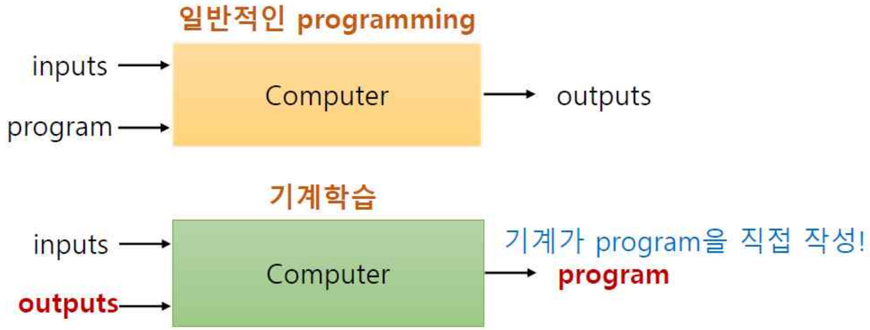
딥러닝을 이용한 악성 도메인 탐지 기법

한국인터넷진흥원
사이버보안빅데이터센터
서상욱

1. 머신러닝 vs 딥러닝

머신러닝이란?

- 컴퓨터에 명시적으로 프로그래밍하지 않고 학습할 수 있는 능력을 부여하는 컴퓨터 과학의 하위분야



일반적인 programming

inputs → Computer → outputs

program →

기계학습

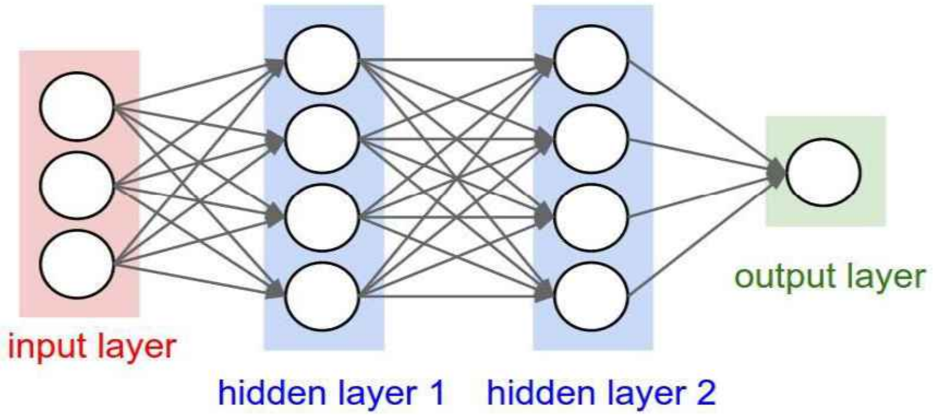
inputs → Computer → program

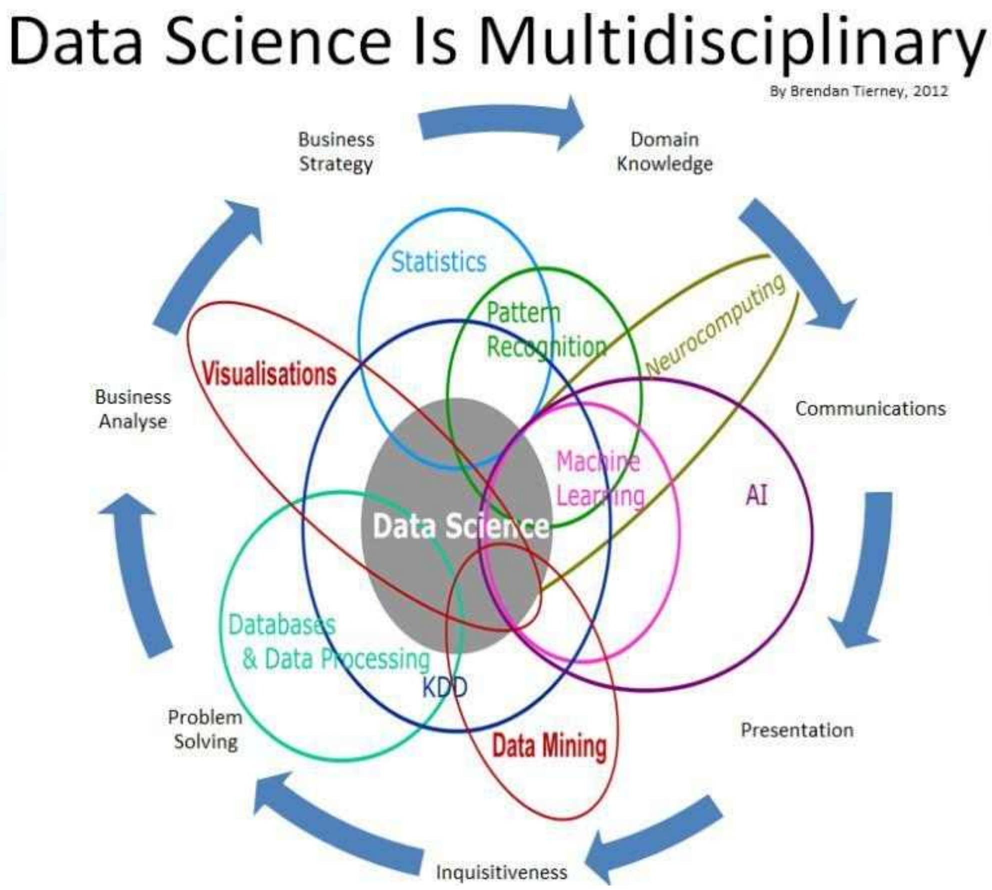
outputs →

기계가 program을 직접 작성!

딥러닝이란?

- 딥러닝은 Deep Neural Network를 통하여 학습하는 것





2. DGA(Domain Generation Algorithm) 개요

DGA 정의 (from Wikipedia)

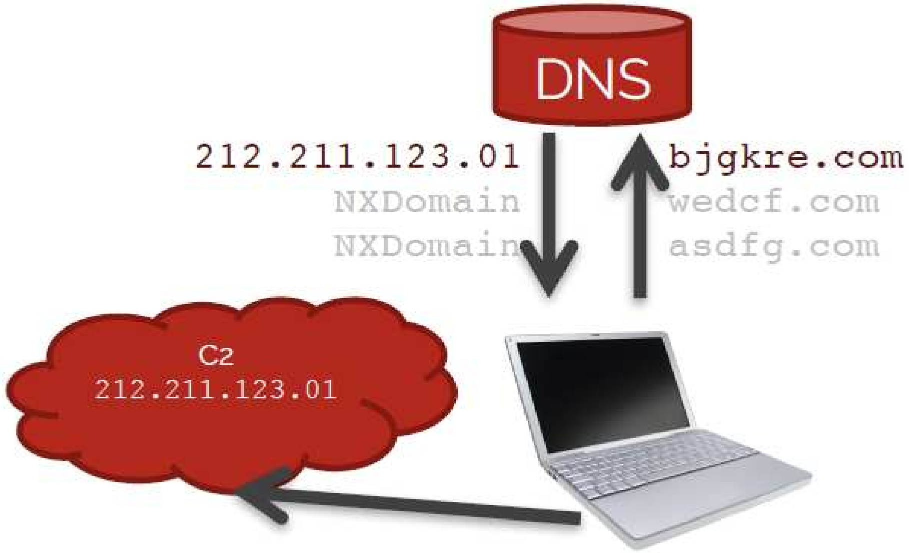
- DGA(Domain Generation Algorithm) - algorithms seen in various families of malware that are used to periodically generate a large number of domain names that can be used as rendezvous points with their command and control servers

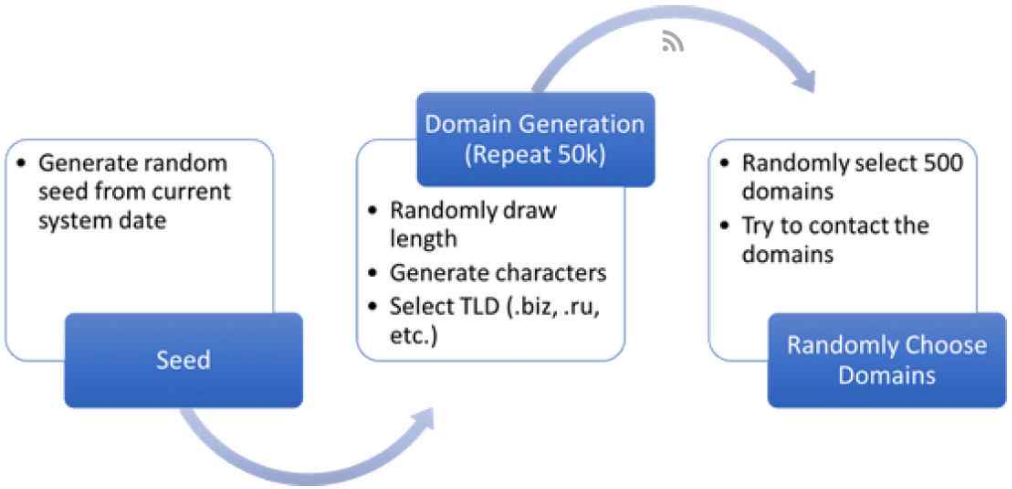
DGA 역사 (from FIRST 2017) [3]

- Early 2008 - Kranken one of the first malware families to use a DGA
- Mid 2008 - World's largest botnet "Srizbi" uses DGA algorithm
  - FireEye sinkholes for two weeks to keep out of criminal hands - abandoned
- Late 2008 - Conficker first discovered
  - Sinkhole efforts successful but malware authors escalate to creating over 250,000 potential domains per day in 2009.
- 2010 - Texas A&M University researchers publish paper on detecting DGA domain names
- 2012 - Georgia Tech and Damballa release whitepapers on new DGA use and detection methods using machine learning
- 2015 - DGA tracker website online
- 2016 - Registrar of last resort stood-up to sinkhole many DGA's

2. DGA(Domain Generation Algorithm) 개요

DGA 구조 [2, 4, 5]





Seed

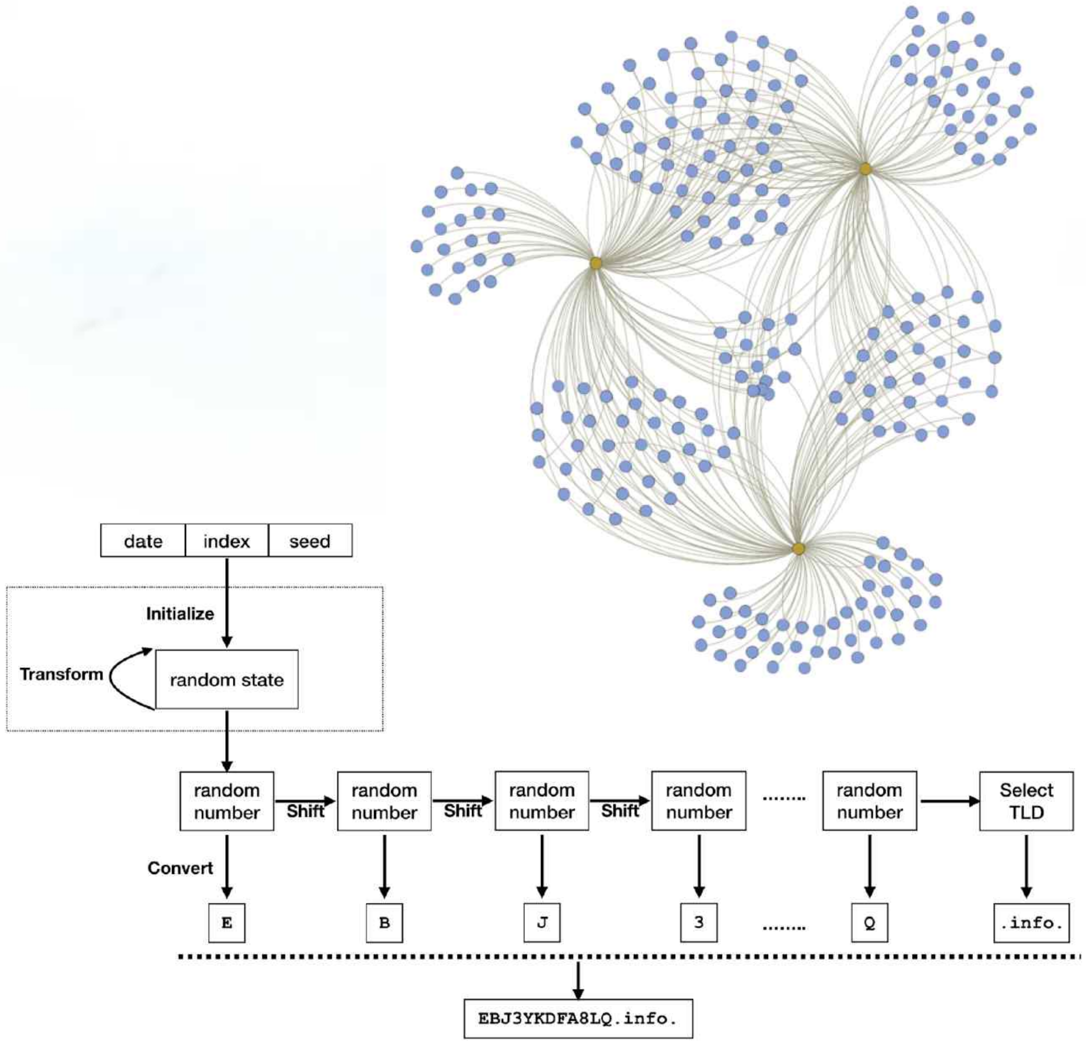
- Generate random seed from current system date

Domain Generation (Repeat 50k)

- Randomly draw length
- Generate characters
- Select TLD (.biz, .ru, etc.)

