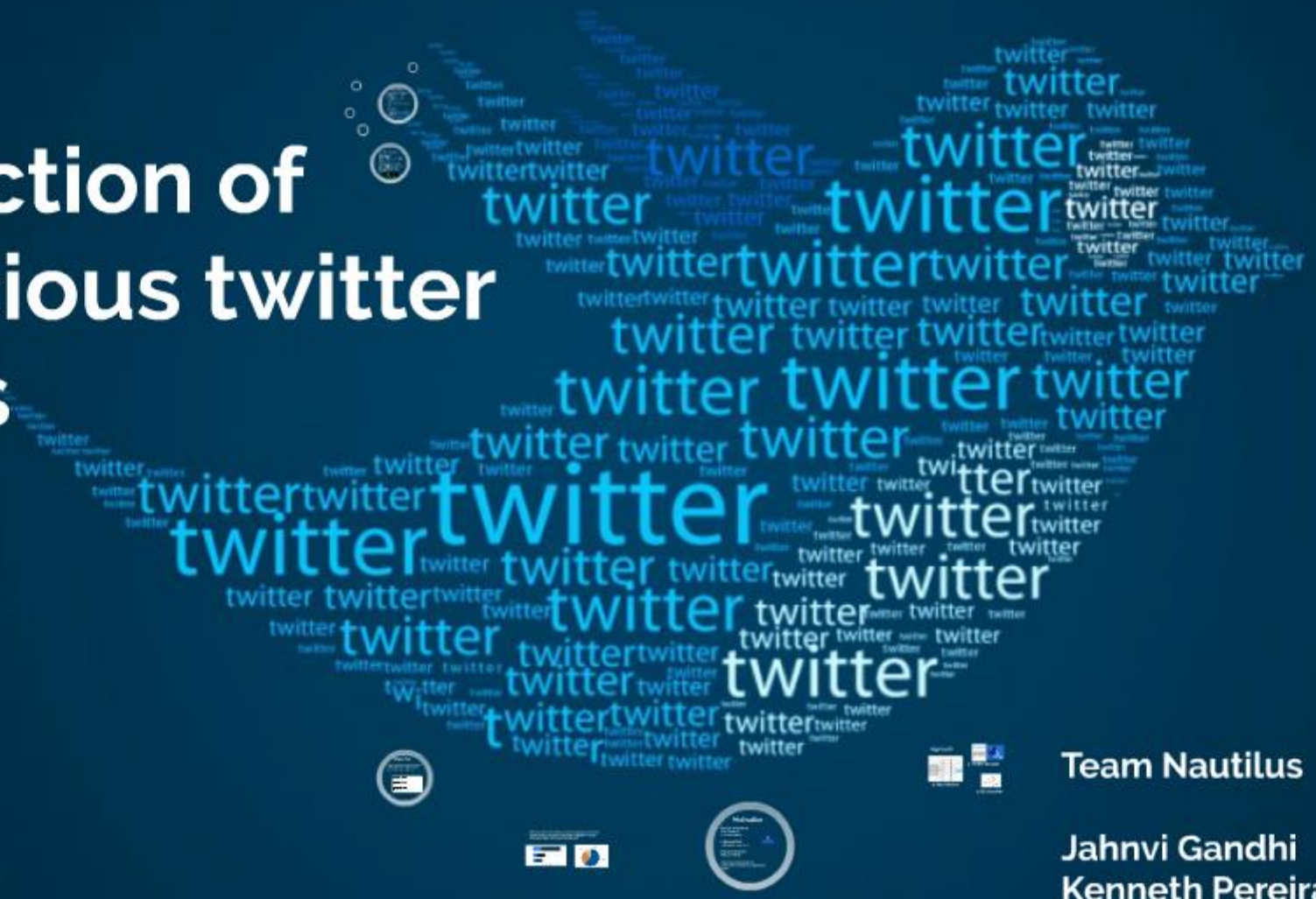
Detection of Malicious twitter feeds



Team Nautilus

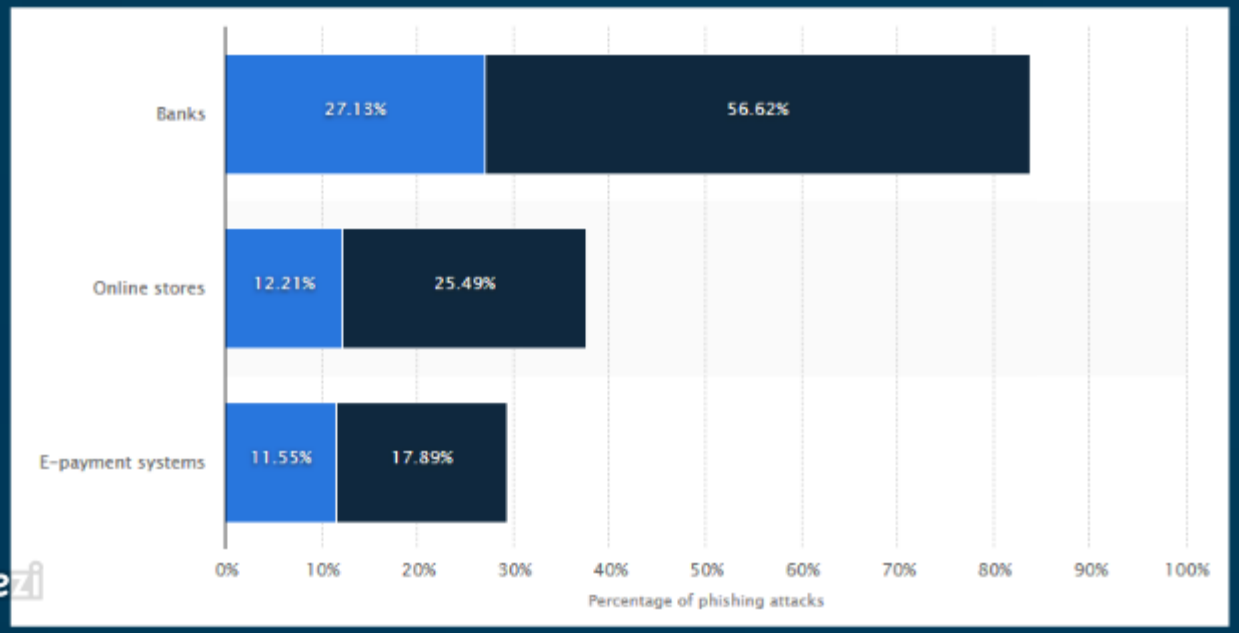
Jahnvi Gandhi
Kenneth Pereira
Mariano Claveria
