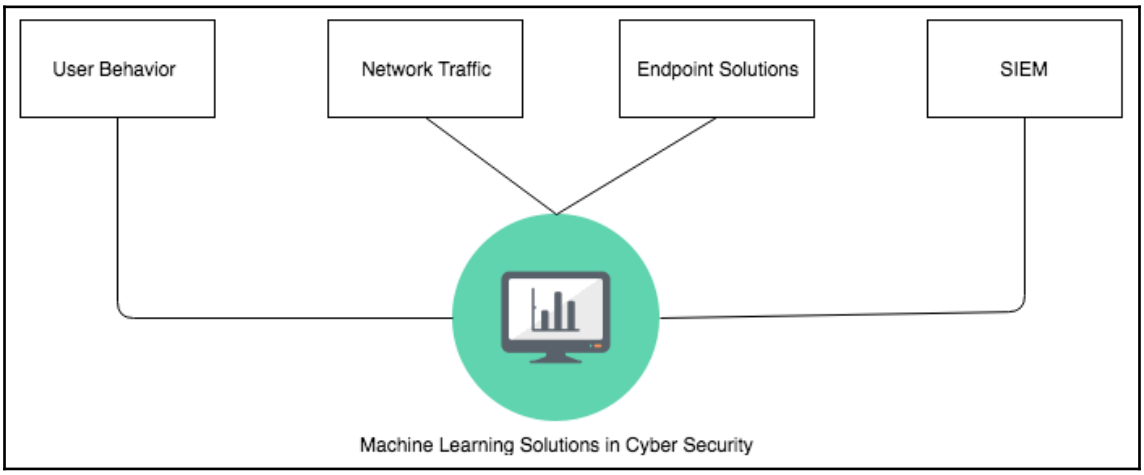
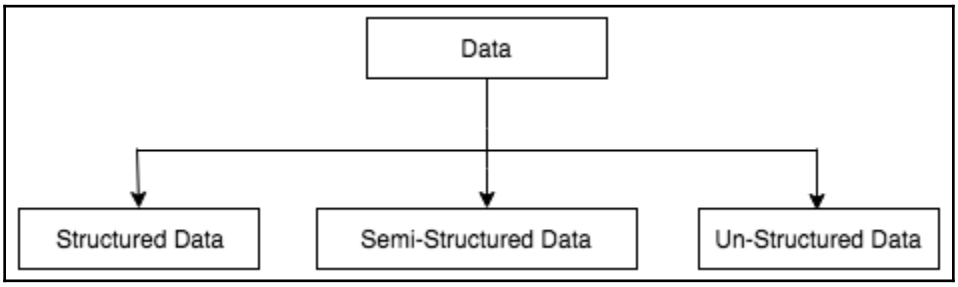
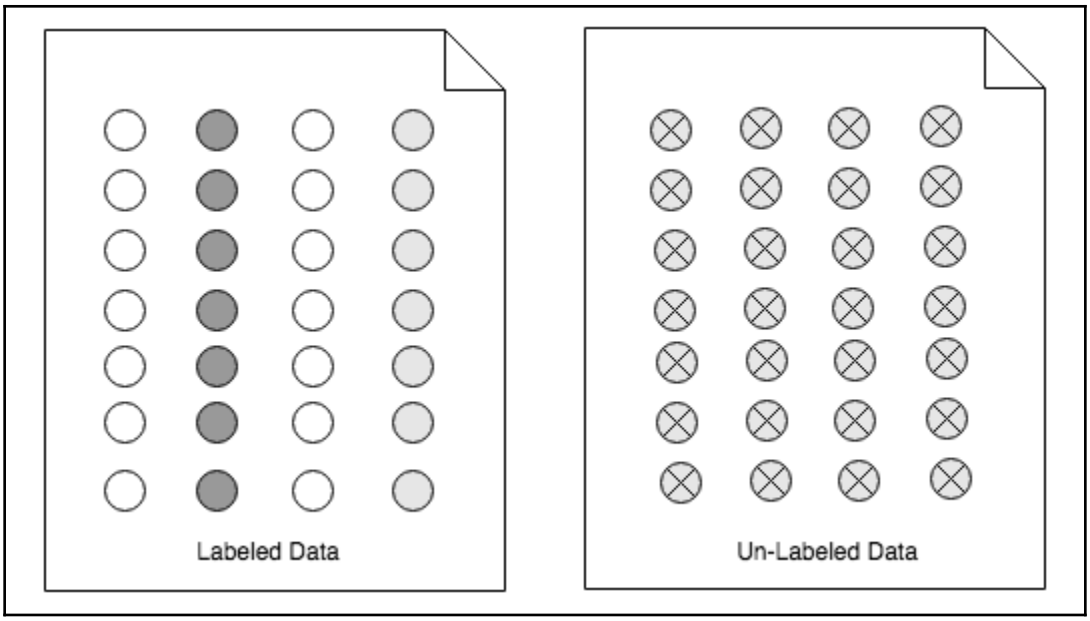
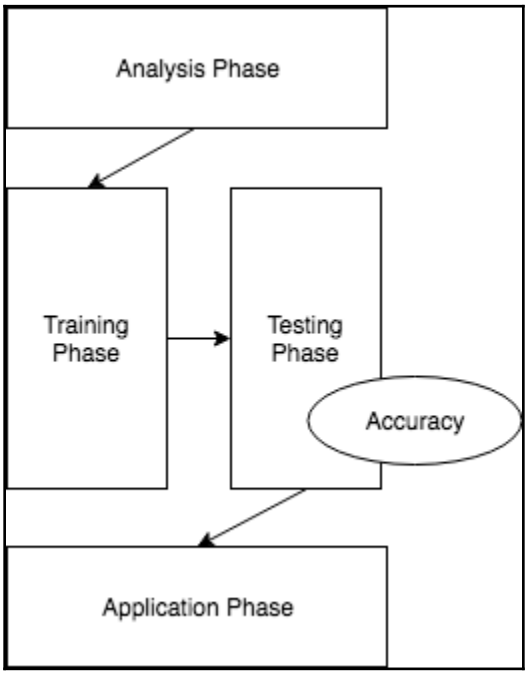
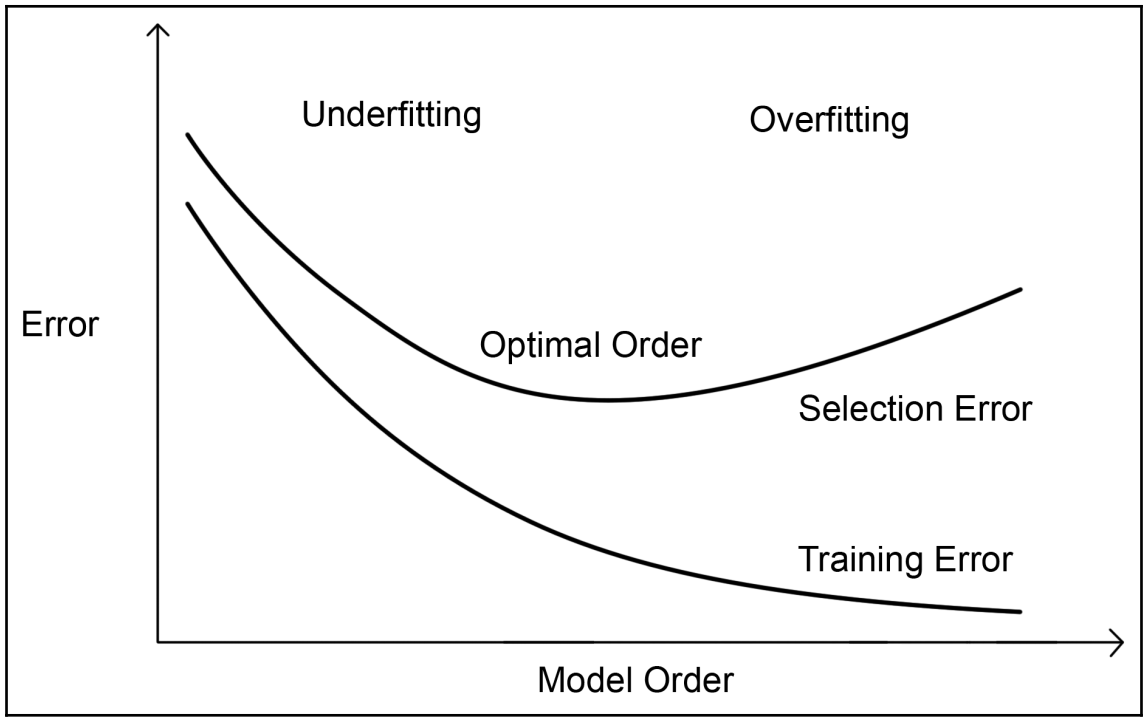
Chapter 1: Basics of Machine Learning in Cybersecurity