Randomly Choose Domains

- Randomly select 500 domains
- Try to contact the domains

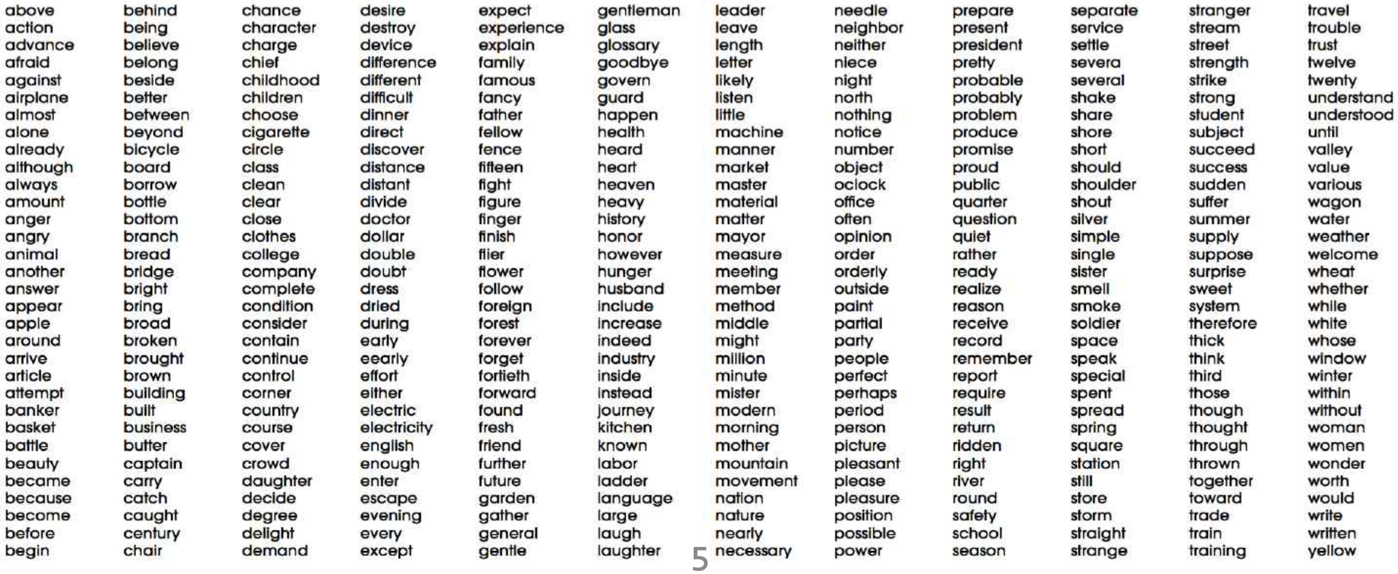


3. 악성 vs 정상 도메인

DGA 및 정상 도메인 샘플 [1]

Cryptolocker	Goz	NewGoz	Legit
etledwndgunmrt	eiaupamojzhlrciwkeqhyxd	1erk1aq2tfv3e1dy8ikv1f0nxs8	fujifilm
obgfmoyfwptep	tkdabqnkrgdozhitdehypz	i5ep531lfuanc1ytynl1mmkio4	dallasdoglife
bugvesrwqxdjoa	uswodcmnvemqfmzxynjdnvhynvbe	zj7llmpk5fo87dtcg81e2j07c	startups
qxavdikemhepxk	ohhyhypphvgtucgiemfqdhai	vehvq1swdu9vuhfqvrcjxr46	askganesha
ohgnphscwbyvuse	ydqwmzhgaxoxfyzvcpvqgmfxro	1ncn8kn675d4o6dc4hh1f0se4r	wildcatdirectory
fbveqghechlth	kbcirszxzxscgeukcizjrntclvp	1v11tu8z5okt61njpiky1xoprmr	cherokeeherald
ihyrtyunnaltjm	eiseiondsgkbnzvgwdehxda	sd345o1rq011a1ms3qlley5yvu	admaster
auxiyeexsfcqj	ytwkpzlobljxkljhushyxkyt	1jz5ktklbpm53r2pdymmri043	directory2009
tknbivcmbekpwh	hswvovkduhlbfugqxpmfnjnzn	17adaod1oih6t91x358vyshspil	theupsstore
gtpjifumwmqpn	vwdjxoqworljhirgetwh	1e95km61jytx813ozodwofkggu	expediamail
cnqgglwrucrgp	xcbeeieymbguwddcabueipzwg	970z95v4nzg1qmt2c37ib43h	dyad-inc
aucdtwkdfyewc	pdqfrsvgkkfuwmvgpvvwayyzleu	5a3d2xgu8lq31bbf72q7l7o6c	qimaging

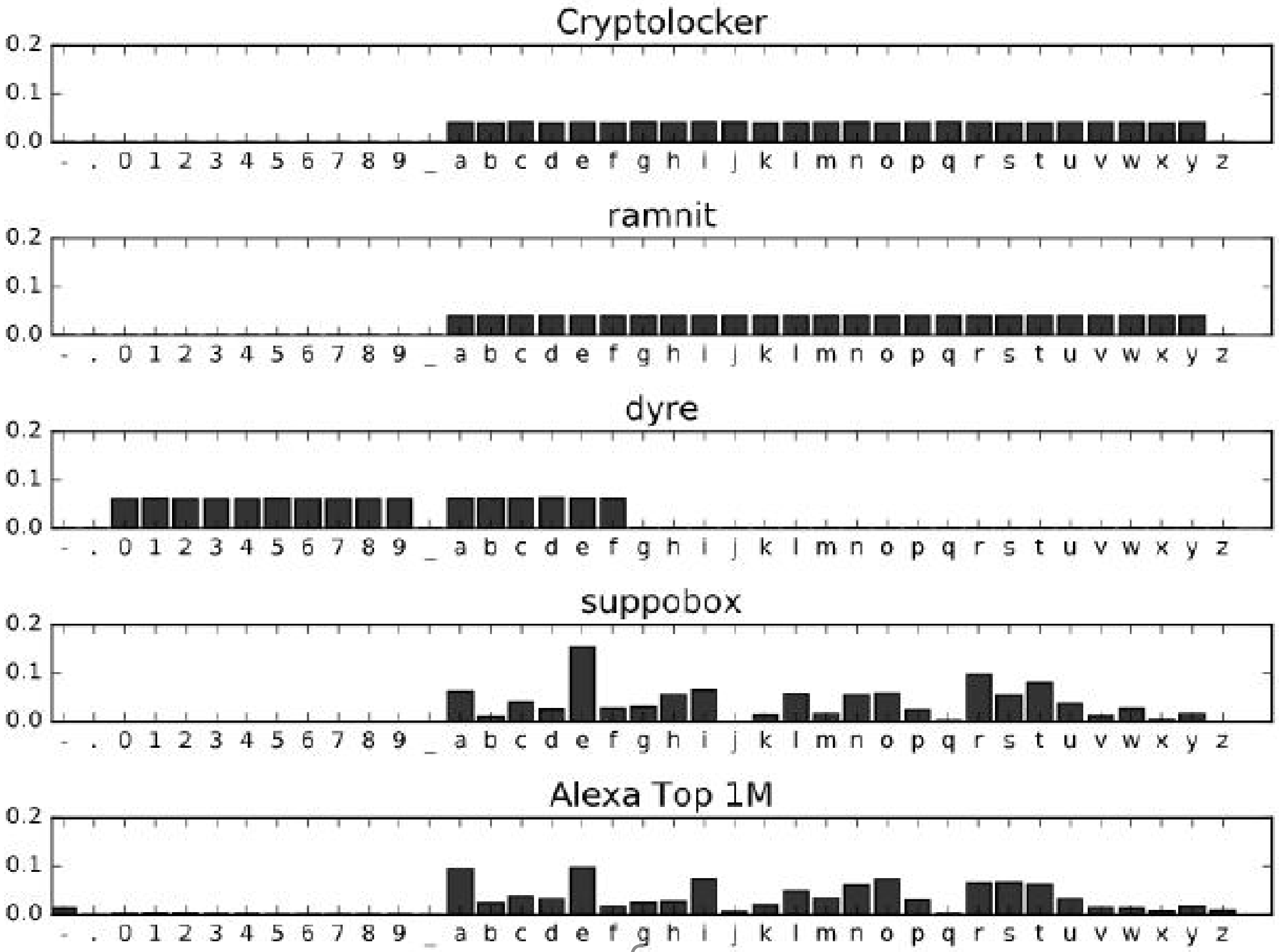
DGA Dictionary [3]



above
action
advance
afraid
against
airplane
almost
alone
already
although
always
amount
anger
angry
animal
another
answer
appear
apple
around
arrive
article
attempt
banker
basket
battle
beauty
became
because
become
before
begin
behind
being
believe
belong
beside
better
between
beyond
bicycle
board
borrow
bottle
bottom
branch
bread
bridge
bright
bring
broad
broken
brought
brown
building
built
business
butter
captain
carry
catch
caught
century
chair
chance
character
charge
chief
childhood
children
choose
cigarette
circle
class
clean
clear
close
clothes
college
company
complete
condition
consider
contain
continue
control
corner
country
course
cover
crowd
daughter
decide
degree
delight
demand
desire
destroy
device
difference
different
difficult
dinner
direct
discover
distance
distant
divide
doctor
dollar
double
doubt
dress
dried
during
early
eearly
effort
either
electric
electricity
english
enough
enter
escape
evening
every
except
expect
experience
explain
family
famous
fancy
father
fellow
fence
fifteen
fight
figure
finger
finish
flier
flower
follow
foreign
forest
forever
forget
fortieth
forward
found
fresh
friend
further
future
garden
gather
general
gentle
gentleman
glass
glossary
goodbye
govern
guard
happen
health
heard
heart
heaven
heavy
history
honor
however
hunger
husband
include
increase
indeed
industry
inside
instead
journey
kitchen
known
labor
ladder
language
large
laugh
laughter
leader
leave
length
letter
likely
listen
little
machine
manner
market
master
material
matter
mayor
measure
meeting
member
method
middle
might
million
minute
mister
modern
morning
mother
mountain
movement
nation
nature
nearly
necessary
needle
neighbor
neither
niece
night
north
nothing
notice
number
object
oclock
office
often
opinion
order
orderly
outside
paint
partial
party
people
perfect
perhaps
period
person
picture
pleasant
please
pleasure
position
possible
power
prepare
present
president
pretty
probable
probably
problem
produce
promise
proud
public
quarter
question
quiet
rather
ready
realize
reason
receive
record
remember
report
require
result
return
ridden
right
river
round
safety
school
season
separate
service
settle
severa
several
shake
share
shore
short
should
shoulder
shout
silver
simple
single
sister
smell
smoke
soldier
space
speak
special
spent
spread
spring
square
station
still
store
storm
straight
strange
stranger
stream
street
strength
strike
strong
student
subject
succeed
success
sudden
suffer
summer
supply
suppose
surprise
sweet
system
therefore
thick
think
third
those
though
thought
through
thrown
together
toward
trade
train
training
travel
trouble
trust
twelve
twenty
understand
understood
until
valley
value
various
wagon
water
weather
welcome
wheat
whether
while
white
whose
window
winter
within
without
woman
women
wonder
worth
would
write
written
yellow

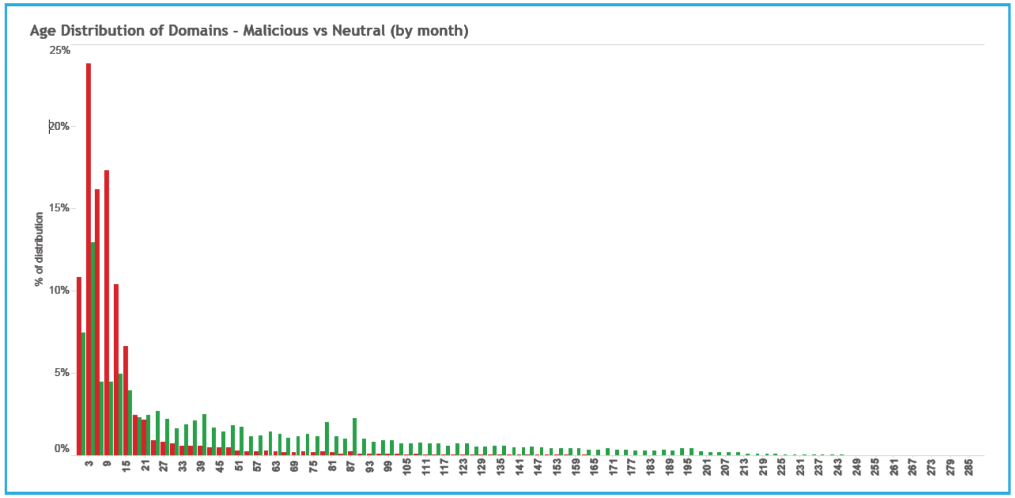
3. 악성 vs 정상 도메인

DGA 및 정상 도메인 샘플 분포 [2]

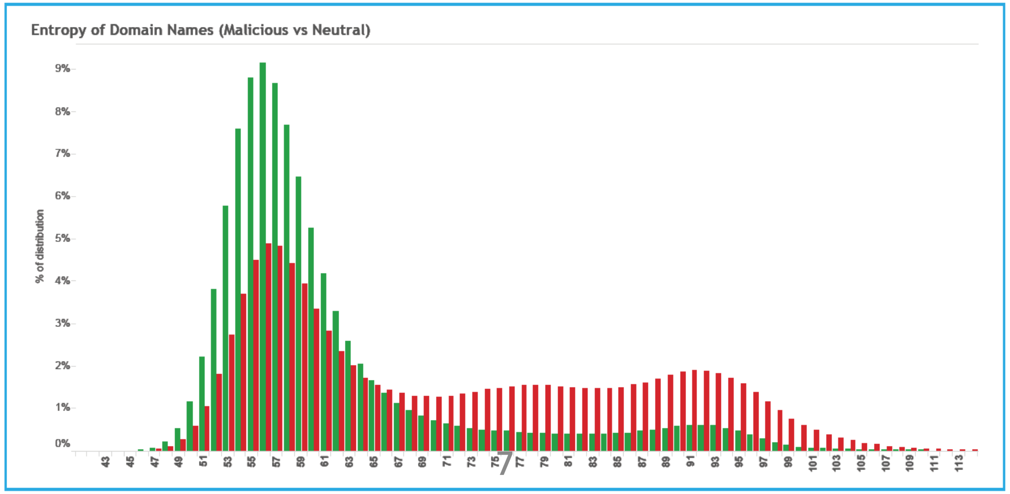


3. 악성 vs 정상 도메인

Age Distribution [6]

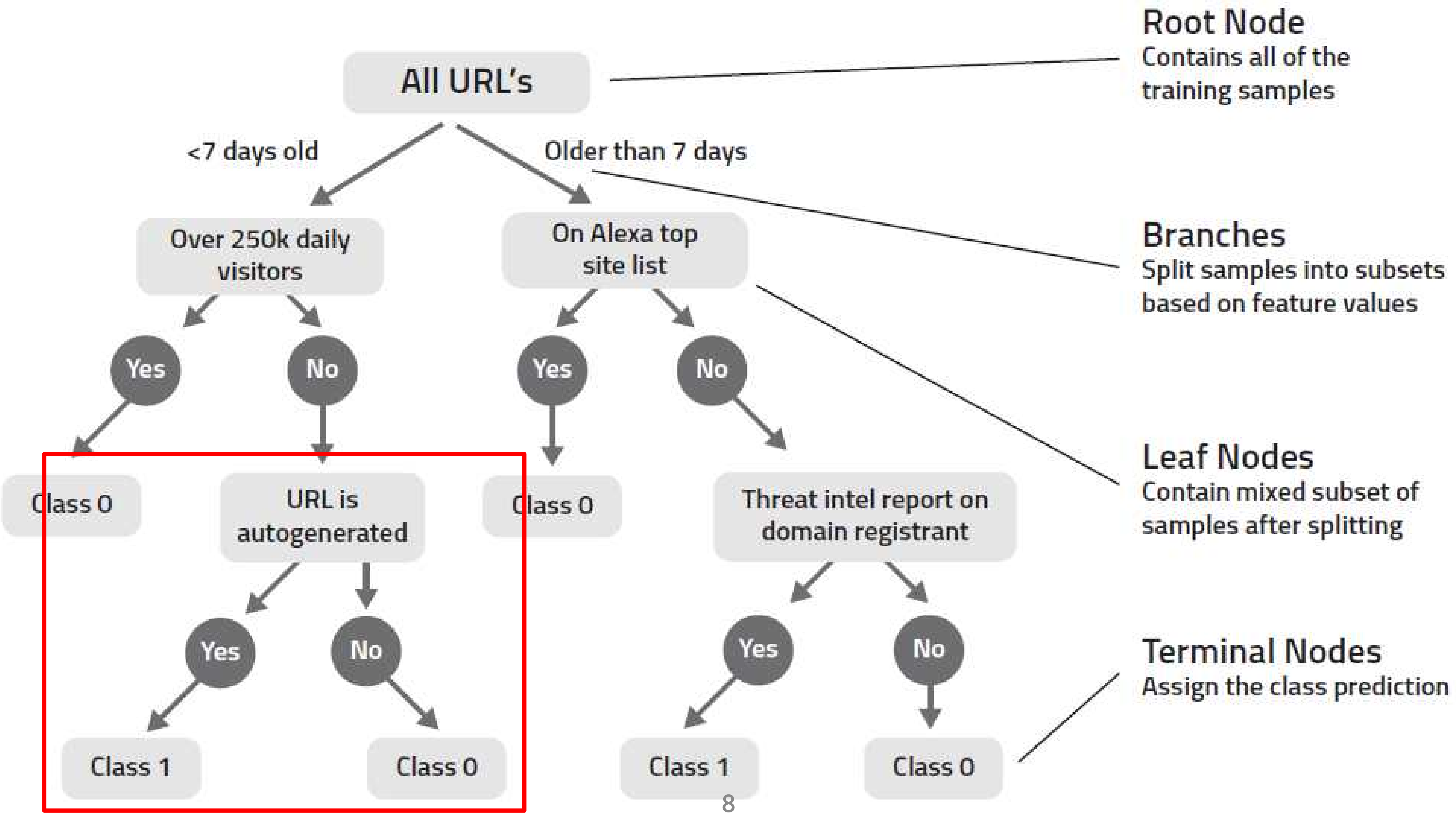


Entropy Distribution [6]