Nuhiya Rafeeq
Priyanka Mishra

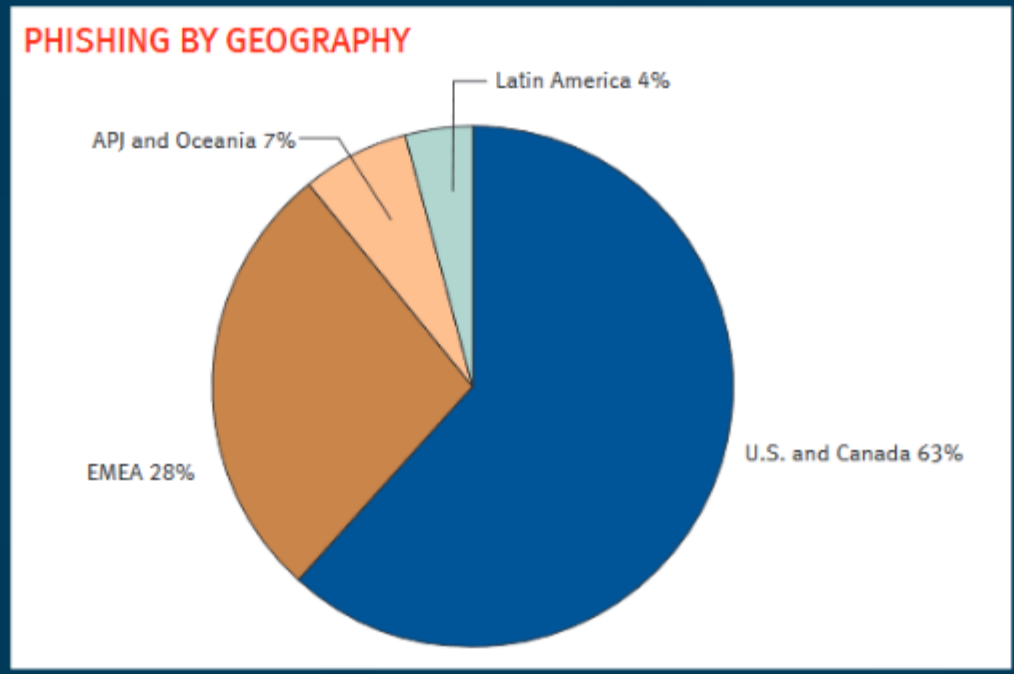
Objective

Build a system that scrapes real-time twitter feeds and classifies shared URLs as phishing or legitimate

Why?