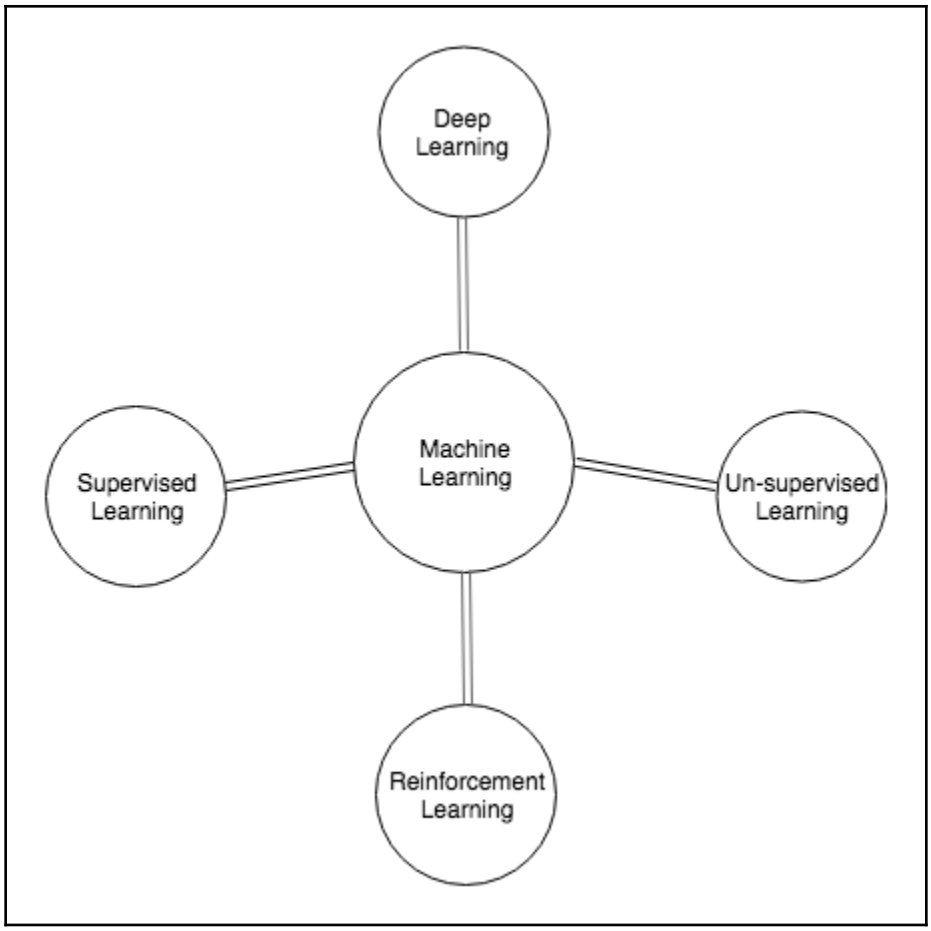


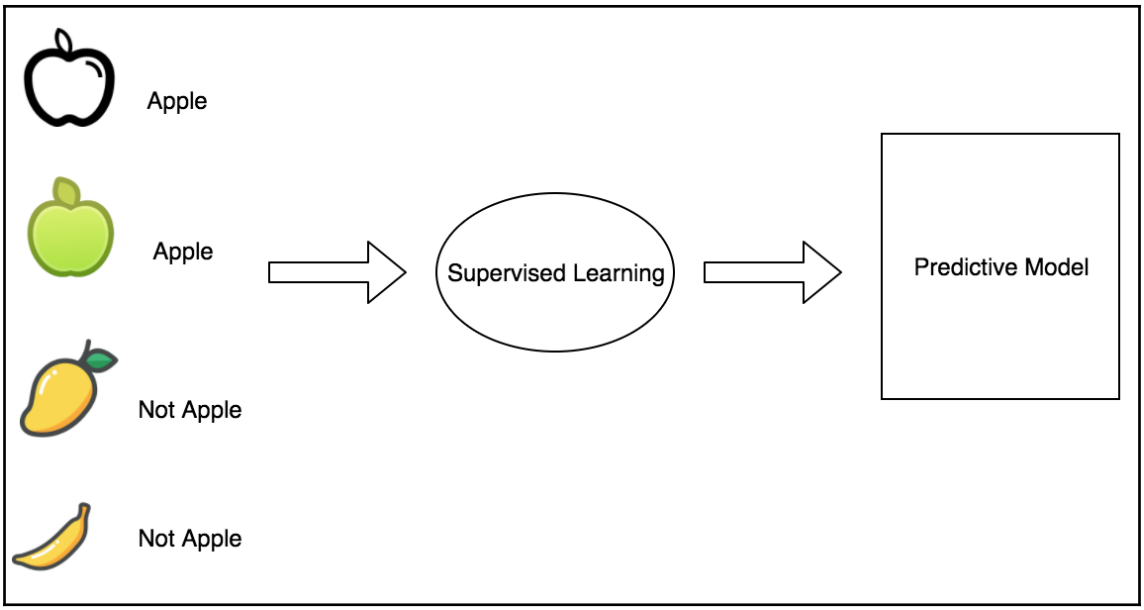


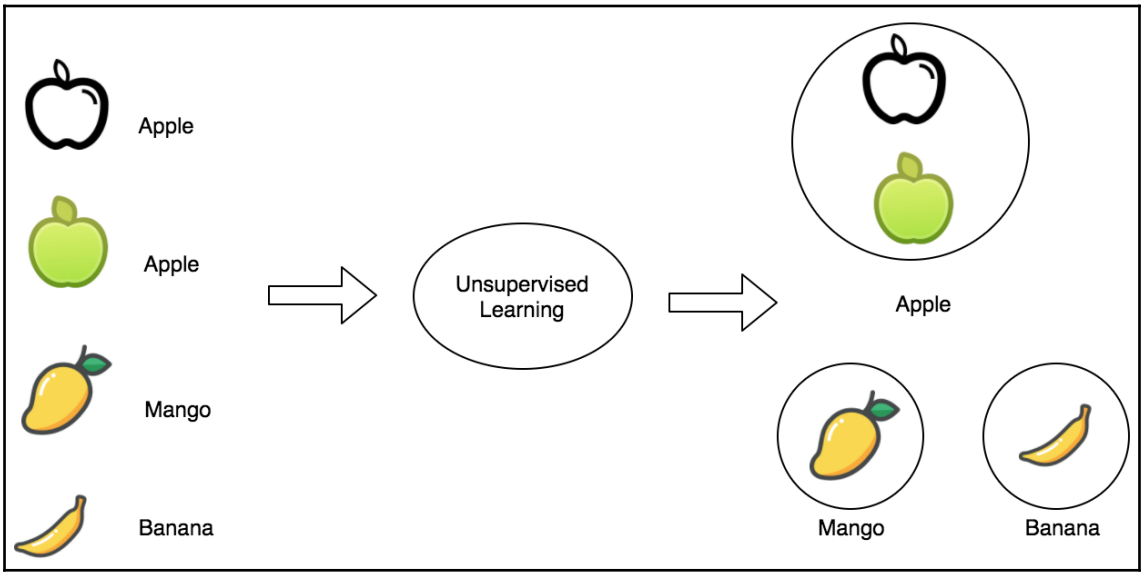

Labeled Data
Un-Labeled Data

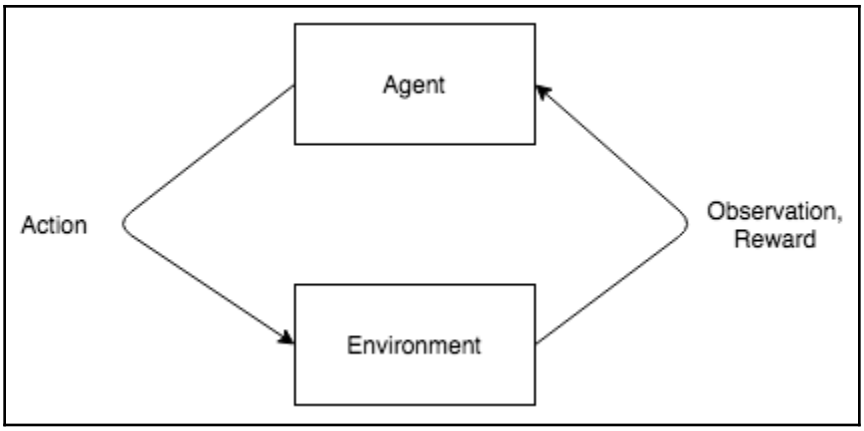

Analysis Phase
Training Phase
Testing Phase
Accuracy
Application Phase