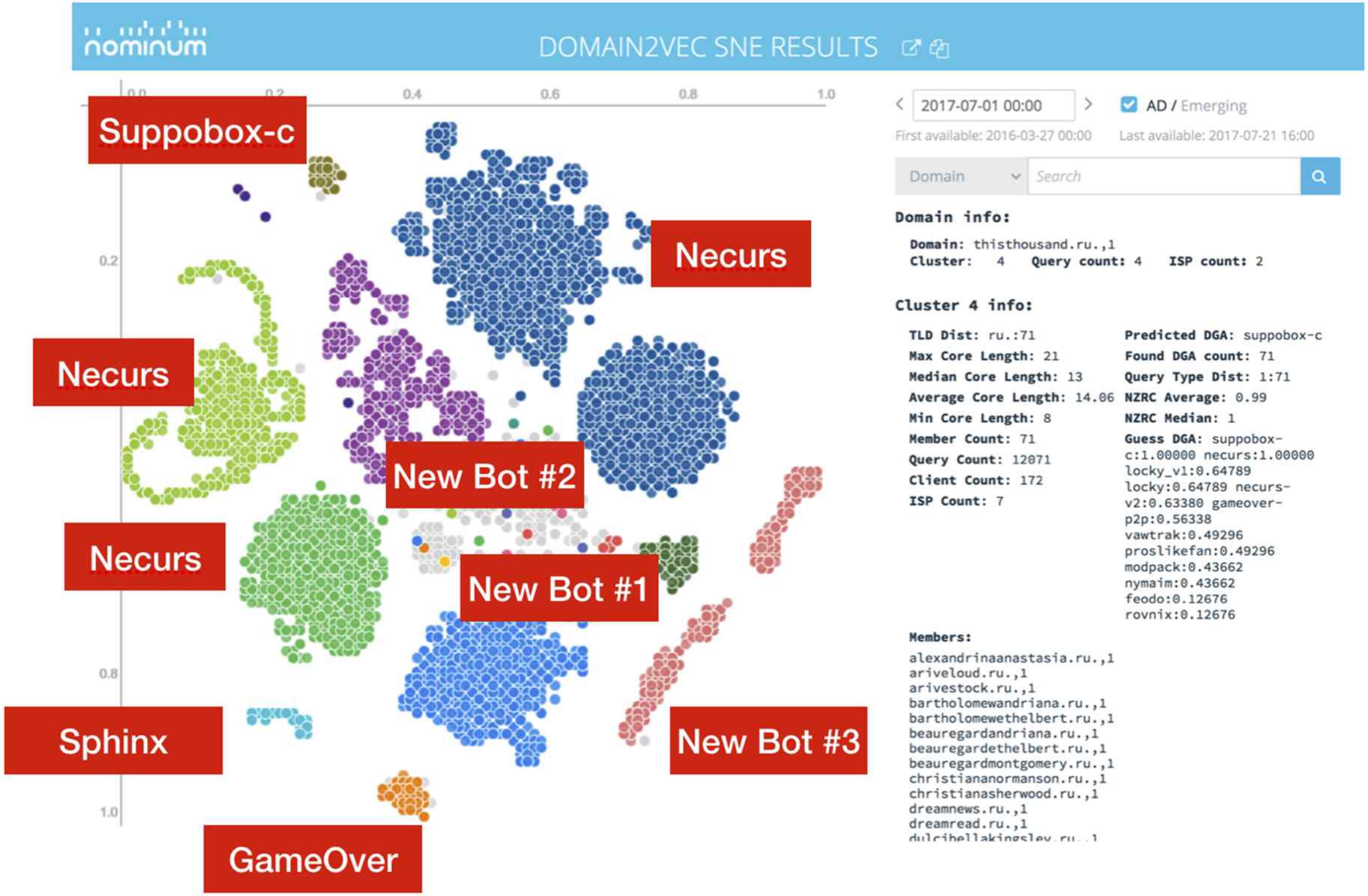
3. 악성 vs 정상 도메인

All Together! [7]



4. DGA 활용 사례

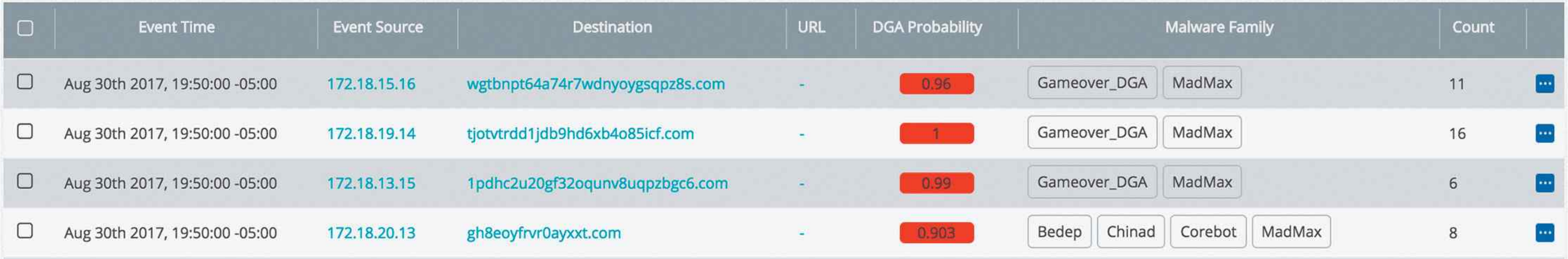
Akamai (Nominum) [4]



```
Domain info:

  Domain: thisthousand.ru.,1
  Cluster:   4   Query count: 4   ISP count: 2

Cluster 4 info:

  TLD Dist: ru.:71                  Predicted DGA: suppobox-c
  Max Core Length: 21               Found DGA count: 71
  Median Core Length: 13            Query Type Dist: 1:71
  Average Core Length: 14.06        NZRC Average: 0.99
  Min Core Length: 8                NZRC Median: 1
  Member Count: 71                  Guess DGA: suppobox-
  Query Count: 12071                c:1.00000 necurs:1.00000
  Client Count: 172                 locky_v1:0.64789
  ISP Count: 7                      locky:0.64789 necurs-
                                    v2:0.63380 gameover-
                                    p2p:0.56338
                                    vawtrak:0.49296
                                    proslikefan:0.49296
                                    modpack:0.43662
                                    nymaim:0.43662
                                    feodo:0.12676
                                    rovnix:0.12676

  Members:

  alexandrinaanastasia.ru.,1
  ariveloud.ru.,1
  arivestock.ru.,1
  bartholomewandriana.ru.,1
  bartholomewethelbert.ru.,1
  beauregardandriana.ru.,1
  beauregardethelbert.ru.,1
  beauregardmontgomery.ru.,1
  christiananormanson.ru.,1
  christianasherwood.ru.,1
  dreamnews.ru.,1
  dreamread.ru.,1
  dulcibellakingsley.ru.,1
```

4. DGA 활용 사례

Anomali Enterprise [8]

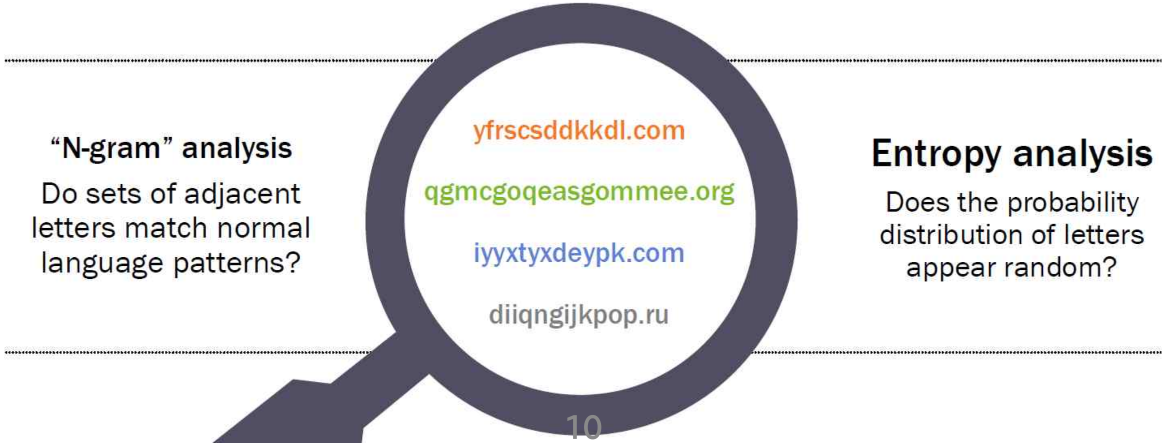


□	Event Time	Event Source	Destination	URL	DGA Probability	Malware Family	Count	
□	Aug 30th 2017, 19:50:00 -05:00	172.18.15.16	wgtbnpt64a74r7wdnyoygsqpz8s.com	-	0.96	Gameover_DGA MadMax	11	···
□	Aug 30th 2017, 19:50:00 -05:00	172.18.19.14	tjotvtrdd1jdb9hd6xb4o85icf.com	-	1	Gameover_DGA MadMax	16	···
□	Aug 30th 2017, 19:50:00 -05:00	172.18.13.15	1pdhc2u20gf32oqunv8uqpzbgc6.com	-	0.99	Gameover_DGA MadMax	6	···
□	Aug 30th 2017, 19:50:00 -05:00	172.18.20.13	gh8eoyfrvr0ayxxt.com	-	0.903	Bedep Chinad Corebot MadMax	8	···

Cisco Umbrella (OpenDNS)

DGA Detection

Identifies malicious domain-squatting and targeted C2 or phishing domains



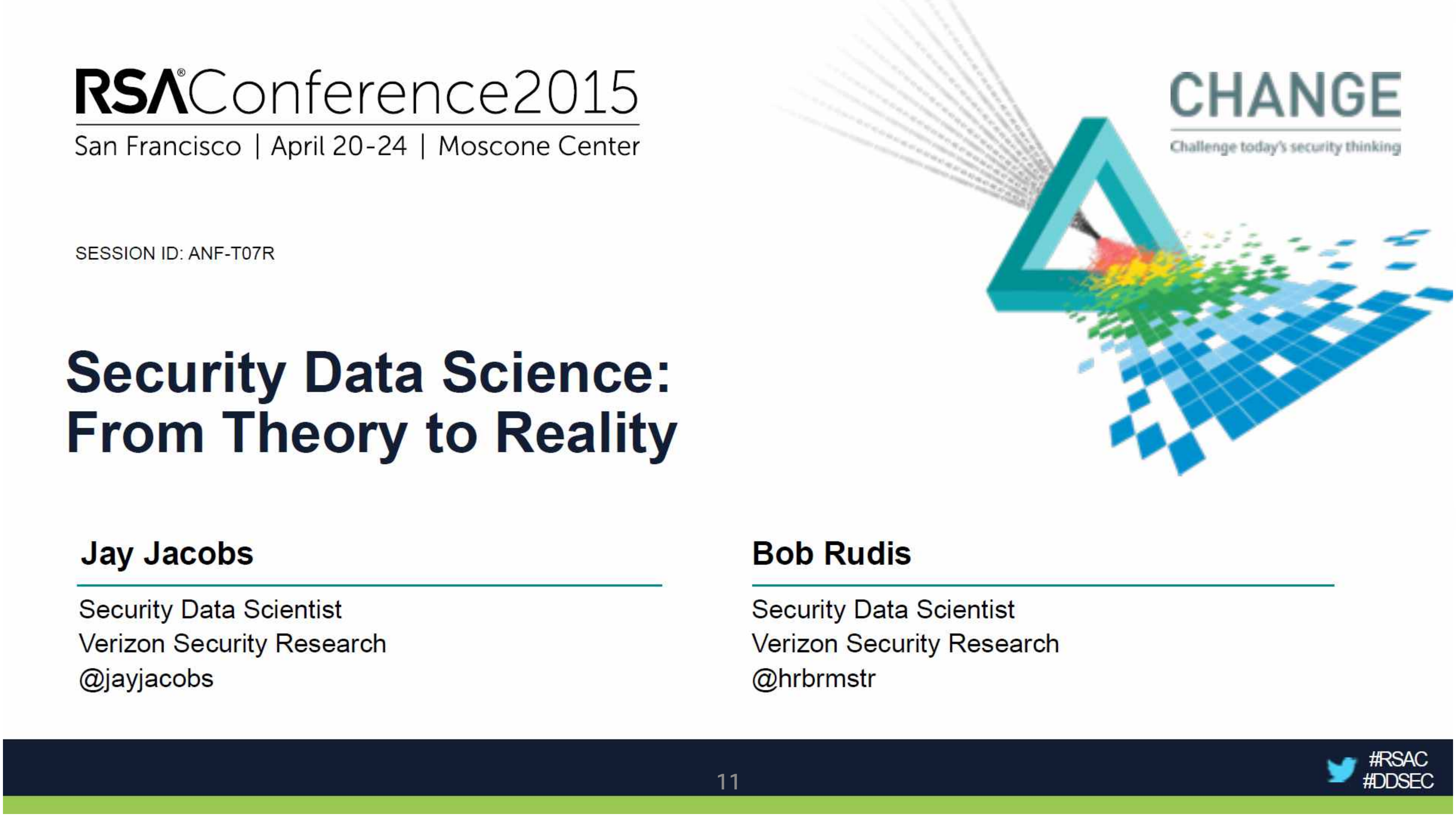
“N-gram” analysis

Do sets of adjacent letters match normal language patterns?

Entropy analysis

Does the probability distribution of letters appear random?