- More than 50% of phishing cases lead to Financial losses
- Organizations encounter more than $3 Billion in losses
- Increase in Data Theft and Identity theft





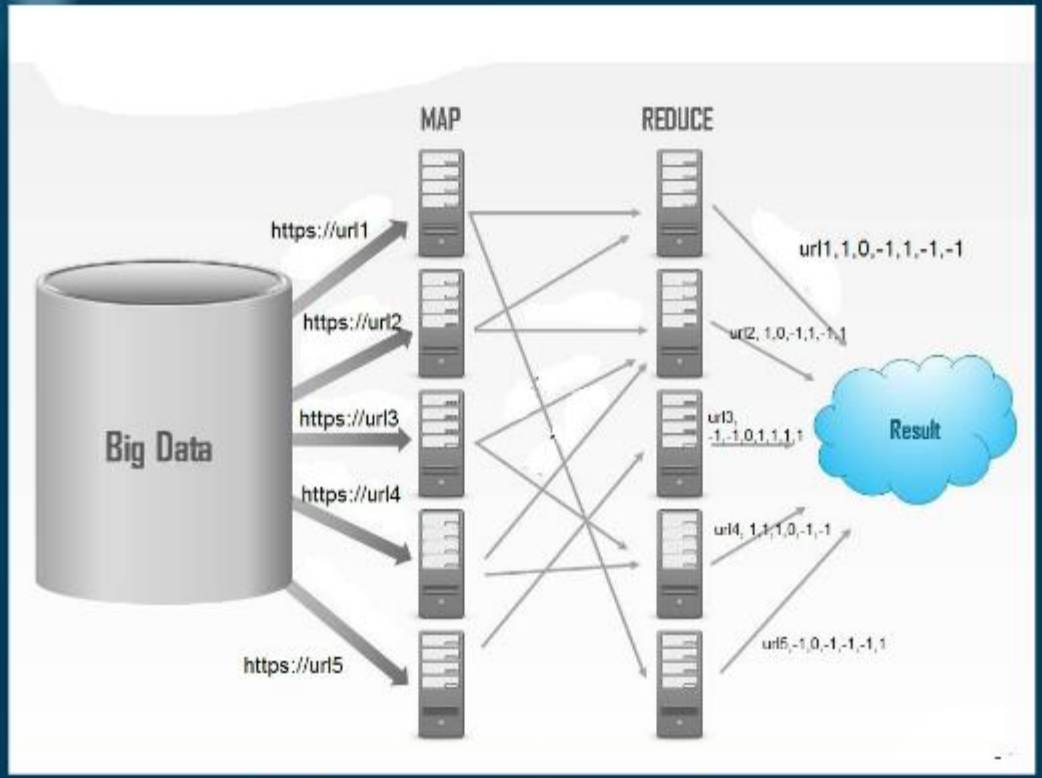
Motivation

- Rami M. Mohammad
  Fadi Thabtah
  Lee McCluskey
- Collected Data using their own tool
- Phishing Websites Dataset (2015)
- Only rules discussed, no instruction as how to implement them

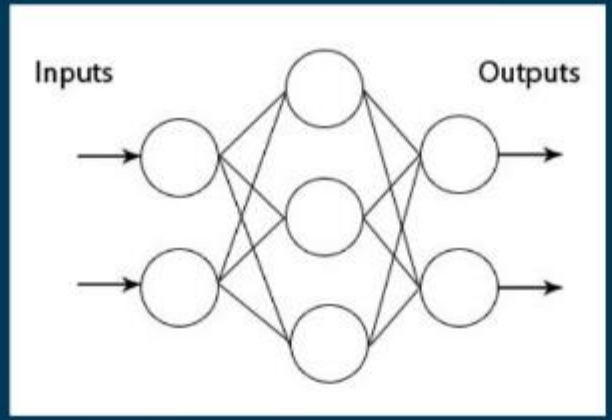
University of HUDDERSFIELD

Approach

1. Twitter Scraper



2. Map Reduce



3. ML classifier

Twitter Scraper

- Built using a twitter API
- API gets feeds
- Output file (Tweet data, userDetails, timestamp, location...)