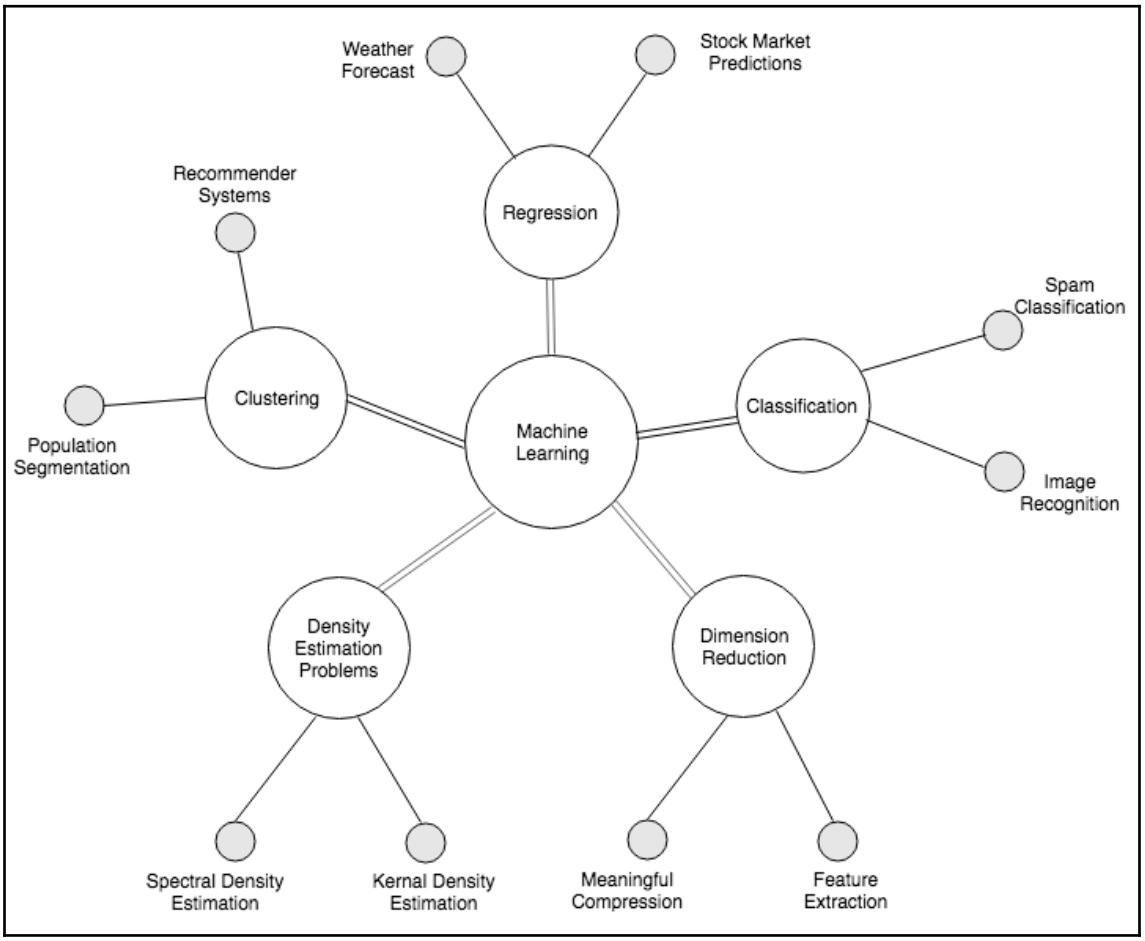

Underfitting
Overfitting
Error
Optimal Order
Selection Error
Training Error
Model Order

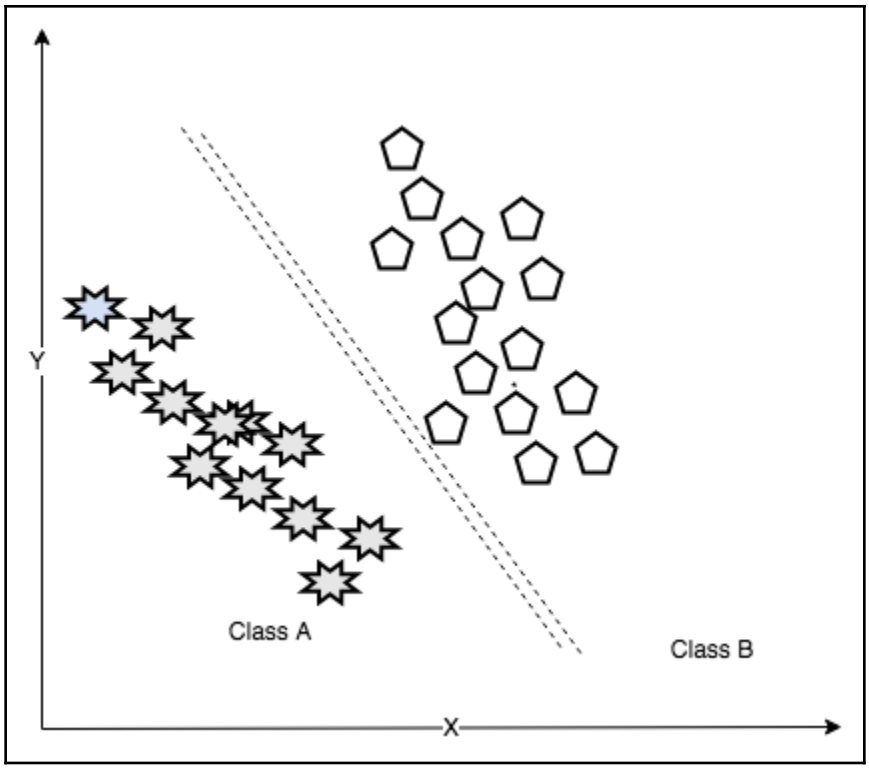

Deep
Learning
Supervised
Learning
Machine
Learning
Un-supervised
Learning
Reinforcement
Learning