5. 머신러닝을 이용한 악성 도메인 탐지 기법 [1]



RSAConference2015

San Francisco | April 20-24 | Moscone Center

CHANGE

Challenge today's security thinking

SESSION ID: ANF-T07R

Security Data Science: From Theory to Reality

Jay Jacobs

Security Data Scientist
Verizon Security Research
@jayjacobs

Bob Rudis

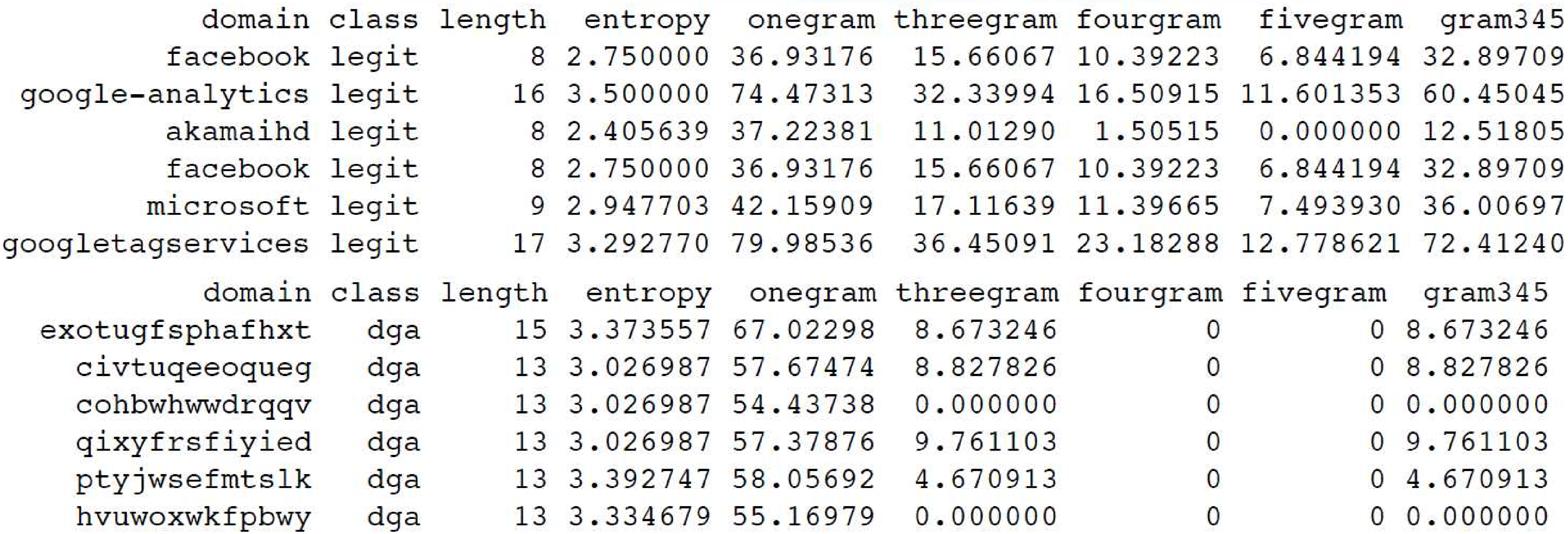
Security Data Scientist
Verizon Security Research
@hrbrmstr

#RSAC
#DDSEC

5. 머신러닝을 이용한 악성 도메인 탐지 기법 [1]

Domain Features

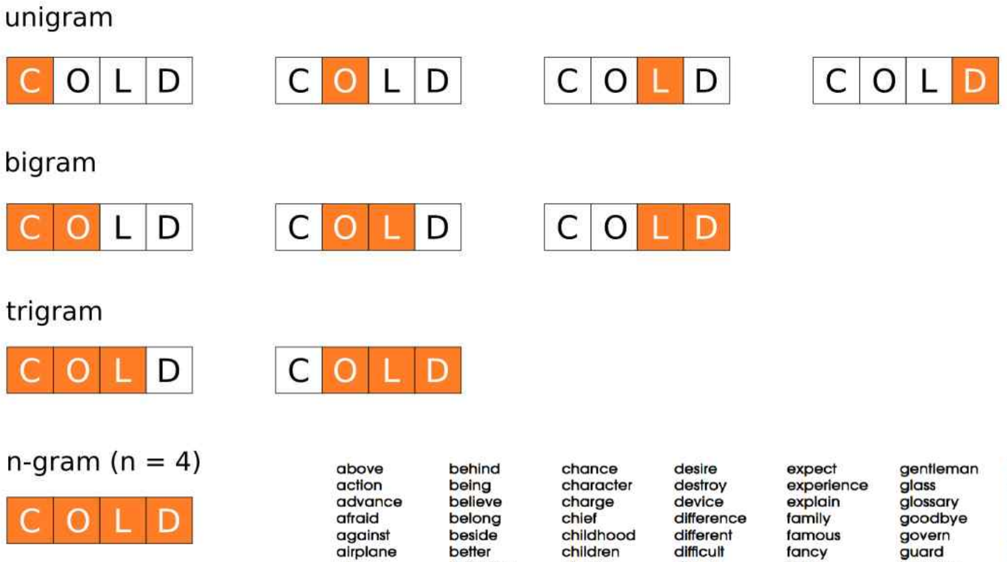
- Length
- Entropy
- letter sequences (n-grams)
- Others?



```
            domain class length  entropy  onegram threegram fourgram  fivegram  gram345
          facebook legit      8 2.750000 36.93176  15.66067 10.39223  6.844194 32.89709
  google-analytics legit     16 3.500000 74.47313  32.33994 16.50915 11.601353 60.45045
          akamaihd legit      8 2.405639 37.22381  11.01290  1.50515  0.000000 12.51805
          facebook legit      8 2.750000 36.93176  15.66067 10.39223  6.844194 32.89709
         microsoft legit      9 2.947703 42.15909  17.11639 11.39665  7.493930 36.00697
googletagservices legit     17 3.292770 79.98536  36.45091 23.18288 12.778621 72.41240

            domain class length  entropy  onegram threegram fourgram fivegram  gram345
   exotugfsphafhxt   dga     15 3.373557 67.02298  8.673246        0        0 8.673246
     civtuqeeoqueg   dga     13 3.026987 57.67474  8.827826        0        0 8.827826
     cohbwhwwdrqqv   dga     13 3.026987 54.43738  0.000000        0        0 0.000000
     qixyfrsfiyied   dga     13 3.026987 57.37876  9.761103        0        0 9.761103
     ptyjwsefmtslk   dga     13 3.392747 58.05692  4.670913        0        0 4.670913
     hvuwoxwkfpbwy   dga     13 3.334679 55.16979  0.000000        0        0 0.000000
```

5. 머신러닝을 이용한 악성 도메인 탐지 기법 [1]

n-grams & entropy



unigram

C O L D   C O L D   C O L D   C O L D

bigram

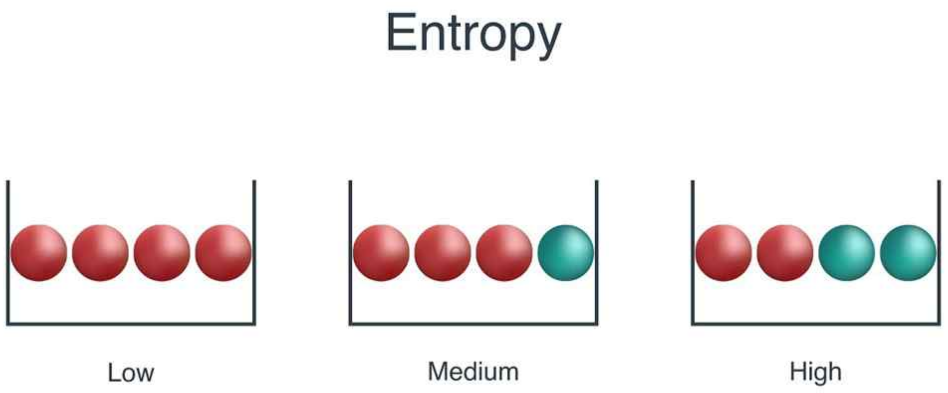
C O L D   C O L D   C O L D

trigram

C O L D   C O L D

n-gram (n = 4)

C O L D



above
action
advance
afraid
against
airplane
almost
alone
already
although
always
amount
anger
angry
animal
another
answer
appear
apple
around
arrive
article
attempt
banker
basket
battle
beauty
became
because
become
before
begin

behind
being
believe
belong
beside
better
between
beyond
bicycle
board
borrow
bottle
bottom
branch
bread
bridge
bright
bring
broad
broken
brought
brown
building
built
business
butter
captain
carry
catch
caught
century
chair

chance
character
charge
chief
childhood
children
choose
cigarette
circle
class
clean
clear
close
clothes
college
company
complete
condition
consider
contain
continue
control
corner
country
course
cover
crowd
daughter
decide
degree
delight
demand

desire
destroy
device
difference
different
difficult
dinner
direct
discover
distance
distant
divide
doctor
dollar
double
doubt
dress
dried
during
early
eearly
effort
either
electric
electricity
english
enough
enter
escape
evening
every
except

expect
experience
explain
family
famous
fancy
father
fellow
fence
fifteen
fight
figure
finger
finish
flier
flower
follow
foreign
forest
forever
forget
fortieth
forward
found
fresh
friend
further
future
garden
gather
general
gentle

gentleman
glass
glossary
goodbye
govern
guard
happen
health
heard
heart
heaven
heavy
history
honor
however
hunger
husband
include
increase
indeed
industry
inside
instead
journey
kitchen
known
labor
ladder
language
large
laugh
laughter

leader
leave
length
letter
likely
listen
little
machine
manner
market
master
material
matter
mayor
measure
meeting
member
method
middle
might
million
minute
mister
modern
morning
mother
mountain
movement
nation
nature
nearly
necessary

needle
neighbor
neither
niece
night
north
nothing
notice
number
object
oclock
office
often
opinion
order
orderly
outside
paint
partial
party
people
perfect
perhaps
period
person
picture
pleasant
please
pleasure
position
possible
power

prepare
present
president
pretty
probable
probably
problem
produce
promise
proud
public
quarter
question
quiet
rather
ready
realize
reason
receive
record
remember
report
require
result
return
ridden
right
river
round
safety
school
season

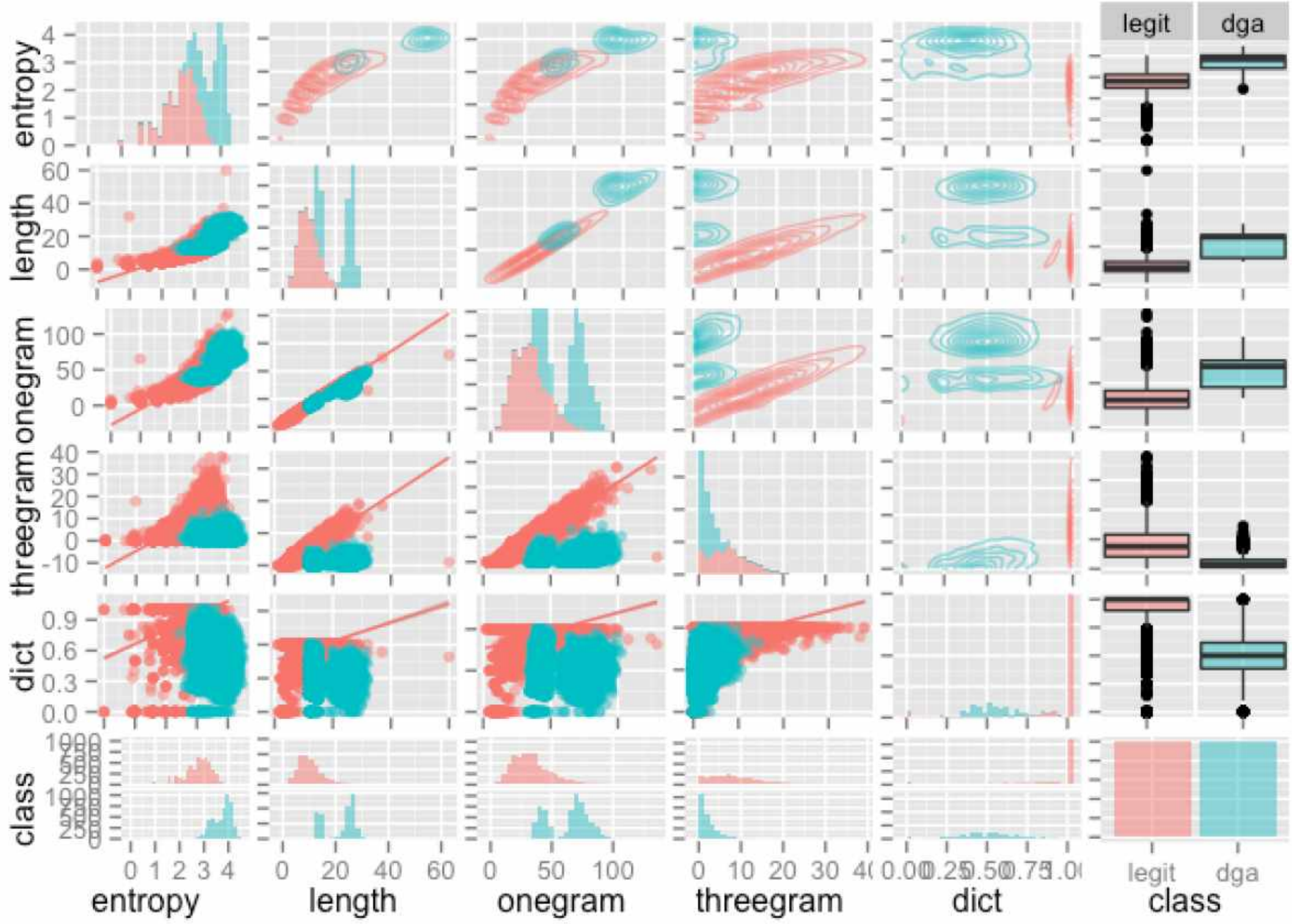
separate
service
settle
severa
several
shake
share
shore
short
should
shoulder
shout
silver
simple
single
sister
smell
smoke
soldier
space
speak
special
spent
spread
spring
square
station
still
store
storm
straight
strange

stranger
stream
street
strength
strike
strong
student
subject
succeed
success
sudden
suffer
summer
supply
suppose
surprise
sweet
system
therefore
thick
think
third
those
though
thought
through
thrown
together
toward
trade
train
training

travel
trouble
trust
twelve
twenty
understand
understood
until
valley
value
various
wagon
water
weather
welcome
wheat
whether
while
white
whose
window
winter
within
without
woman
women
wonder
worth
would
write
written
yellow

