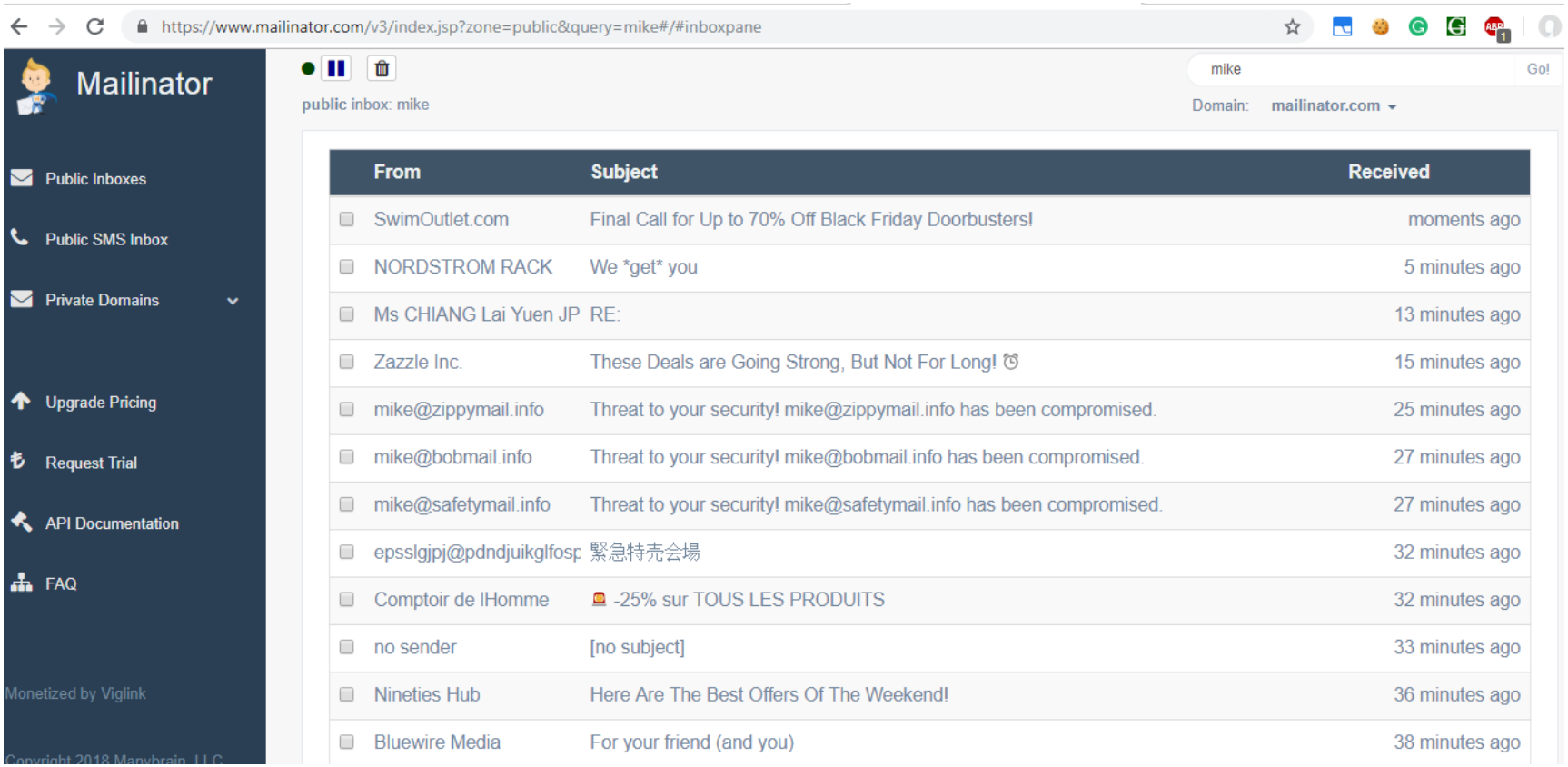
Collecting, categorizing and analyzing Emails from top Free Disposable Email Address Services

CSE508: Network Security

By Shubham Jindal
SBU ID: 112129688

DEA Services

- Unique public email address
- Disposable if compromised (spam, phishing)
- Free or paid
- Uses: Online registration/middle layer mailbox



mike@mailinator.com

Problem

- Analyze the kinds of mails present in DEA mailboxes.

- Requirements:
  - Collection of mails
  - Categorization of mails
  - Draw insights

Challenges

- **C1**: pages took time to load
- **C2**: mailinator has request limit and is protected by cloudFlare, number one service for protection against DOS and DDOS attacks.
- **C3**: mailinator blocks TOR proxy
- **C4**: Access lost after crossing limits.
- **C5**: Mailinator empties the email data at a certain point of time every day.
- **C6:** every run of the mailinator started fetching the emails from the start.