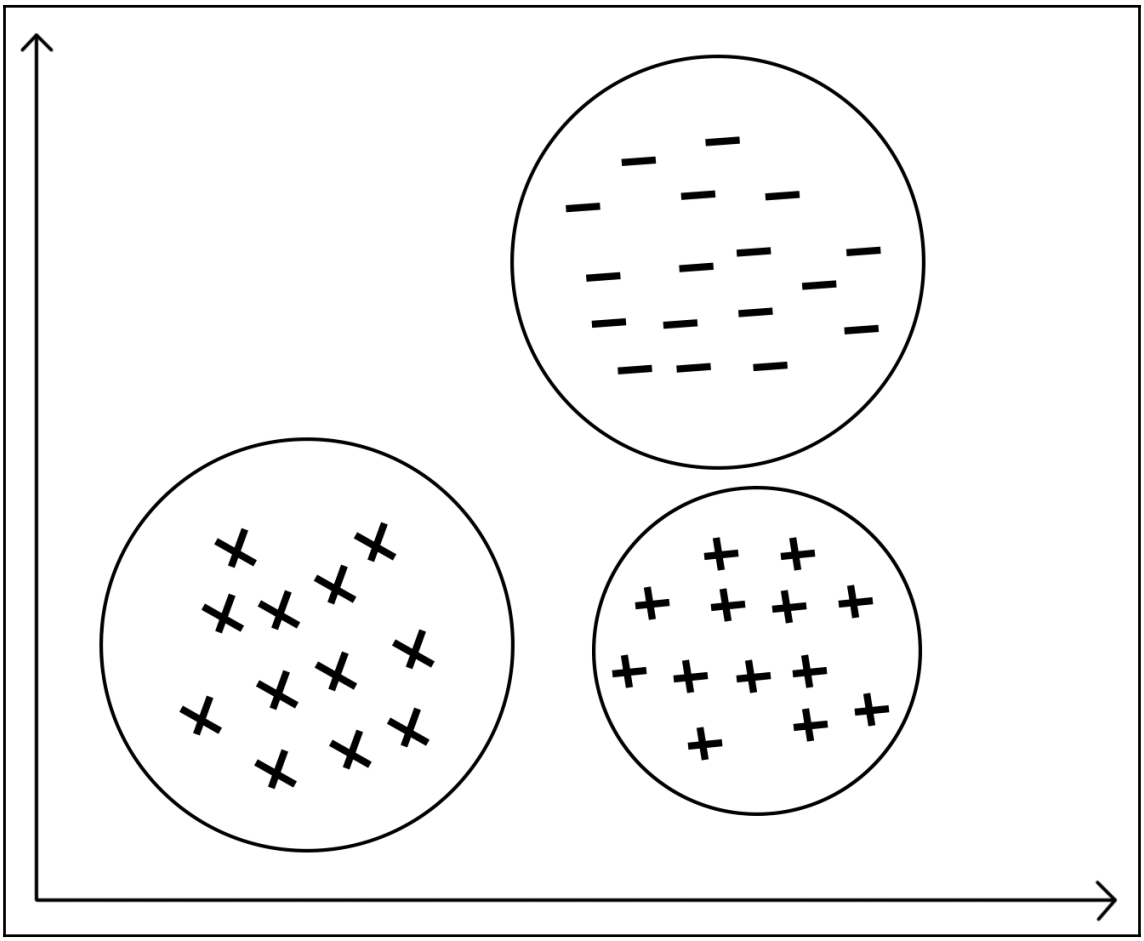

Apple
Apple
Not Apple
Not Apple
Supervised Learning
Predictive Model

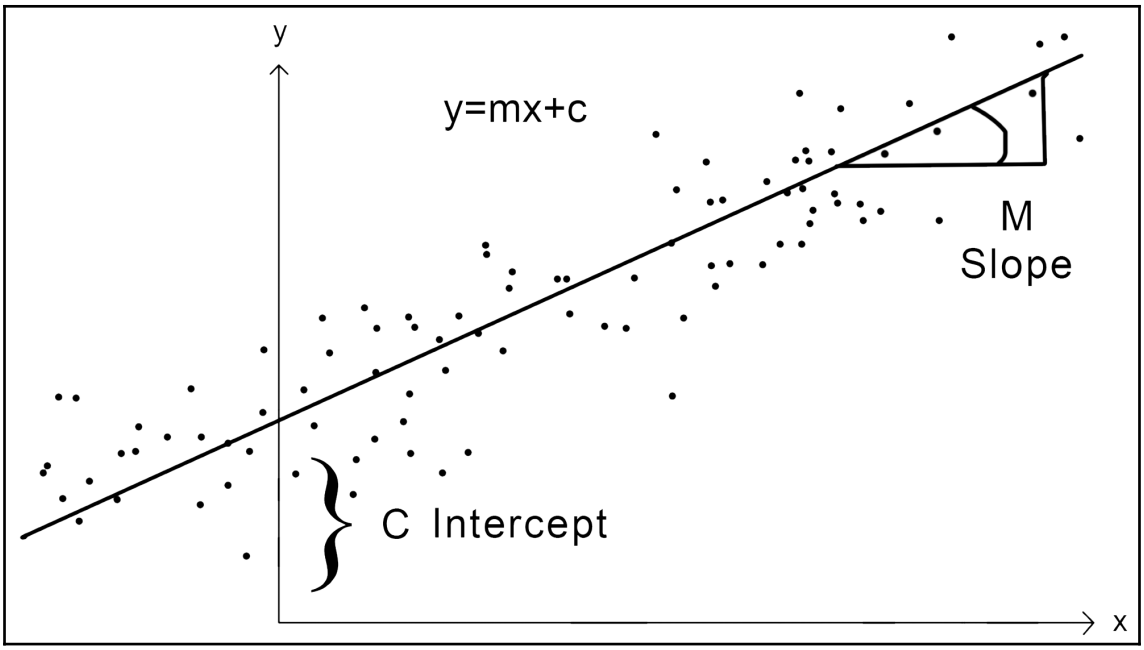

Apple
Apple
Mango
Banana
Unsupervised Learning
Apple
Mango
Banana

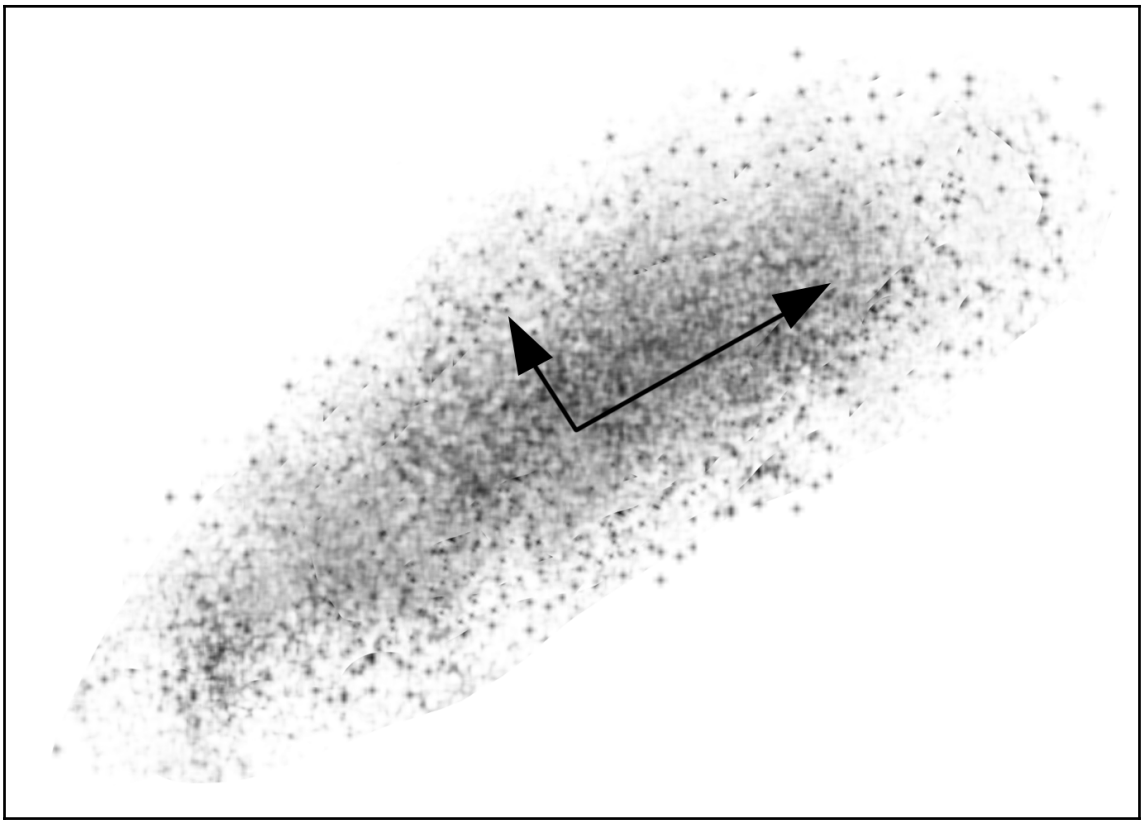

Agent
Action
Observation,
Reward
Environment

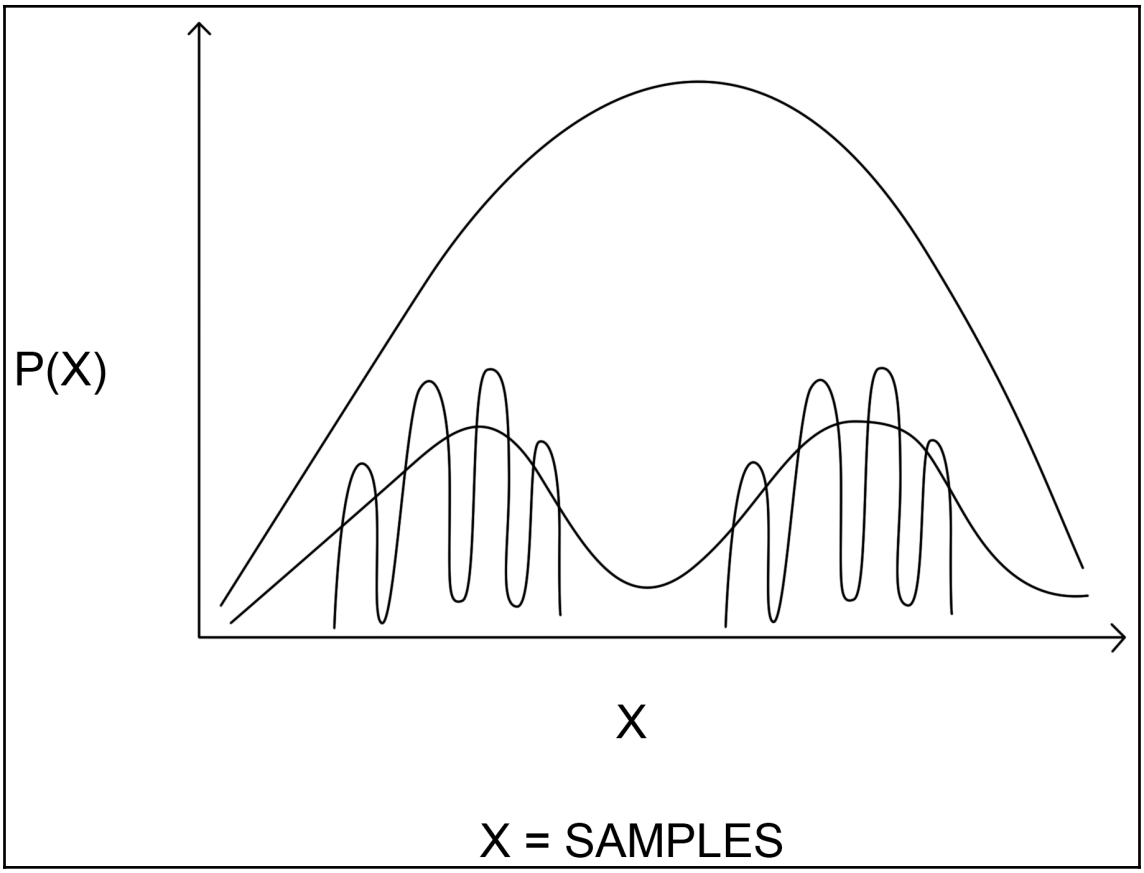

Weather
Forecast
Stock Market
Predictions
Recommender
Systems
Regression
Spam
Classification
Clustering
Machine
Learning
Classification
Population
Segmentation
Image
Recognition
Density
Estimation
Problems
Dimension
Reduction
Spectral Density
Estimation
Kernal Density
Estimation
Meaningful
Compression
Feature
Extraction