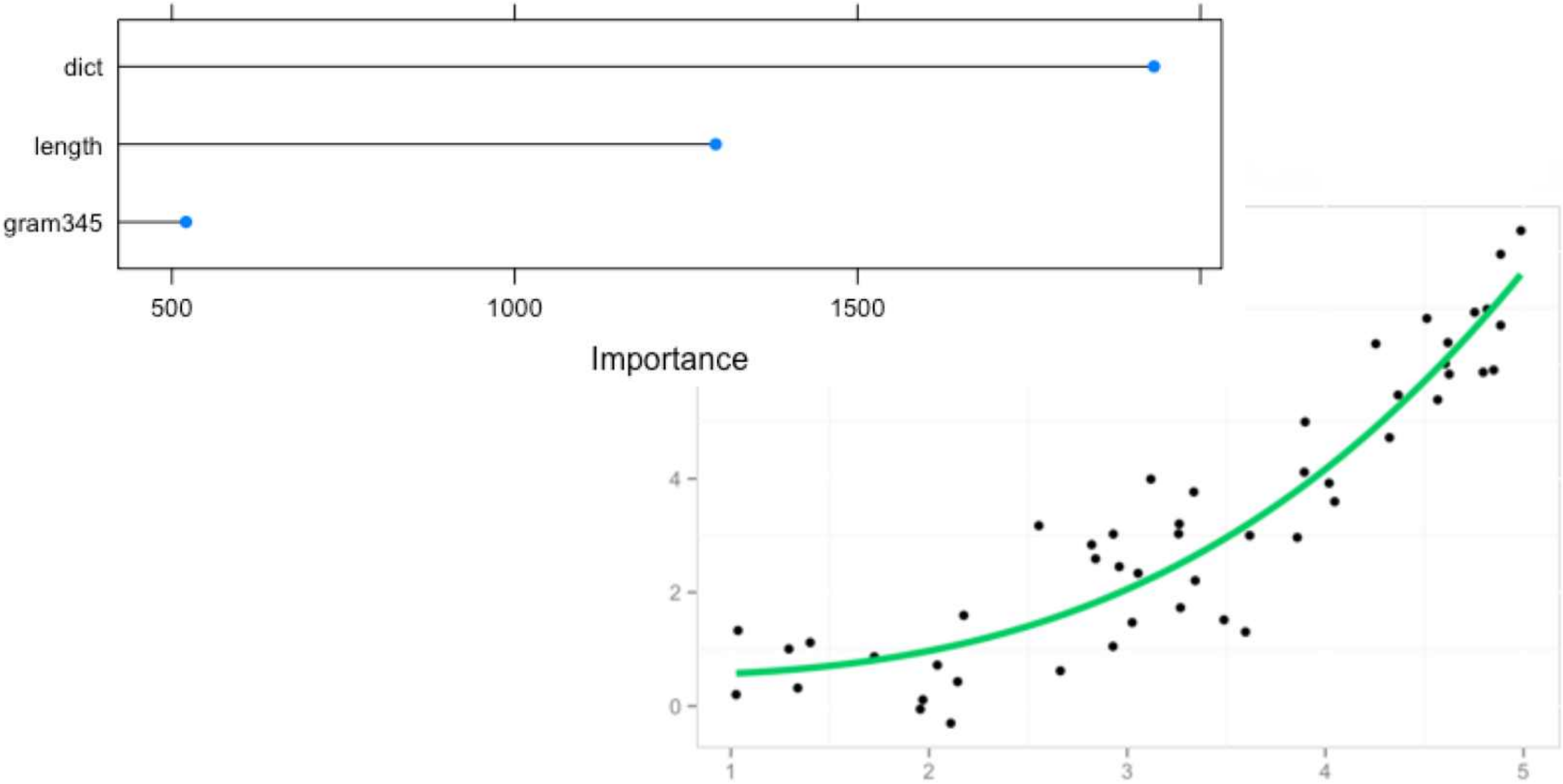
5. 머신러닝을 이용한 악성 도메인 탐지 기법 [1]

Comparing all the Features…



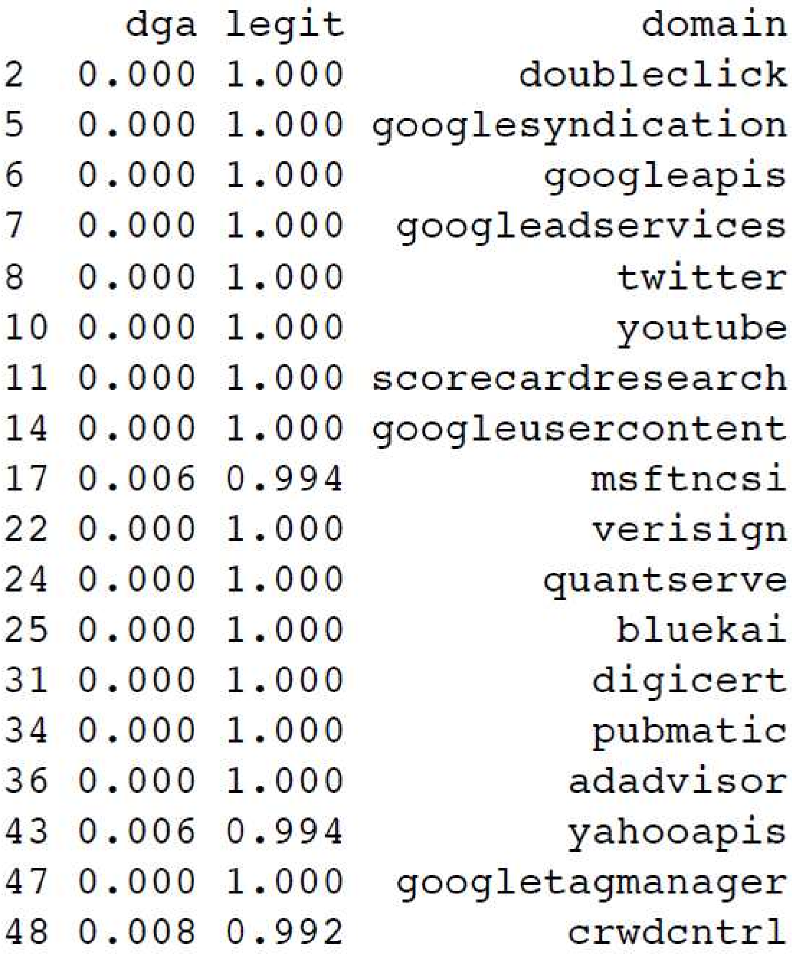
5. 머신러닝을 이용한 악성 도메인 탐지 기법 [1]

Training with selected Features

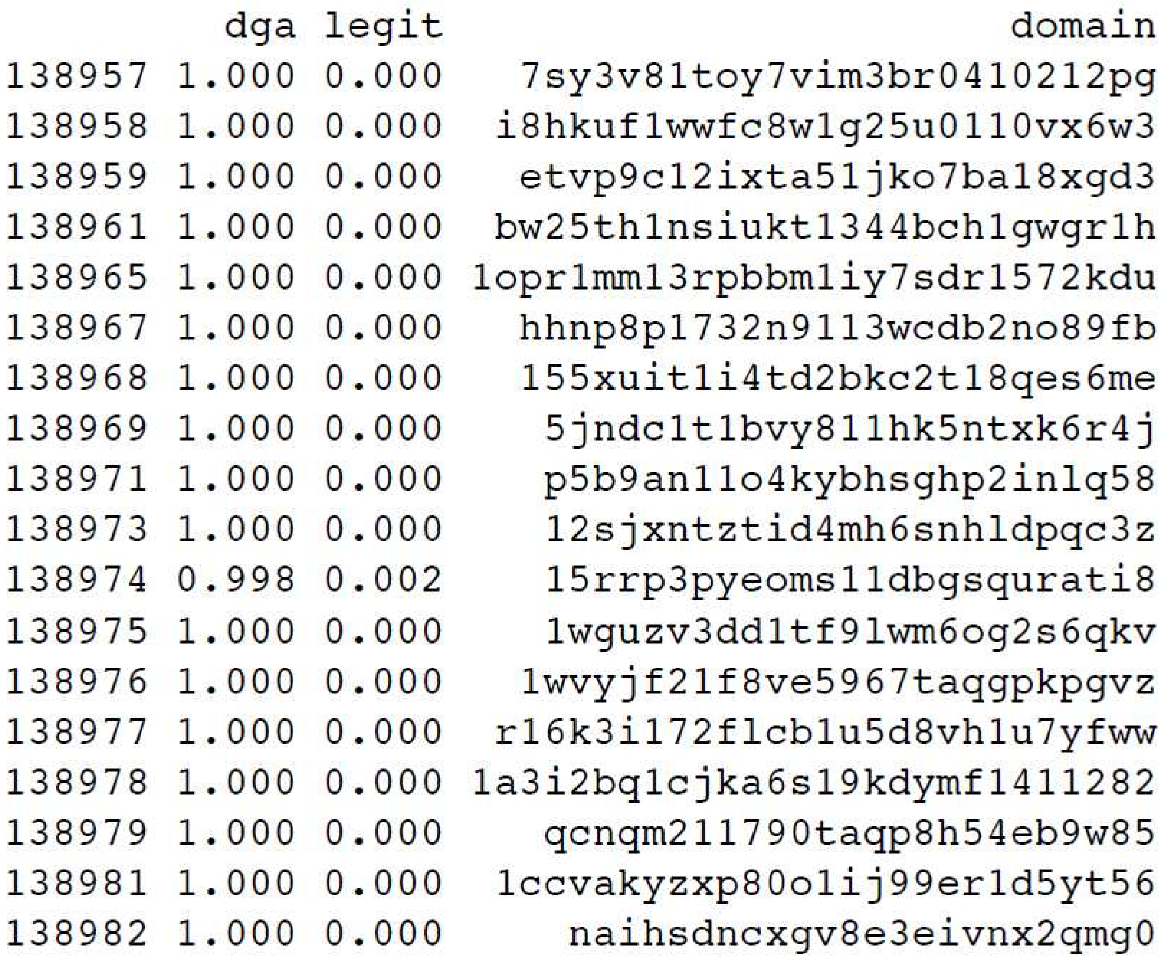


5. 머신러닝을 이용한 악성 도메인 탐지 기법 [1]

The Result (Black & White)



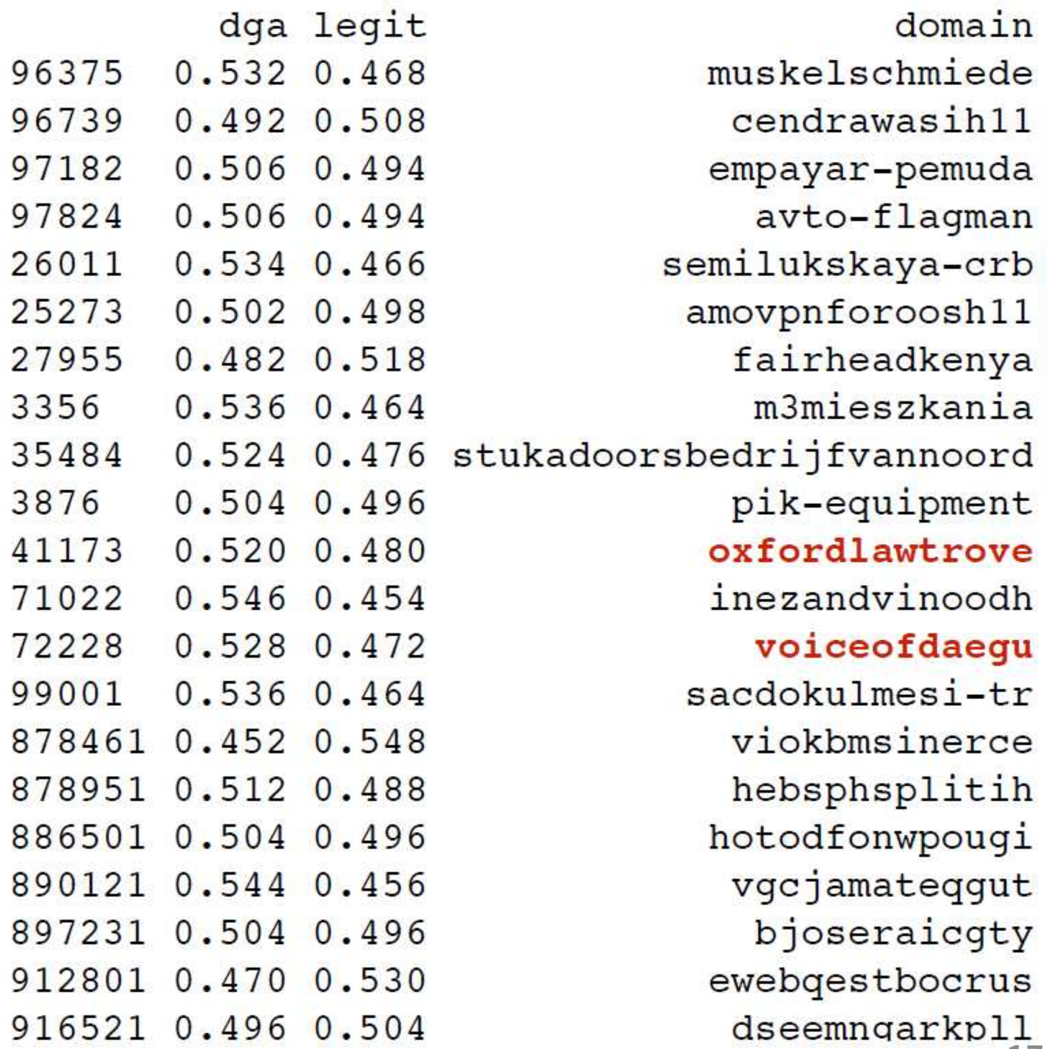
```
     dga legit            domain
2  0.000 1.000       doubleclick
5  0.000 1.000 googlesyndication
6  0.000 1.000        googleapis
7  0.000 1.000  googleadservices
8  0.000 1.000           twitter
10 0.000 1.000           youtube
11 0.000 1.000 scorecardresearch
14 0.000 1.000 googleusercontent
17 0.006 0.994          msftncsi
22 0.000 1.000          verisign
24 0.000 1.000        quantserve
25 0.000 1.000           bluekai
31 0.000 1.000          digicert
34 0.000 1.000         pubmatic
36 0.000 1.000         adadvisor
43 0.006 0.994        yahooapis
47 0.000 1.000  googletagmanager
48 0.008 0.992          crwdcntrl
```



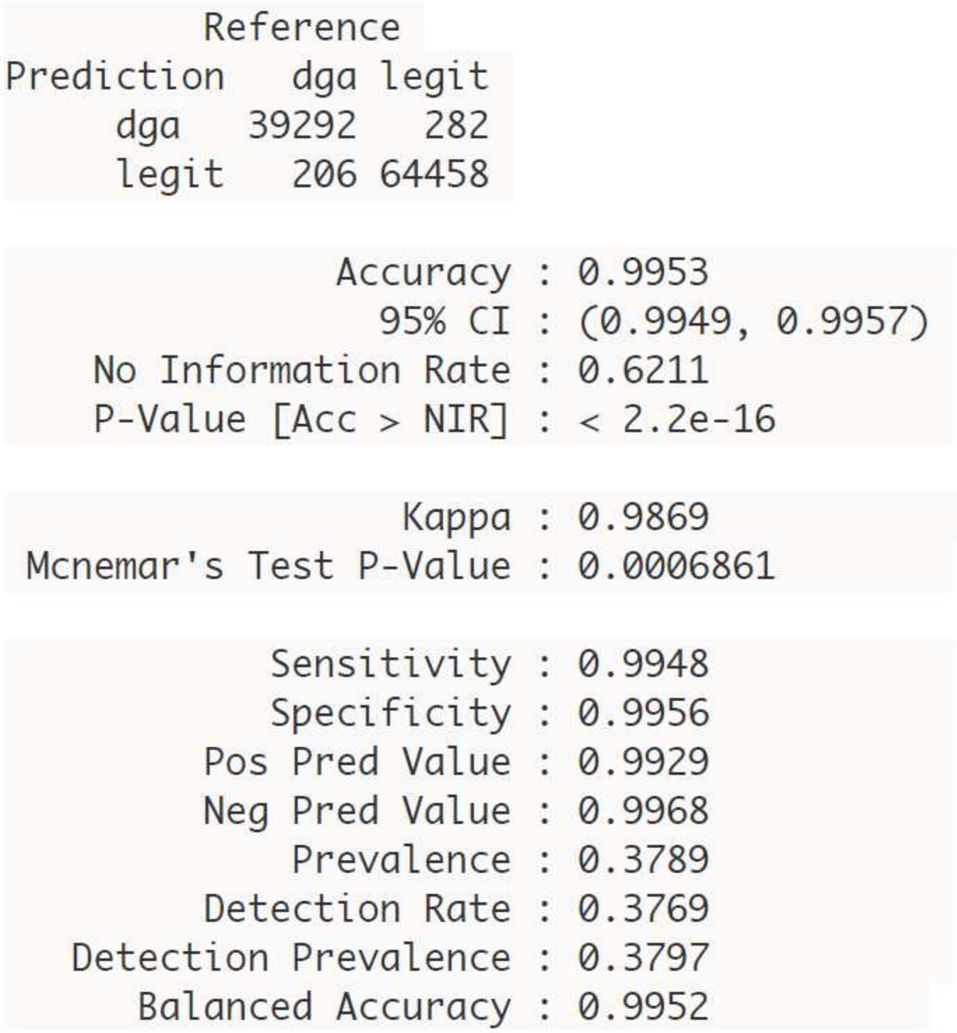
```
         dga legit                       domain
138957 1.000 0.000   7sy3v81toy7vim3br0410212pg
138958 1.000 0.000  i8hkuf1wwfc8w1g25u0110vx6w3
138959 1.000 0.000   etvp9c12ixta51jko7ba18xgd3
138961 1.000 0.000  bw25th1nsiukt1344bch1gwgr1h
138965 1.000 0.000 1opr1mm13rpbbm1iy7sdr1572kdu
138967 1.000 0.000   hhnp8p1732n9113wcdb2no89fb
138968 1.000 0.000   155xuit1i4td2bkc2t18qes6me
138969 1.000 0.000    5jndc1t1bvy811hk5ntxk6r4j
138971 1.000 0.000    p5b9an11o4kybhsghp2inlq58
138973 1.000 0.000    12sjxntztid4mh6snhldpqc3z
138974 0.998 0.002    15rrp3pyeoms11dbgsqurati8
138975 1.000 0.000    1wguzv3dd1tf91wm6og2s6qkv
138976 1.000 0.000   1wvyjf21f8ve5967taqgpkpgvz
138977 1.000 0.000  r16k3i172f1cb1u5d8vh1u7yfww
138978 1.000 0.000 1a3i2bq1cjka6s19kdymf1411282
138979 1.000 0.000    qcnqm211790taqp8h54eb9w85
138981 1.000 0.000  1ccvakyzxp80o1ij99er1d5yt56
138982 1.000 0.000     naihsdncxgv8e3eivnx2qmg0
```

