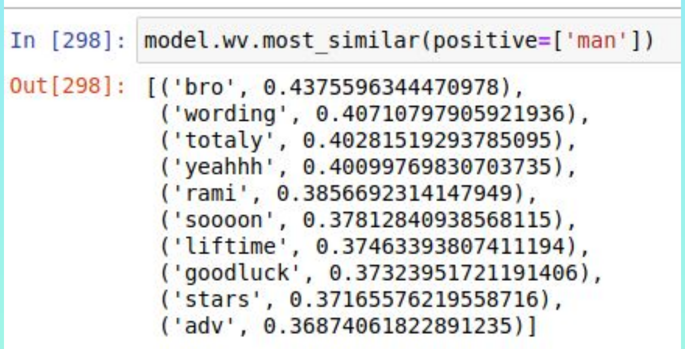
following <= 0.039
gini = 0.061
samples = 252
value = [8, 244]
class = Exploitation

gini = 0.0
samples = 5
value = [5, 0]
class = Others

(...) (...) (...) (...) (...) (...) (...) (...) (...) (...) (...) (...)

Doc2Vec Insight

What exactly is the word “man” doing on the decision tree? Word2Vec to investigate! (notice Doc2Vec was not necessary in this case. Doc2Vec is built on top of Word2Vec, so we have access to it with the Doc2Vec model)



```
In [298]: model.wv.most_similar(positive=['man'])

Out[298]: [('bro', 0.4375596344470978),
           ('wording', 0.40710797905921936),
           ('totaly', 0.40281519293785095),
           ('yeahhh', 0.40099769830703735),
           ('rami', 0.3856692314147949),
           ('soooon', 0.37812840938568115),
           ('liftime', 0.37463393807411194),
           ('goodluck', 0.37323951721191406),
           ('stars', 0.37165576219558716),
           ('adv', 0.36874061822891235)]
```

The most similar word is “bro”, so we can be sure it is really the noun we were thinking. By the experience gained with the manual labeling, we know exploitation threads usually have friendly interactions directed to an author who announced a pack or a specific exploitation. Now the word “man” on the decision tree makes sense.

These posts were found in exploitation threads:

“Ok **man** I wanna buy this, I have to wait a till i get my refund but I can pay via paypal.”

“**man** i just cant wait i need this lol”

“Thanks for the vouch **man**! Glad you're satisfied with the product.”

“Hello **man**,
I'm interested in buying this.”

Lessons Learned

- Doc2Vec provides better scores than tf-idf but less interpretability.
- Doc2Vec can help to understand the words on the decision tree though
- Manual labeling must be reliable, otherwise the researcher himself can start to question his own results. To that aim, it’s important to define a set of well defined rules and follow them through the process.
- It’s also important to automate heuristics and apply them to the dataset. If the scores are good, then the manual labeling is reliable.
- 3 classes provide interesting decision tree but worse scores
- Oversampling vs Undersampling didn’t seem to affect the final scores. I was not sure if oversampling would improve or decrease the scores. None happened. To be tested with more samples.