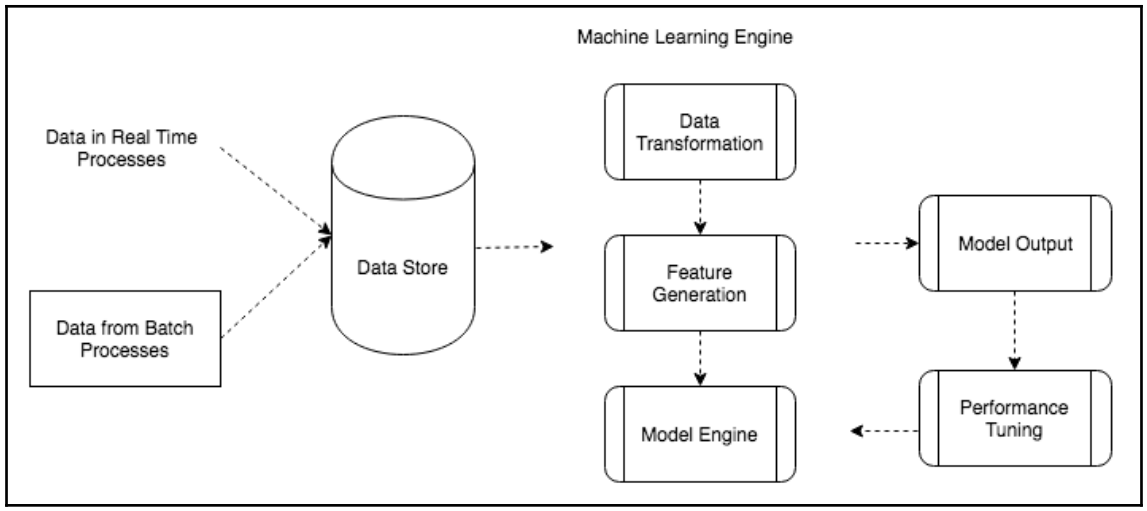

Y
Class A
Class B
X

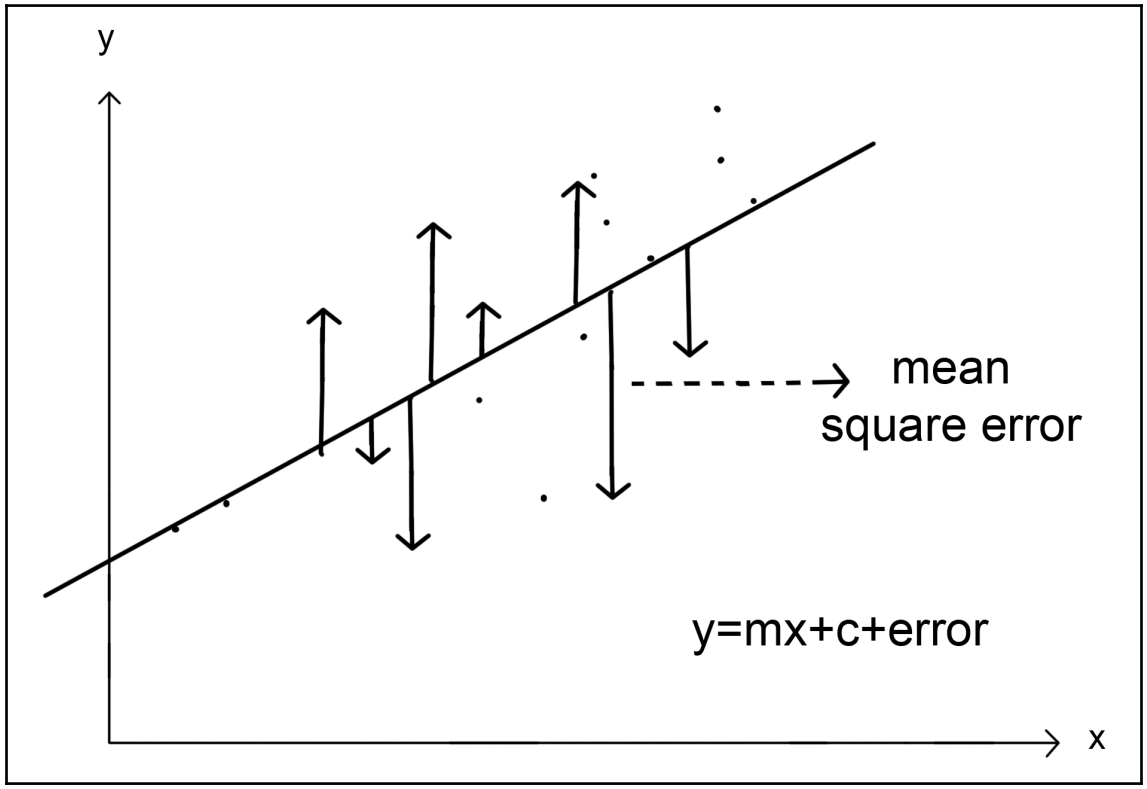


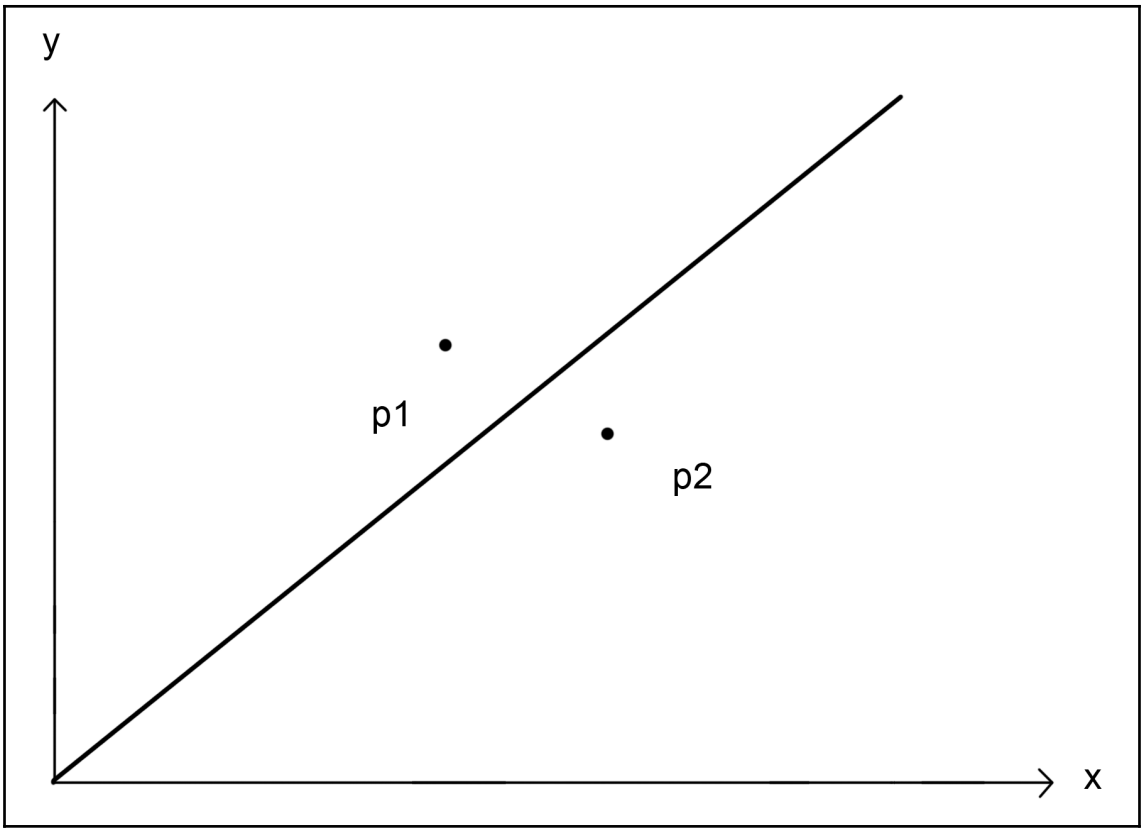

y
y=mx+c
M
Slope
C Intercept
x

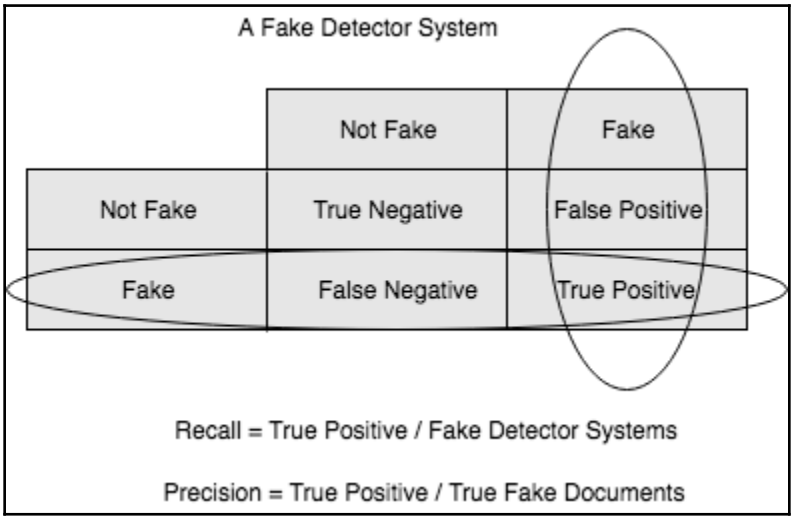


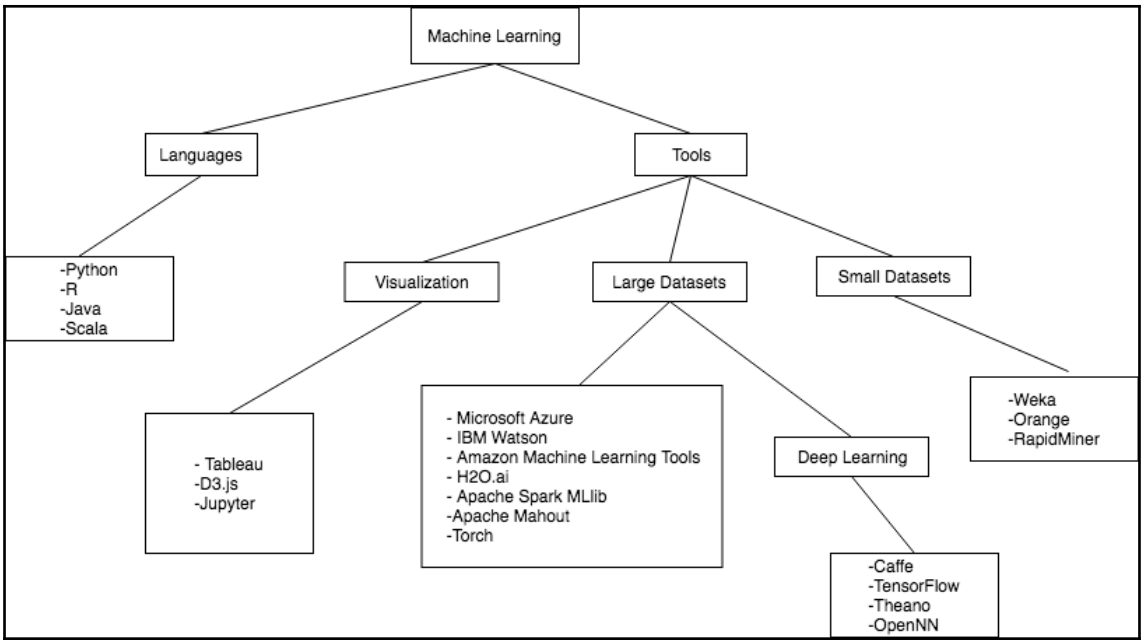