5. 머신러닝을 이용한 악성 도메인 탐지 기법 [1]

The Result (Gray)



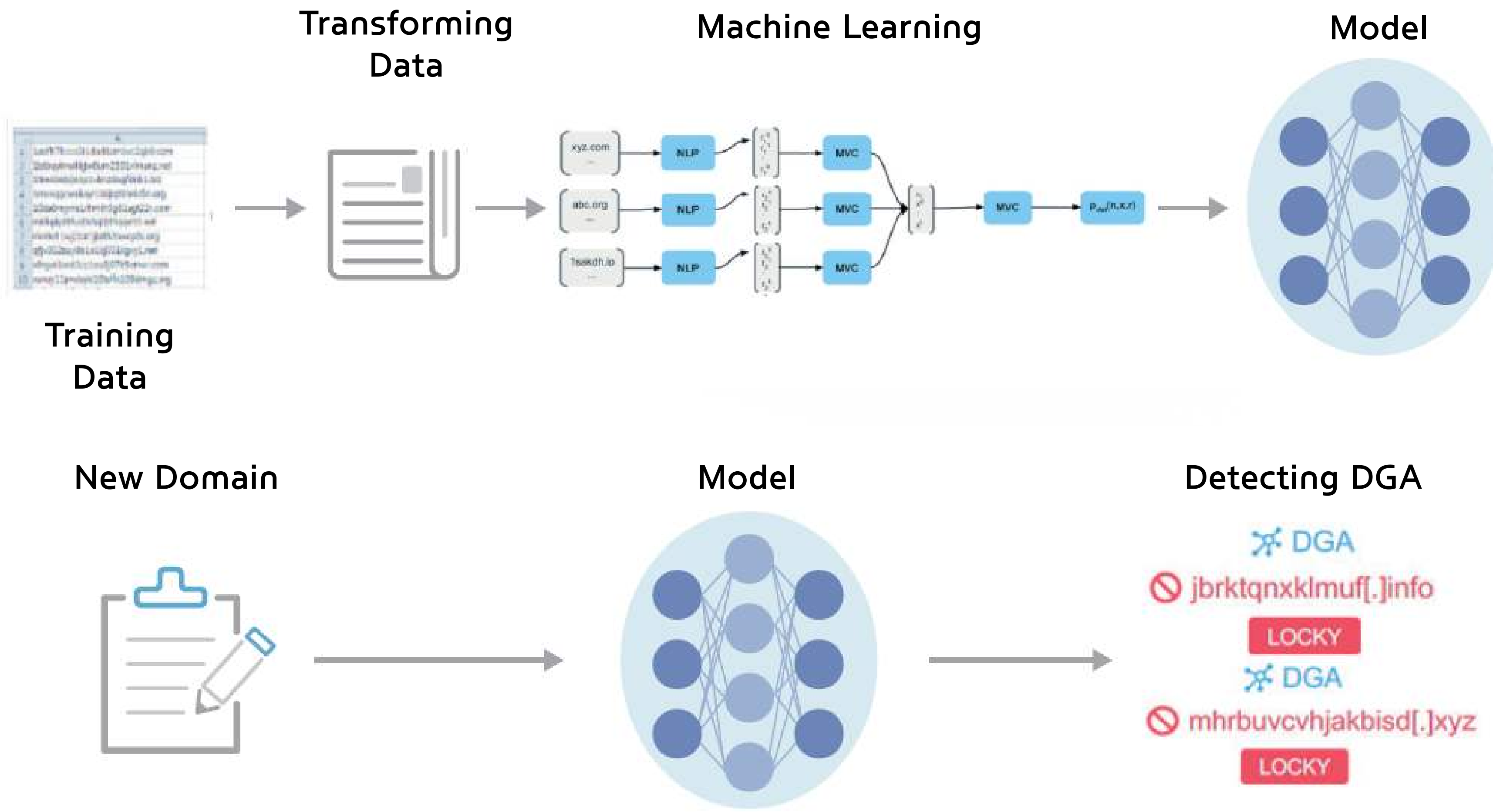
```
         dga  legit                      domain
96375   0.532 0.468              muskelschmiede
96739   0.492 0.508               cendrawasih11
97182   0.506 0.494              empayar-pemuda
97824   0.506 0.494                avto-flagman
26011   0.534 0.466            semilukskaya-crb
25273   0.502 0.498             amovpnforoosh11
27955   0.482 0.518                fairheadkenya
3356    0.536 0.464                 m3mieszkania
35484   0.524 0.476 stukadoorsbedrijfvannoord
3876    0.504 0.496                pik-equipment
41173   0.520 0.480               oxfordlawtrove
71022   0.546 0.454               inezandvinoodh
72228   0.528 0.472                 voiceofdaegu
99001   0.536 0.464             sacdokulmesi-tr
878461  0.452 0.548                viokbmsinerce
878951  0.512 0.488                hebsphsplitih
886501  0.504 0.496               hotodfonwpougi
890121  0.544 0.456                vgcjamateqgut
897231  0.504 0.496                 bjoseraicgty
912801  0.470 0.530               ewebqestbocrus
916521  0.496 0.504                dseemnqarkpll
```



```
          Reference
Prediction   dga legit
     dga   39292   282
     legit   206 64458

               Accuracy : 0.9953
                 95% CI : (0.9949, 0.9957)
    No Information Rate : 0.6211
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.9869
 Mcnemar's Test P-Value : 0.0006861

            Sensitivity : 0.9948
            Specificity : 0.9956
         Pos Pred Value : 0.9929
         Neg Pred Value : 0.9968
             Prevalence : 0.3789
         Detection Rate : 0.3769
   Detection Prevalence : 0.3797
      Balanced Accuracy : 0.9952
```

6. 딥러닝을 이용한 악성 도메인 탐지 기법 [2]

DGA 도메인 탐지 기법 개요



Transforming Data

Machine Learning

Model

Training Data

New Domain

Model

Detecting DGA

DGA

jbrktqnxklmuf[.]info

LOCKY

DGA

mhrbuvcvhjakbisd[.]xyz

LOCKY

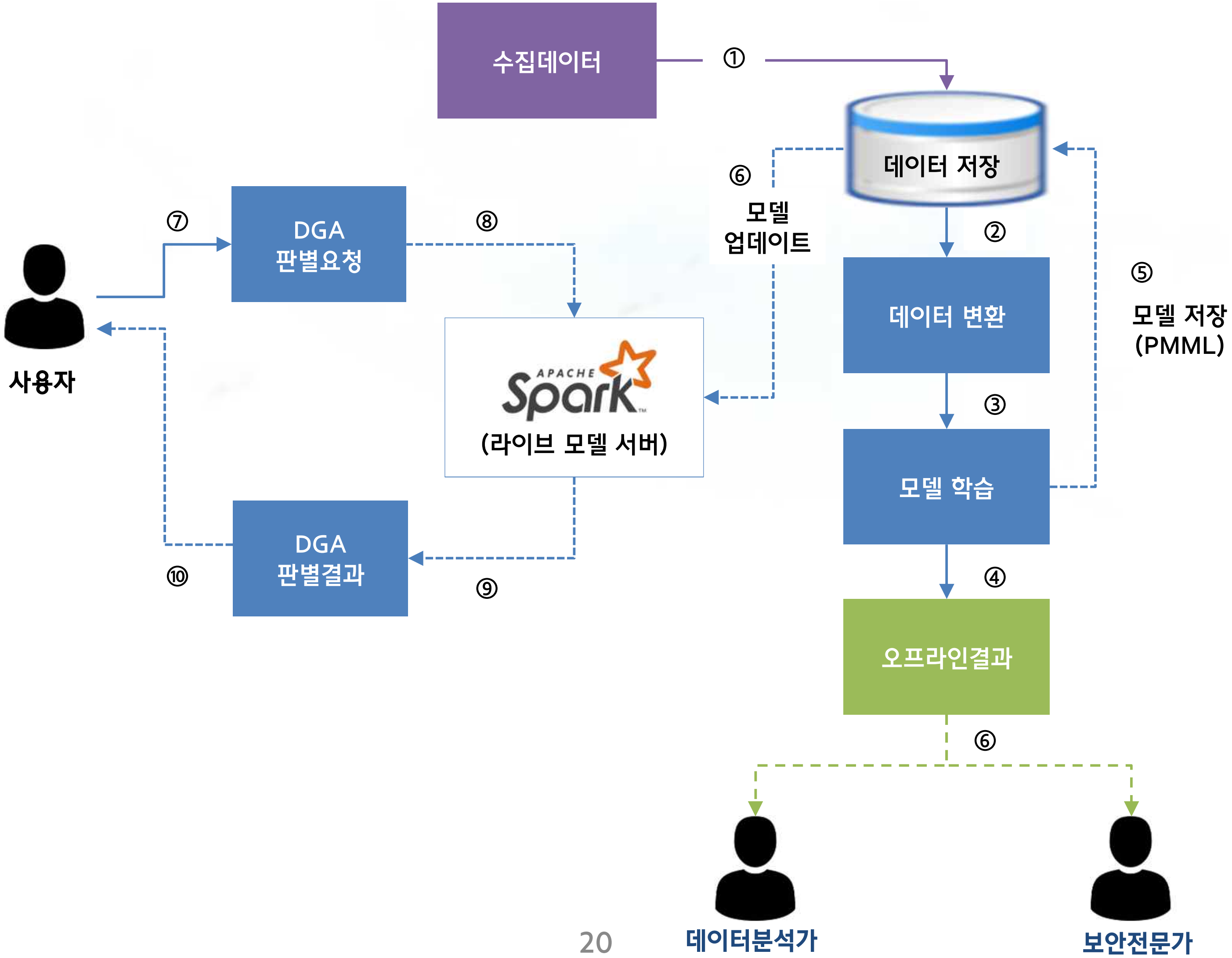
6. 딥러닝을 이용한 악성 도메인 탐지 기법 [2]

딥러닝 학습과정

수집방법	• 국내 · 외 위협 인텔리전스 수집 • KISC 침해사고 분석 및 대응 업무 • C-TAS 위협정보 수집
학습 데이터	• 악성 도메인에 대한 분석 정보 • 국내 · 외 정상 도메인 등록정보
학습 모델	① 악성 도메인에 대한 분석 정보를 이용하여 자동 생성된 도메인 분류 ※ 악성 도메인 명을 자동으로 생성하는 DGA(Domain Generation Algorithm) 알고리즘 및 분류(Classification) 알고리즘 사용 ② 자동 생성된 도메인과 정상 도메인에 대한 정보를 학습하여 신규 도메인에 대한 악성 여부 판단 ※ 신규 도메인에 대한 악성 여부 판단을 위해 딥러닝 사용
KISC 활용	① 악성 도메인에 대한 자동생성 여부 판단 및 관련정보 공유 ② 신규 등록된 국내 · 외 도메인에 대한 악성 여부 판단 및 사전 차단

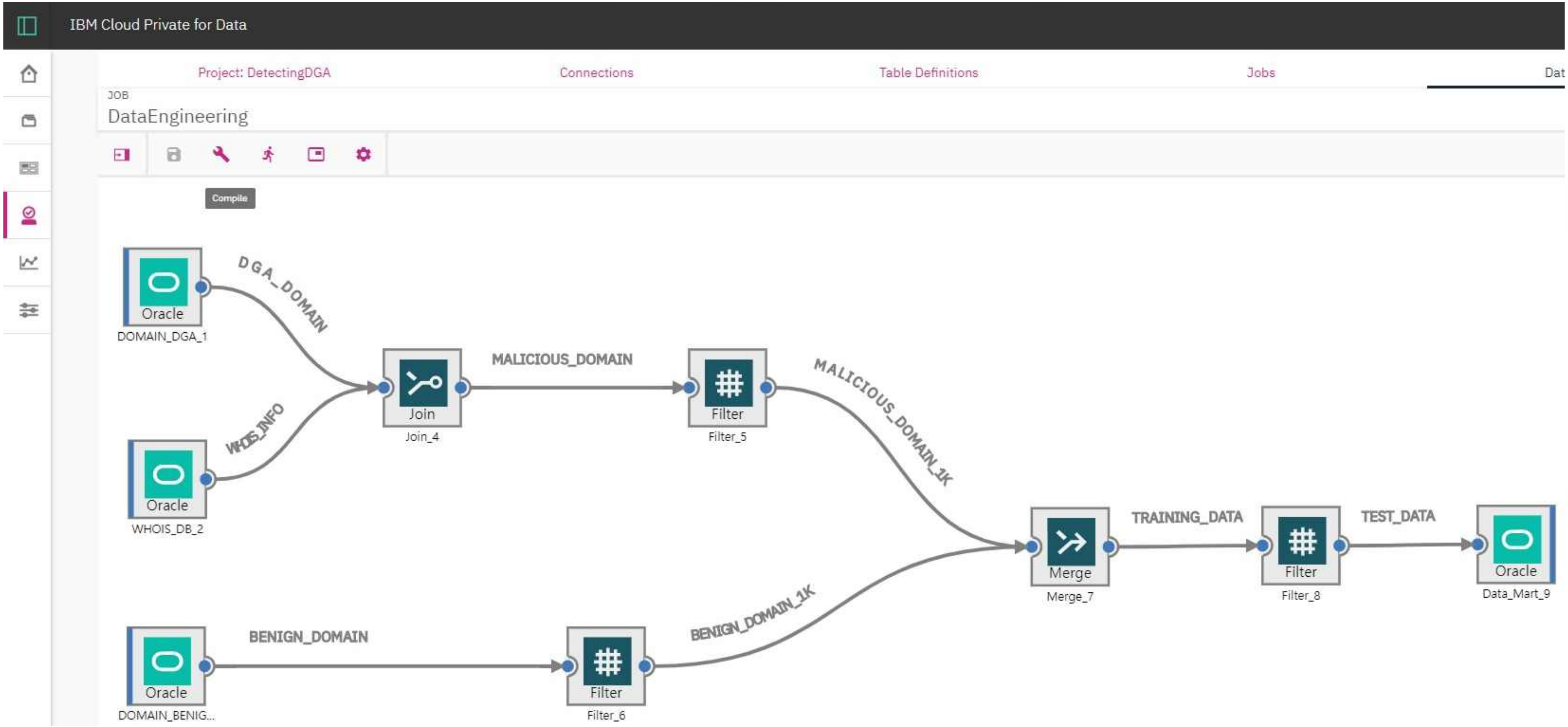
6. 딥러닝을 이용한 악성 도메인 탐지 기법 [2]