Roadblocks:

- API calls restricted to 180 requests every 15 minutes
- Missing data from the Twitter feeds due to User Account Settings
- Shortend URLs

```
Output URLS: 165122
Waiting 15 minutes to continue
Tweets read: 310968
Output URLS: 165122
ubuntu@ip-172-31-20-54:~$
```

Map Reduce

- MR Model takes in Twitter feed
- Reducer checks and applies URL rules
- MR Output is a file with 6 attributes
- ex: URL, 1, -1, 0, 1, 1, 1

Adv: Easy detection of repeated URLs

Roadblocks:

- Consistent data required, if any scrapped field has missing data, MR job fails
- Ex: location

MapReduce

- ML output

```
http://1.usa.gov/1Xd5hV1       ,1,1,0,1,1,1,10
http://10.Weather      ,1,1,0,1,1,1,56
http://100-singalong-songs-for-kids.i9-news.co
utm_source=dlvr.it&utm_medium=twitter ,-1,1,0,
http://1000oldies.de   ,1,1,0,1,1,1,8
http://1000phones.com/phone/1-617-830-7579    ,
http://1000rockhits.de ,1,1,0,1,1,1,18
http://1001.tl/122813  ,1,1,0,1,1,1,2
http://100resilientcities.org/blog/entry/bosto
```

Instance State	Status Checks	Public DNS (IPv4)
running	2/2 checks ...	ec2-52-15-123-175.us-...

Reducer

URL rules:

- Using the IP Address

"http://0x58.0xCC.0xCA.0x62/2/paypal.ca/index.html"

$$\underline{Rule}: \text{IF}\begin{cases}\text{If The Domain Part has an IP Address} \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate}\end{cases}$$

- Long URL to Hide the Suspicious Part

http://federmacedoadv.com.br/3f/aze/ab51e2e319e51502f416dbe46b773a5e/?cmd=_home&dispatch=11004d58f5b74f8dc1e7c2e8dd4105e811004d58f5b74f8dc1e7c2e8dd4105e8@phishing.website.html

$$Rule: IF\begin{cases} URL\ length < 54 \rightarrow feature = \text{Legitimate} \\ else\ if\ URL\ length \geq 54\ and\ \leq 75 \rightarrow feature = Suspicious \\ otherwise \rightarrow feature = \text{Phishing}\end{cases}$$

- URL's having "@" Symbol

$$\text{Rule: IF}\begin{cases}\text{Url Having @ Symbol} \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate}\end{cases}$$

Reducer

URL rules (cont):

- Redirecting using "//"

http://www.legitimate.com//http://www.phishing.com

$$\text{Rule: IF} \begin{cases} \text{ThePosition of the Last Occurrence of "//" in the URL} > 7 \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate} \end{cases}$$

- Adding Prefix or Suffix Separated by (-) to the Domain

http://www.Confirme-paypal.com/

$$\text{Rule: IF} \begin{cases} \text{ThePosition of the Last Occurrence of "//" in the URL} > 7 \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate} \end{cases}$$

- The Existence of "HTTPS" Token in the Domain Part of the URL

http://https-www-paypal-it-webapps-mpp-home.soft-hair.com/

$$\text{Rule: IF} \begin{cases} \text{Using HTTP Token in Domain Part of The URL} \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate} \end{cases}$$

MapReduce findings

- Mapper Input: Input file (includes Urls, location, etc)
- Mapper Output: URLs , Count
- Reducer input: URLs, Count
- Reducer output: URLs, Attributes Values

- Six URL rules out of 30 were achievable for predicting the phishing websites
- Ex for non-achievable rules: Favicon, HTTPS certificate, age of Domain, website traffic

- Parallelization achieved
- Map-Reduce helped in observing the duplicate URLs.