P(X)
X
X = SAMPLES

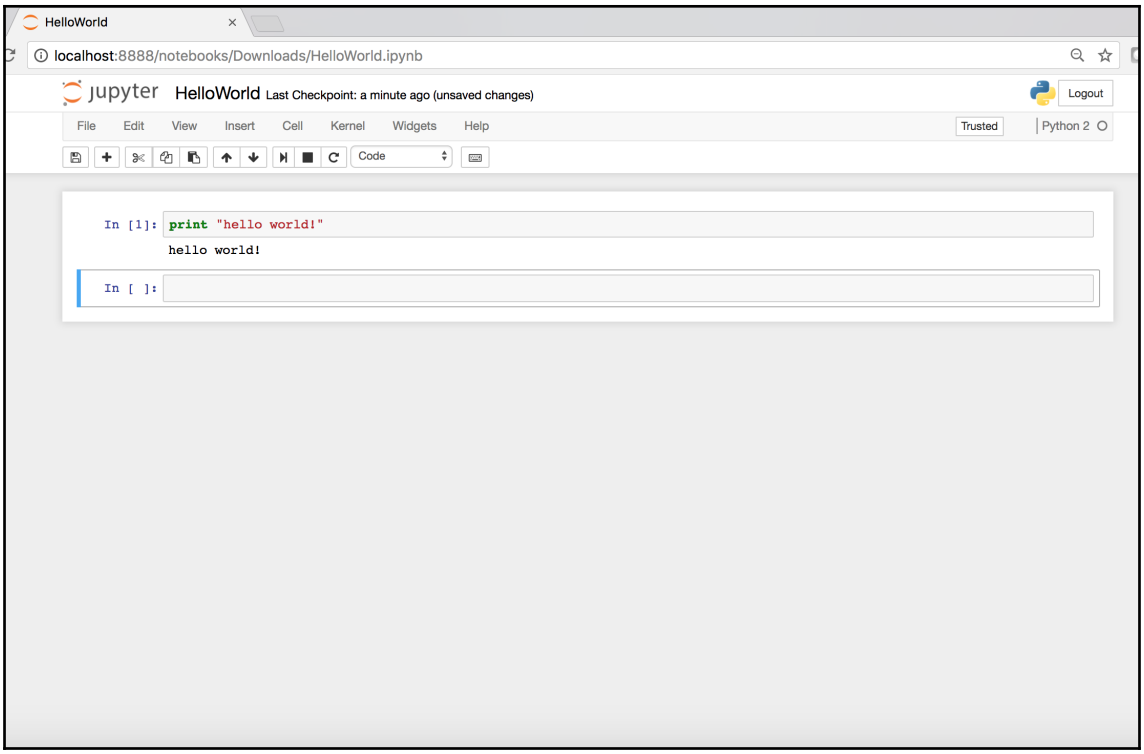

Machine Learning Engine
Data in Real Time Processes
Data from Batch Processes
Data Store
Data Transformation
Feature Generation
Model Engine
Model Output
Performance Tuning