딥러닝 학습과정



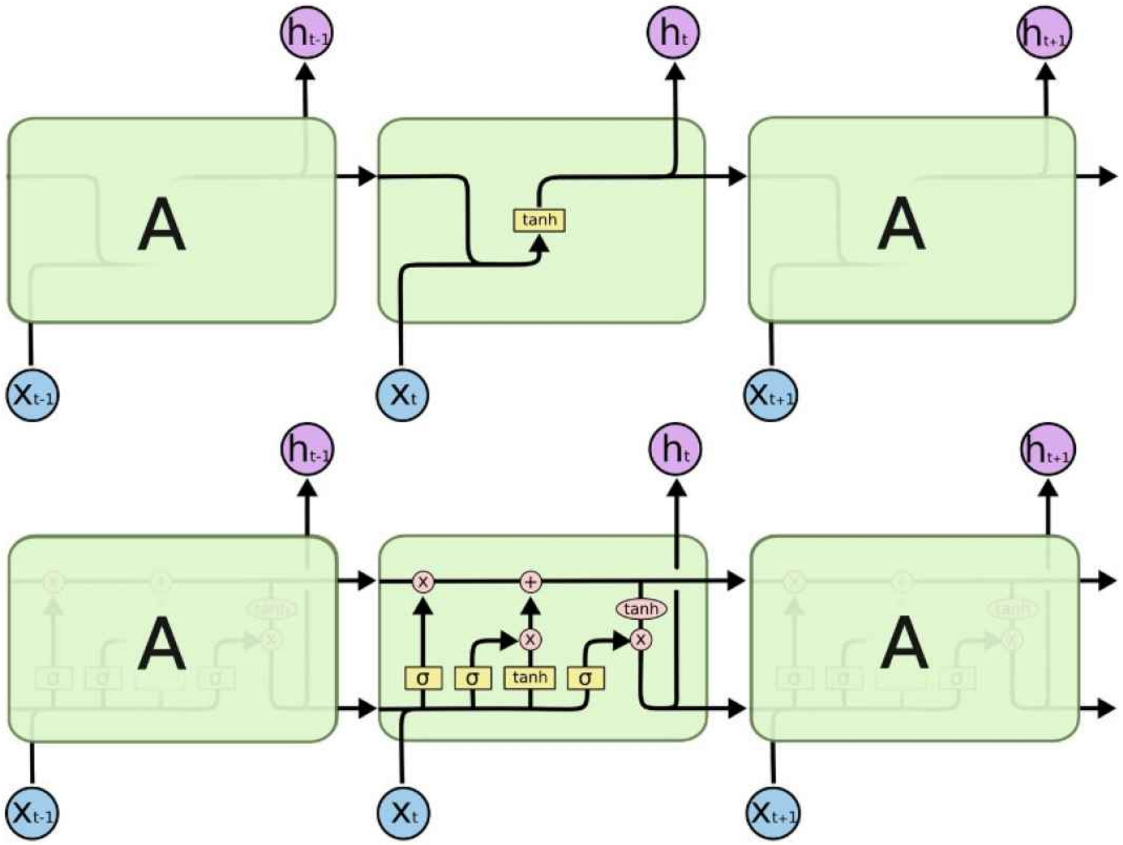
6. 딥러닝을 이용한 악성 도메인 탐지 기법 [2]

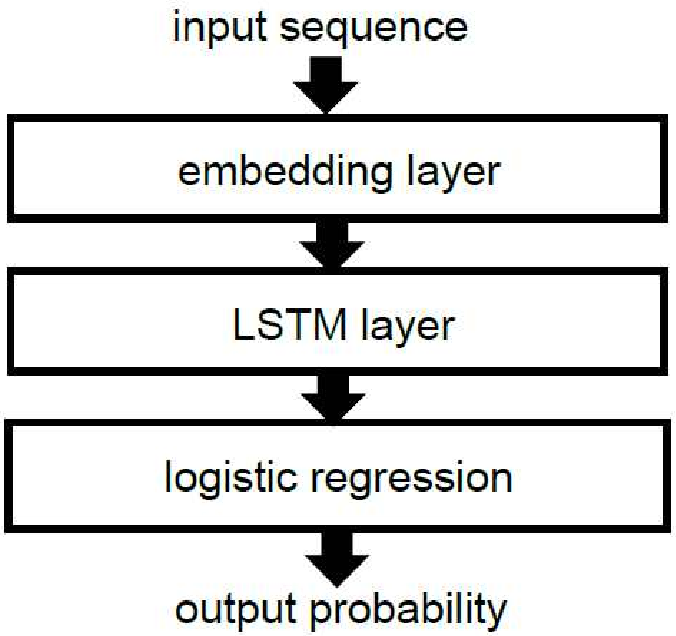
DGA 및 정상 도메인 샘플 정제



6. 딥러닝을 이용한 악성 도메인 탐지 기법 [2]

Long Short-Term Memory




input sequence

embedding layer

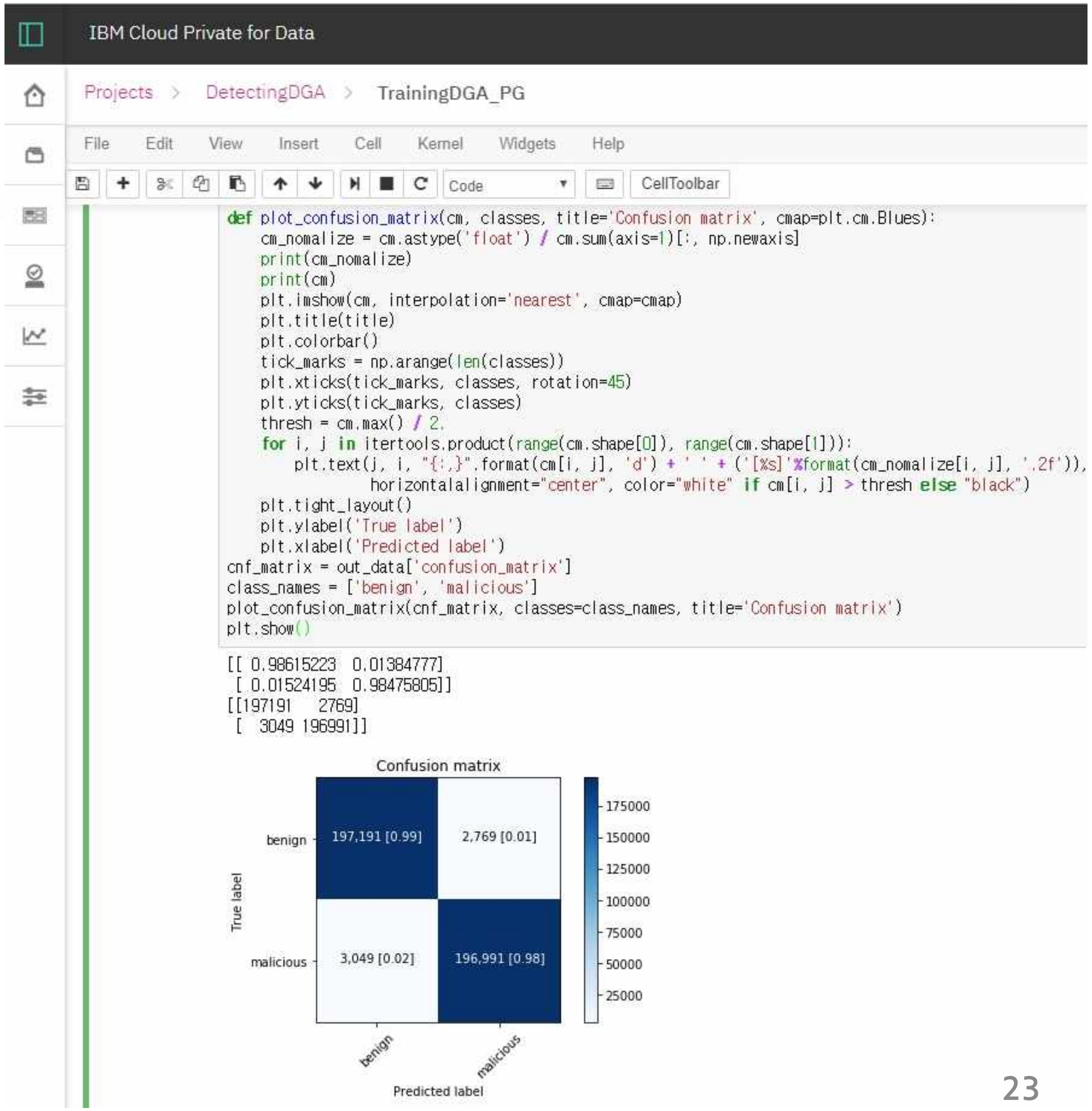
LSTM layer

logistic regression

output probability

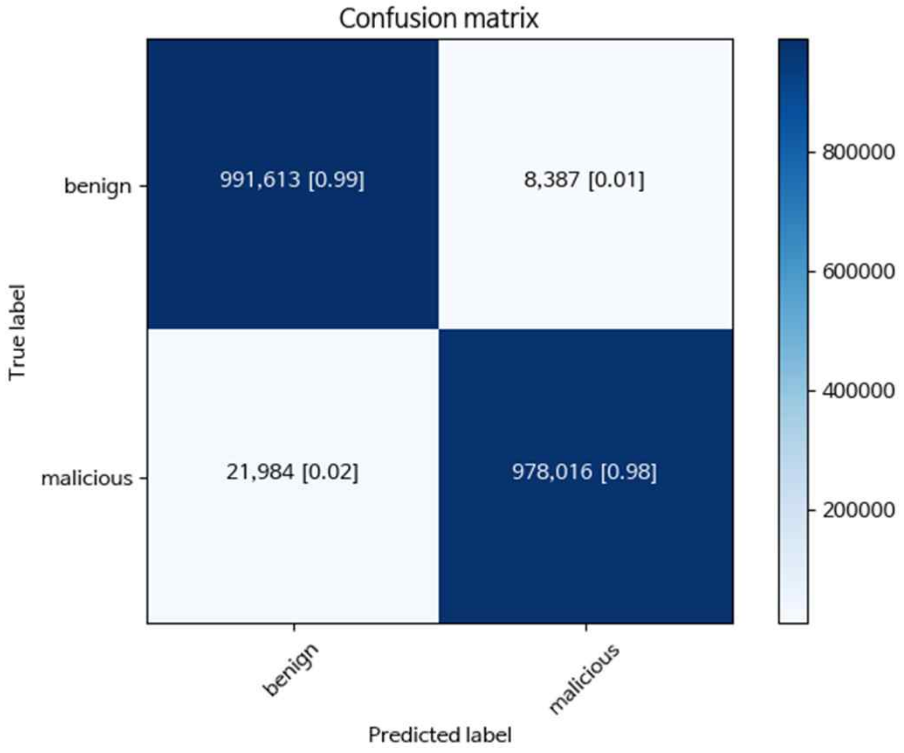
6. 딥러닝을 이용한 악성 도메인 탐지 기법 [2]

DGA 및 정상 도메인 딥러닝 학습

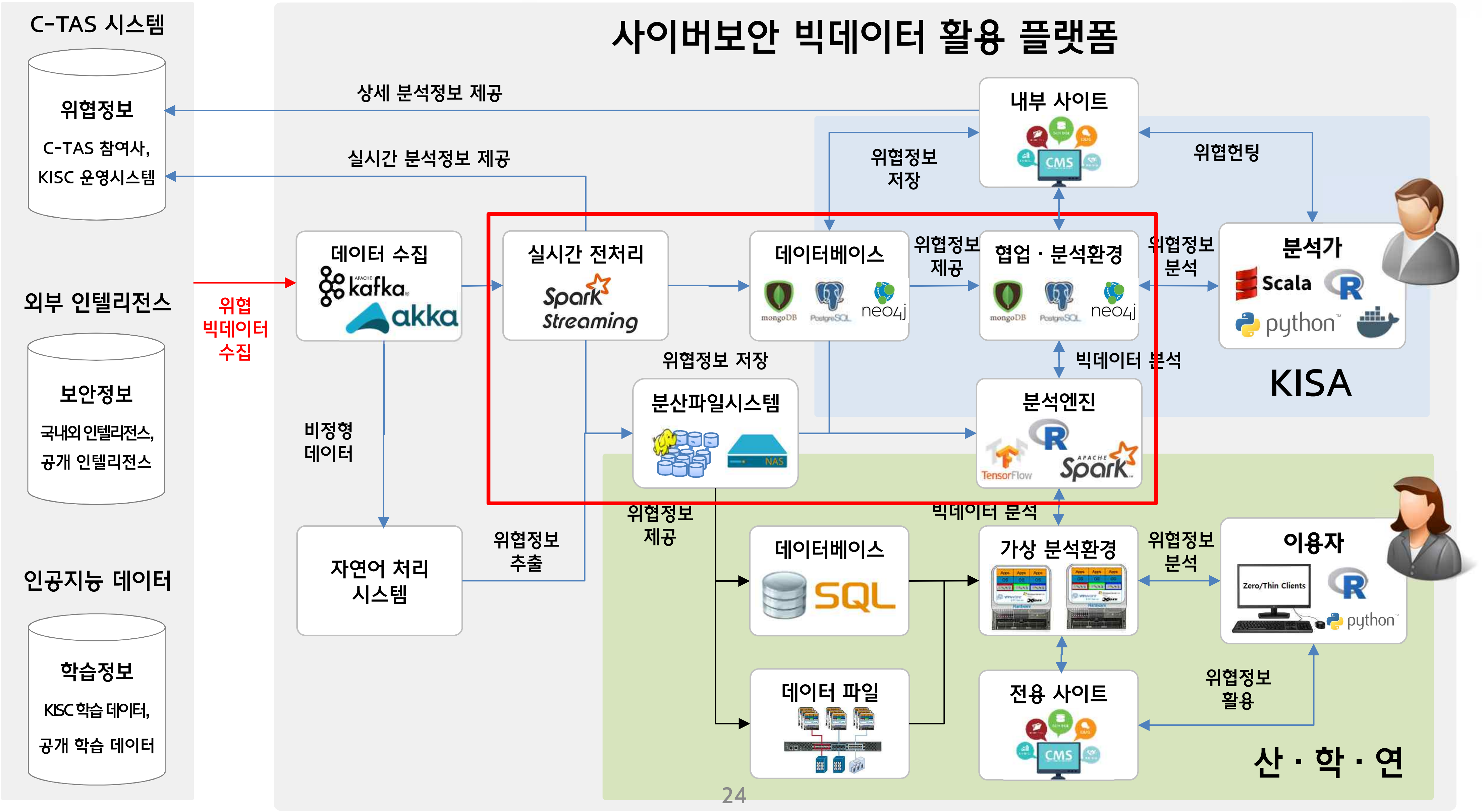


```
def plot_confusion_matrix(cm, classes, title='Confusion matrix', cmap=plt.cm.Blues):
    cm_nomalize = cm.astype('float') / cm.sum(axis=1)[:, np.newaxis]
    print(cm_nomalize)
    print(cm)
    plt.imshow(cm, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
    tick_marks = np.arange(len(classes))
    plt.xticks(tick_marks, classes, rotation=45)
    plt.yticks(tick_marks, classes)
    thresh = cm.max() / 2.
    for i, j in itertools.product(range(cm.shape[0]), range(cm.shape[1])):
        plt.text(j, i, "{:,}".format(cm[i, j], 'd') + ' ' + ('[%s]'%format(cm_nomalize[i, j], '.2f')),
                 horizontalalignment="center", color="white" if cm[i, j] > thresh else "black")
    plt.tight_layout()
    plt.ylabel('True label')
    plt.xlabel('Predicted label')
cnf_matrix = out_data['confusion_matrix']
class_names = ['benign', 'malicious']
plot_confusion_matrix(cnf_matrix, classes=class_names, title='Confusion matrix')
plt.show()
```

```
[[ 0.98615223  0.01384777]
 [ 0.01524195  0.98475805]]
[[197191   2769]
 [  3049 196991]]
```