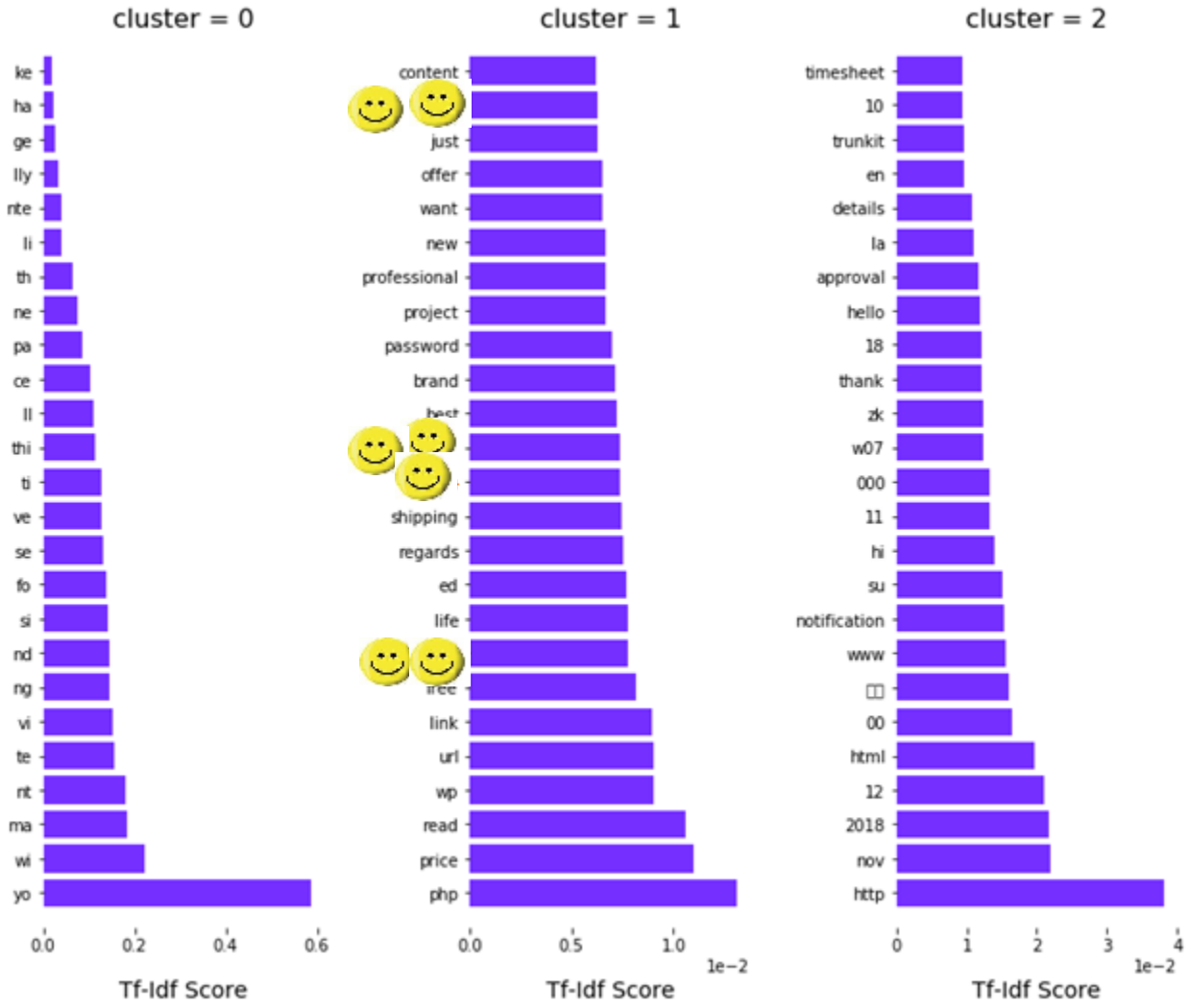
- Usage limit
- Abuse limit
- Blocks IP after abuse limit is crossed

Results

- Collection: Selenium + python + cron + timing adjustments
- Classification: python + ML(scikit learn)
- 100 common names
- Top Google bad words:1703
- Collection results:

Total Mailboxes checked	:	101
Total Mails	:	2808

Classification Result and Insights



SimpleClassifier:

Profane Mail	:539/2808
Pharma Mails	:15
Profanity	:19%
Pharma	:0.5%
DEA Used	:{'yopmail.com': 5}

MLClassifier:

-Top categories : Profanity/promotion/offer

-Some mails in Mandarin and Russian

-26 extortion mails by hackers with BTC Wallet id.

-Presence of vulnerable urls

Future Work

- Crawling links in mail.
- Analyzing source email addresses, time stamp.
- Image and OCR for mails
- Translating emails from Language Unicode to English (Google translater api - 1000000 words/ $20).
- Include scripts for yopmail, tempr.email.

Project link:
https://github.com/sjindal94/DEAemailscraper