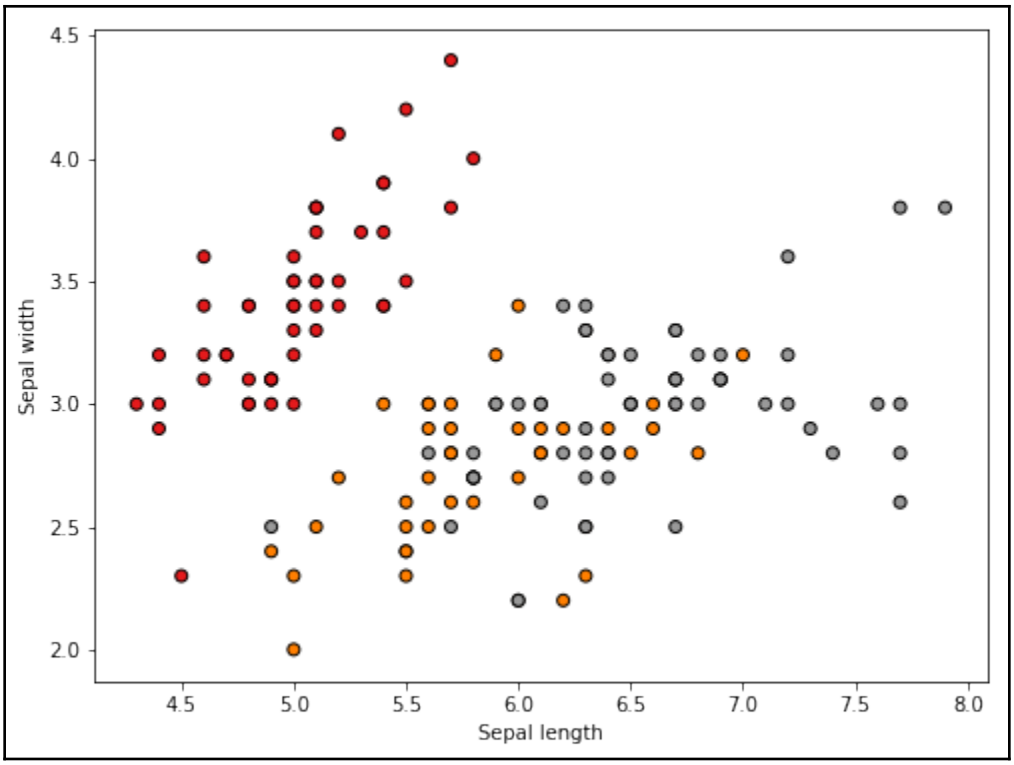

y
mean
square error
y=mx+c+error
x

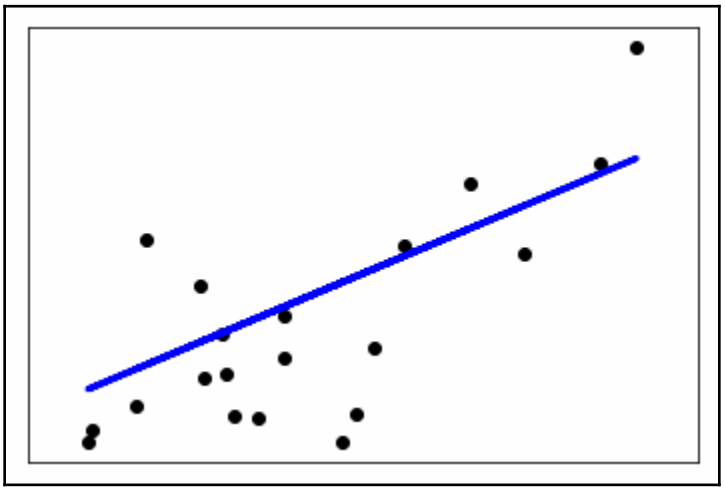

y
p1
p2
x


A Fake Detector System
Not Fake
Fake
Not Fake
True Negative
False Positive
Fake
False Negative
True Positive
Recall = True Positive / Fake Detector Systems
Precision = True Positive / True Fake Documents

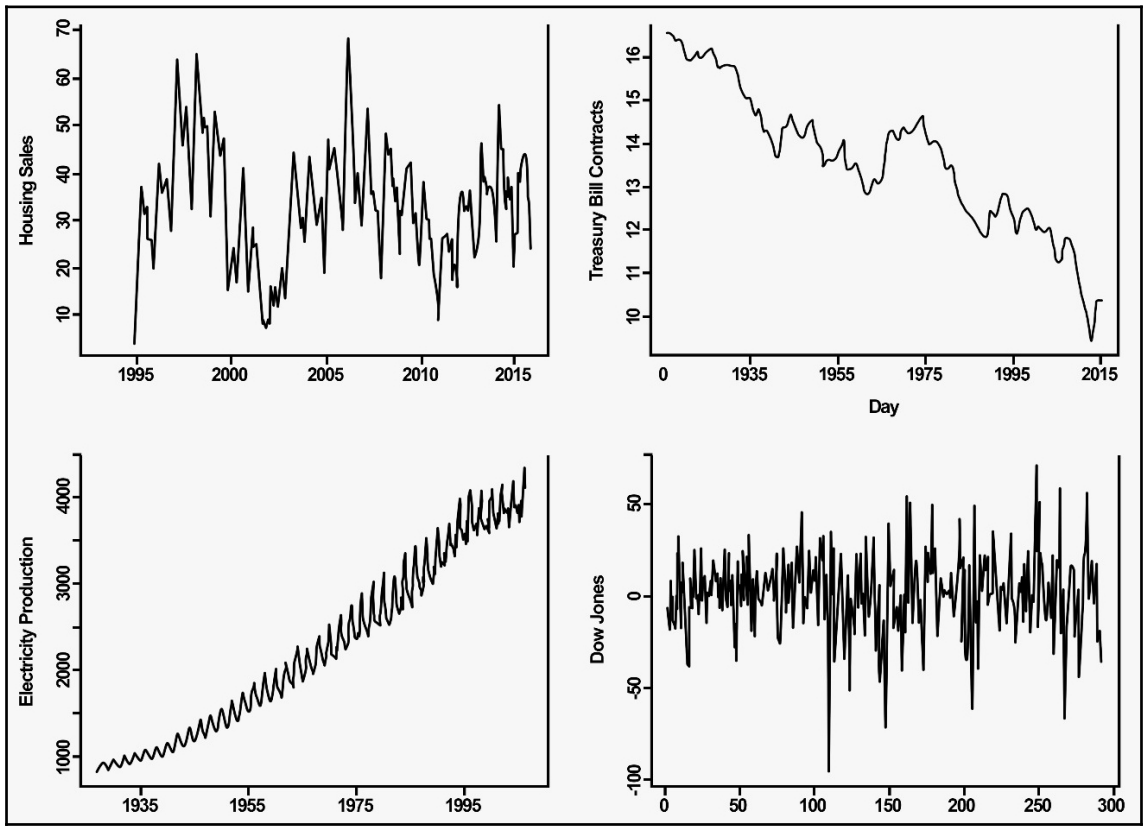

Machine Learning
Languages
Tools
-Python
-R
-Java
-Scala
Visualization
Large Datasets
Small Datasets
- Tableau
-D3.js
-Jupyter
- Microsoft Azure
- IBM Watson
- Amazon Machine Learning Tools
- H2O.ai
- Apache Spark MLlib
-Apache Mahout
-Torch
Deep Learning
-Weka
-Orange
-RapidMiner
-Caffe
-TensorFlow
-Theano
-OpenNN