6. 딥러닝을 이용한 악성 도메인 탐지 기법 [2]



C-TAS 시스템

위협정보

C-TAS 참여사,
KISC 운영시스템

외부 인텔리전스

보안정보

국내외 인텔리전스,
공개 인텔리전스

인공지능 데이터

학습정보

KISC 학습 데이터,
공개 학습 데이터

위협
빅데이터
수집

사이버보안 빅데이터 활용 플랫폼

상세 분석정보 제공

실시간 분석정보 제공

데이터 수집

kafka

akka

실시간 전처리

Spark Streaming

데이터베이스

mongoDB PostgreSQL neo4j

위협정보
제공

협업 · 분석환경

mongoDB PostgreSQL neo4j

위협정보
분석

분석가

Scala R python

내부 사이트

CMS

위협정보
저장

위협헌팅

빅데이터 분석

KISA

비정형
데이터

위협정보 저장

분산파일시스템

NAS

분석엔진

R TensorFlow Spark

자연어 처리
시스템

위협정보
추출

위협정보
제공

빅데이터 분석

데이터베이스

SQL

가상 분석환경

위협정보
분석

이용자

Zero/Thin Clients

R python

데이터 파일

전용 사이트

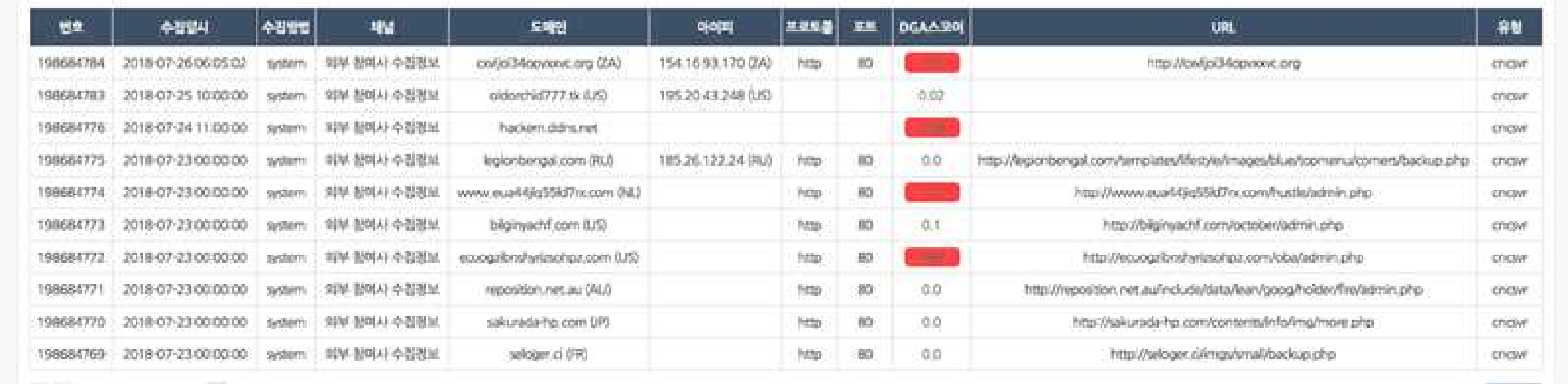
CMS

위협정보
활용

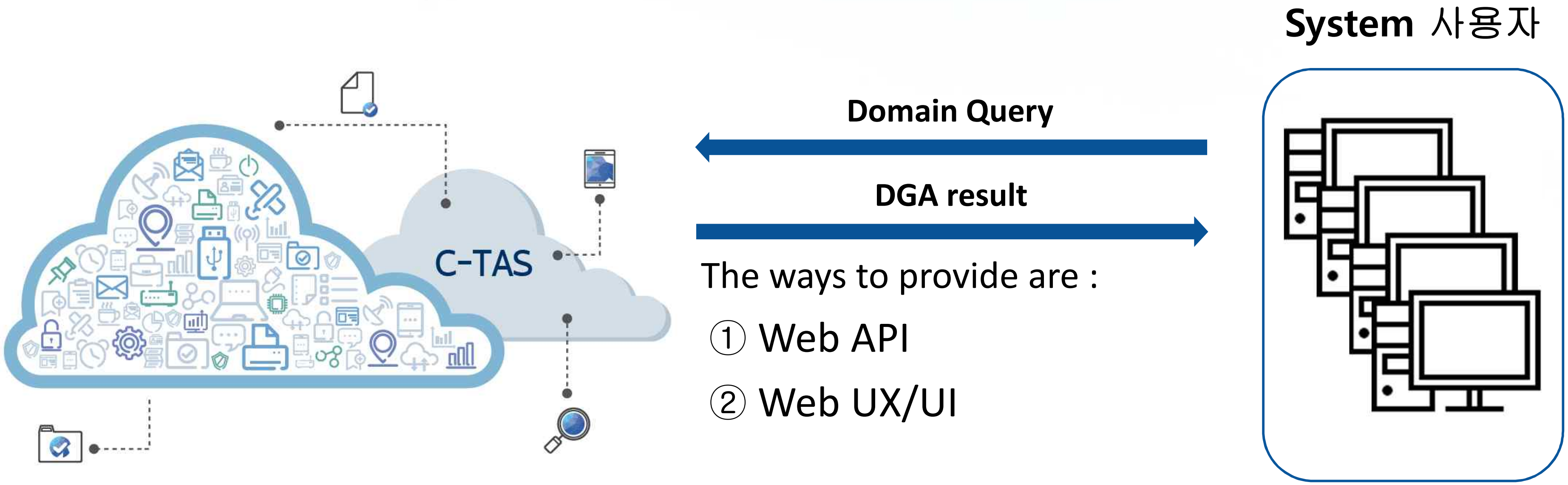
산 · 학 · 연

6. 딥러닝을 이용한 악성 도메인 탐지 기법 [2]

C-TAS를 통한 도메인 악성여부 확인



번호	수집일시	수집방법	채널	도메인	아이피	프로토콜	포트	DGA스코어	URL	유형
198684784	2018-07-26 06:05:02	system	외부 참여사 수집정보	oxvljoi34opvxxvc.org (ZA)	154.16.93.170 (ZA)	http	80	[illegible]	http://oxvljoi34opvxxvc.org	cncsvr
198684783	2018-07-25 10:00:00	system	외부 참여사 수집정보	oldorchid777.tk (US)	195.20.43.248 (US)			0.02		cncsvr
198684776	2018-07-24 11:00:00	system	외부 참여사 수집정보	hackem.ddns.net				[illegible]		cncsvr
198684775	2018-07-23 00:00:00	system	외부 참여사 수집정보	legionbengal.com (RU)	185.26.122.24 (RU)	http	80	0.0	http://legionbengal.com/templates/lifestyle/images/blue/topmenu/corners/backup.php	cncsvr
198684774	2018-07-23 00:00:00	system	외부 참여사 수집정보	www.eua44jiq55kd7rx.com (NL)		http	80	[illegible]	http://www.eua44jiq55kd7rx.com/hustle/admin.php	cncsvr
198684773	2018-07-23 00:00:00	system	외부 참여사 수집정보	bilginyachf.com (US)		http	80	0.1	http://bilginyachf.com/october/admin.php	cncsvr
198684772	2018-07-23 00:00:00	system	외부 참여사 수집정보	ecuogzibnshyrizsohpz.com (US)		http	80	[illegible]	http://ecuogzibnshyrizsohpz.com/oba/admin.php	cncsvr
198684771	2018-07-23 00:00:00	system	외부 참여사 수집정보	reposition.net.au (AU)		http	80	0.0	http://reposition.net.au/include/data/lean/goog/holder/fire/admin.php	cncsvr
198684770	2018-07-23 00:00:00	system	외부 참여사 수집정보	sakurada-hp.com (JP)		http	80	0.0	http://sakurada-hp.com/contents/info/img/more.php	cncsvr
198684769	2018-07-23 00:00:00	system	외부 참여사 수집정보	seloger.ci (FR)		http	80	0.0	http://seloger.ci/imgs/small/backup.php	cncsvr



The ways to provide are :

① Web API

② Web UX/UI

7. 그래프 분석을 이용한 악성 도메인 탐지 기법 [9]

NOW
MATTERS

#RSAC

RSAConference2018

San Francisco | April 16 – 20 | Moscone Center

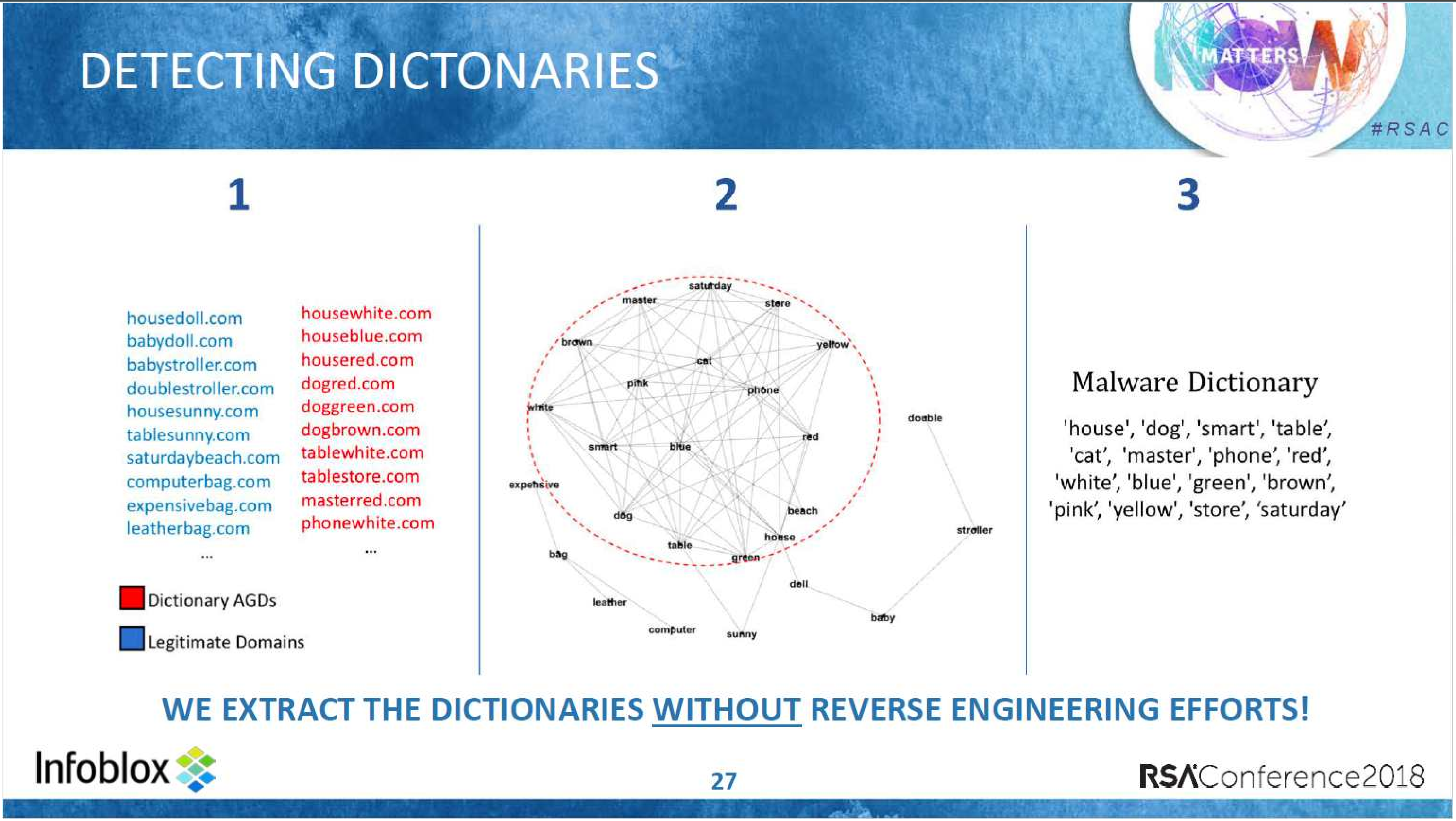
SESSION ID: MLN-F01

FIGHTING MALWARE WITH GRAPH ANALYTICS: AN END-TO-END CASE STUDY

Mayana Pereira

Data Scientist
Infoblox Inc.

7. 그래프 분석을 이용한 악성 도메인 탐지 기법 [9]



DETECTING DICTONARIES

1

- housedoll.com
- babydoll.com
- babystroller.com
- doublestroller.com
- housesunny.com
- tablesunny.com
- saturdaybeach.com
- computerbag.com
- expensivebag.com
- leatherbag.com
- ...

- housewhite.com
- houseblue.com
- housered.com
- dogred.com
- doggreen.com
- dogbrown.com
- tablewhite.com
- tablestore.com
- masterred.com
- phonewhite.com
- ...

🟥 Dictionary AGDs

🟦 Legitimate Domains

2

3

Malware Dictionary

'house', 'dog', 'smart', 'table', 'cat', 'master', 'phone', 'red', 'white', 'blue', 'green', 'brown', 'pink', 'yellow', 'store', 'saturday'

WE EXTRACT THE DICTIONARIES WITHOUT REVERSE ENGINEERING EFFORTS!

8. 참고문헌

References

[1] Jay Jacobs, "Security Data Science: From Theory to Reality", RSA Conference 2015
[2] Hyrum Anderson, "DeepDGA: Adversarially-Tuned Domain Generation and Detection", AISec, 2016
[3] Rod Rasmussen, "DNS is NOT Boring! Using DNS to Expose and Thwart Attacks", FIRST Conference, 2017
[4] AkamAI Research, "A DEATH MATCH OF DOMAIN GENERATION ALGORITHMS", 2017
[5] AkamAI Research, "SPOTLIGHT ON MALWARE DGA COMMUNICATION TECHNIQUE", 2017
[6] DomainTools, "The Distribution of Malicious Domain", 2016
[7] Cylance, "Introduction to Artificial Intelligence for Security Professionals", 2017
[8] Anomali, "Hacker Tactics - Part 1: Domain Generation Algorithms", 2017
[9] Mayana Pereira, "Fighting Malware with Graph Analytics-An End-to-End Case Study", 2018

감사합니다.