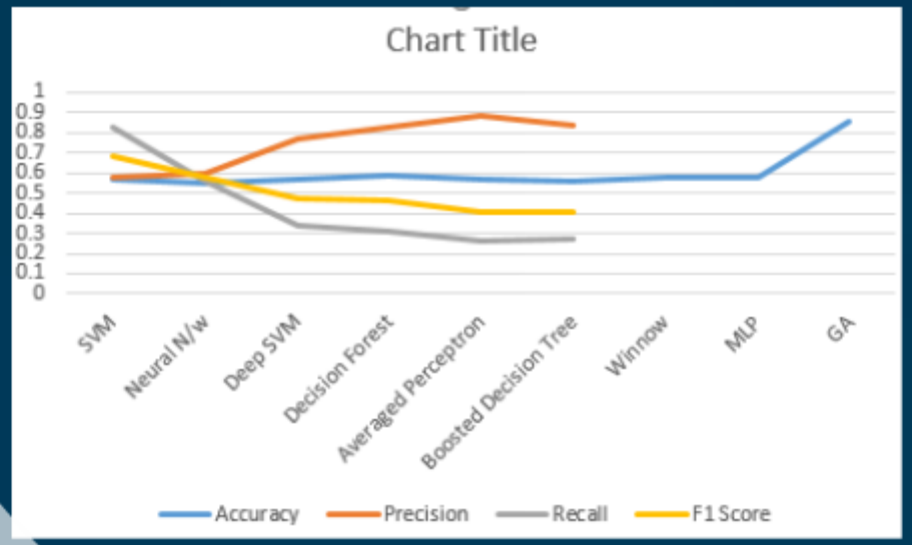
ML classifier

- Two class classification
  - 0 = Phishing
  - 1 = Legitimate
- Input = URL and 6 attributes
  ex: url, 1, -1, 0, 1, -1, 1
- Output = URL, 6 attributes, target Value
  ex: url, 1, -1, 0, 1, -1, 1, 1

```
https://www.youtube.com/watch?v=4di9SQ90KAU&feature=share,0,1,0,1,1,1,1
https://www.youtube.com/watch?v=4xAjrQ7wb3Q,1,1,0,1,1,1,1
https://www.youtube.com/watch?v=56qdgDRDdUM,1,1,0,1,1,1,1
https://www.youtube.com/watch?v=6EPwRdVg5Ug&feature=share,0,1,0,1,1,1,1
https://www.youtube.com/watch?v=6I1bNZ5OdvQ,1,1,0,1,1,1,1
https://www.youtube.com/watch?v=6X0jxo3jlok&feature=share,0,1,0,1,1,1,1
```

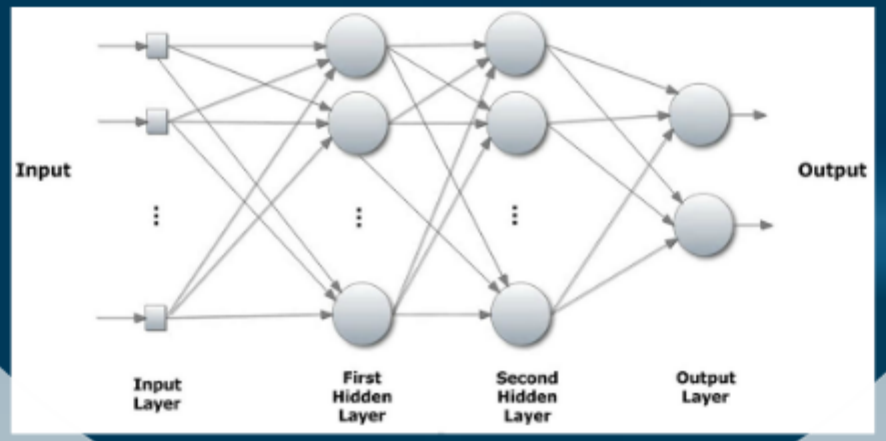
ML findings

- MLP perceptron (deepLearning4j)
- Microsoft Azure ML
  6 algorithms applied and tested
- Winnow, Logistic Classfier, GA



MLP

- Feedforward artificial neural network
- Each layer fully connected to the next one
- Each node is a neuron
- Nonlinear activation function
- Supervised learning technique called backpropagation
- Modification of linear perceptron and can distinguish data that are not linearly separable



ML charts

Twitter Data_Final_Algos › Score Model › Scored dataset

rows 98046 columns 10

double_slash_redirecting	Prefix_Suffix	HTTPS_token	Count	Scored Labels	Scored Probabilities
1	1	1	10	1	0.990307
1	1	1	56	1	0.990307
-1	1	1	2	1	0.999087
1	1	1	8	1	0.990307
1	1	1	2	1	0.990307
1	1	1	18	1	0.990307

2 class Neural Nw

Twitter Data_Final_Algos › Score Model › Scored dataset

rows 98046 columns 10

double_slash_redirecting	Prefix_Suffix	HTTPS_token	Count	Scored Labels	Scored Probabilities
1	1	1	10	1	0.958993
1	1	1	56	1	0.958993
-1	1	1	2	1	0.815083
1	1	1	8	1	0.958993
1	1	1	2	1	0.958993
1	1	1	18	1	0.958993

2 class Support Vector Machine

Future Scope

- Automate twitter API to look for new feeds and classify URLs as and when new feeds are available
- Look for more sophisticated algorithms (deep learning techniques) to get better accuracy
- URL advisor
- ChatBot - scrap data from page being viewed and inform user about URLs not to be clicked
- Sanitize the phishing URLs

Thank You