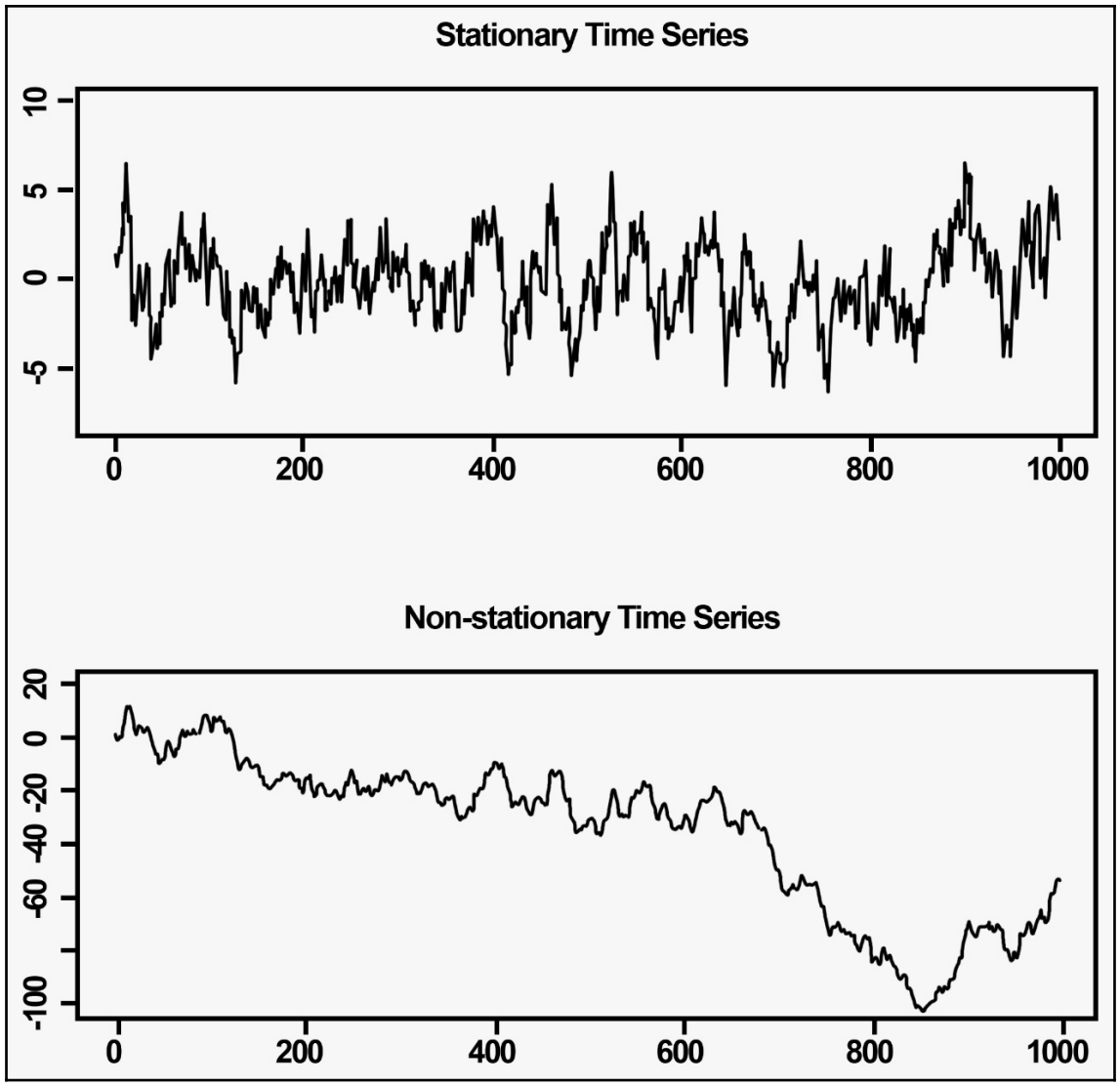

HelloWorld
localhost:8888/notebooks/Downloads/HelloWorld.ipynb
jupyter
HelloWorld
Last Checkpoint: a minute ago (unsaved changes)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 2
Code
In [1]:
print "hello world!"
hello world!
In [ ]:

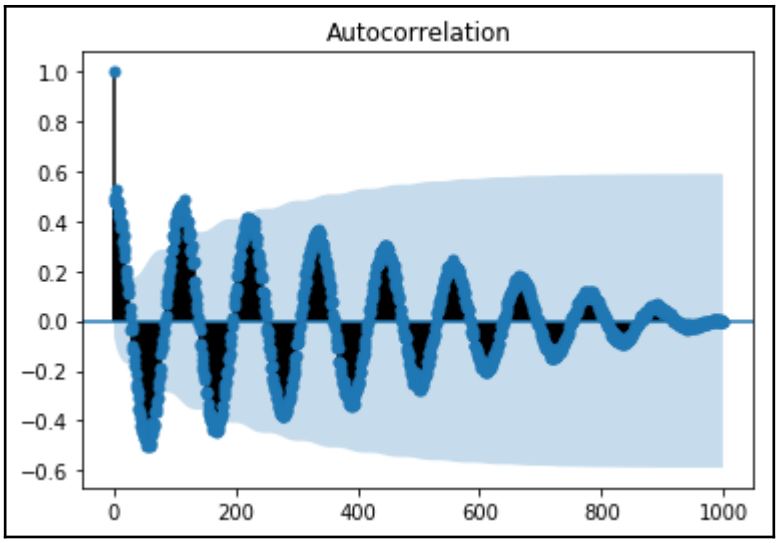

4.5
4.0
3.5
3.0
2.5
2.0
Sepal width
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
Sepal length

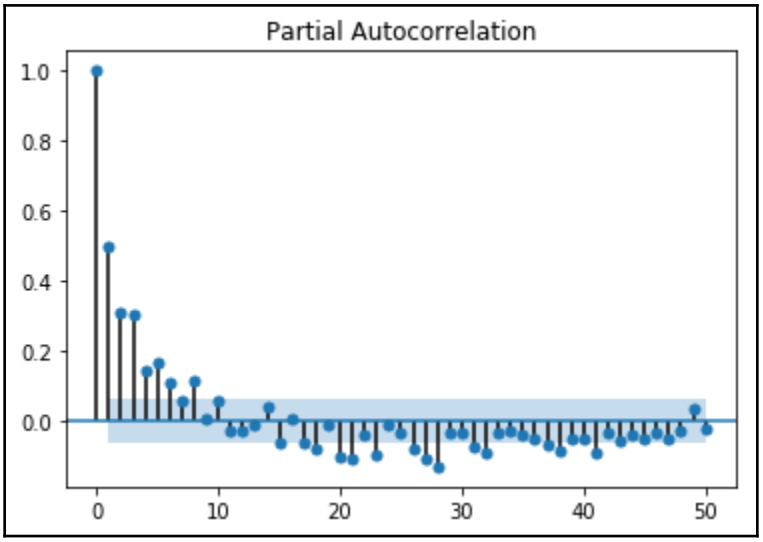


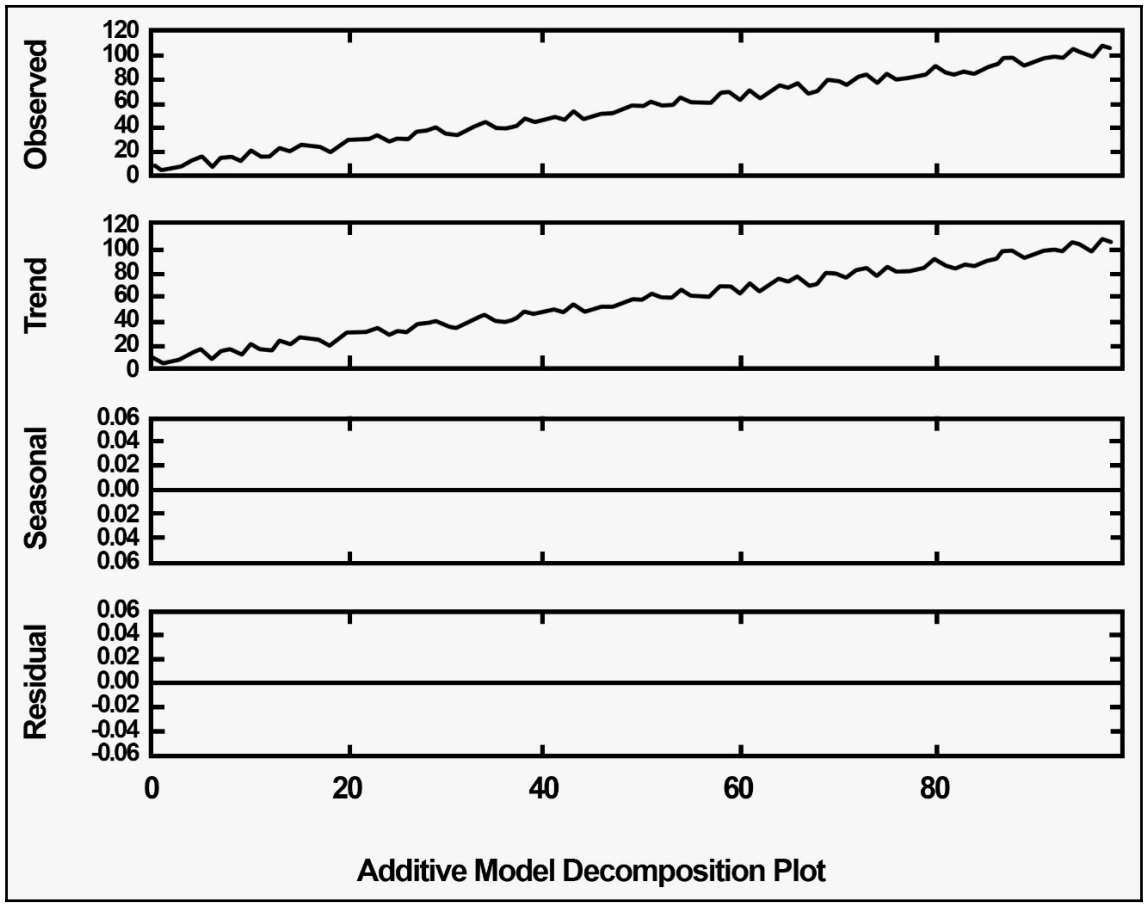
Chapter 2: Time Series Analysis and Ensemble Modeling

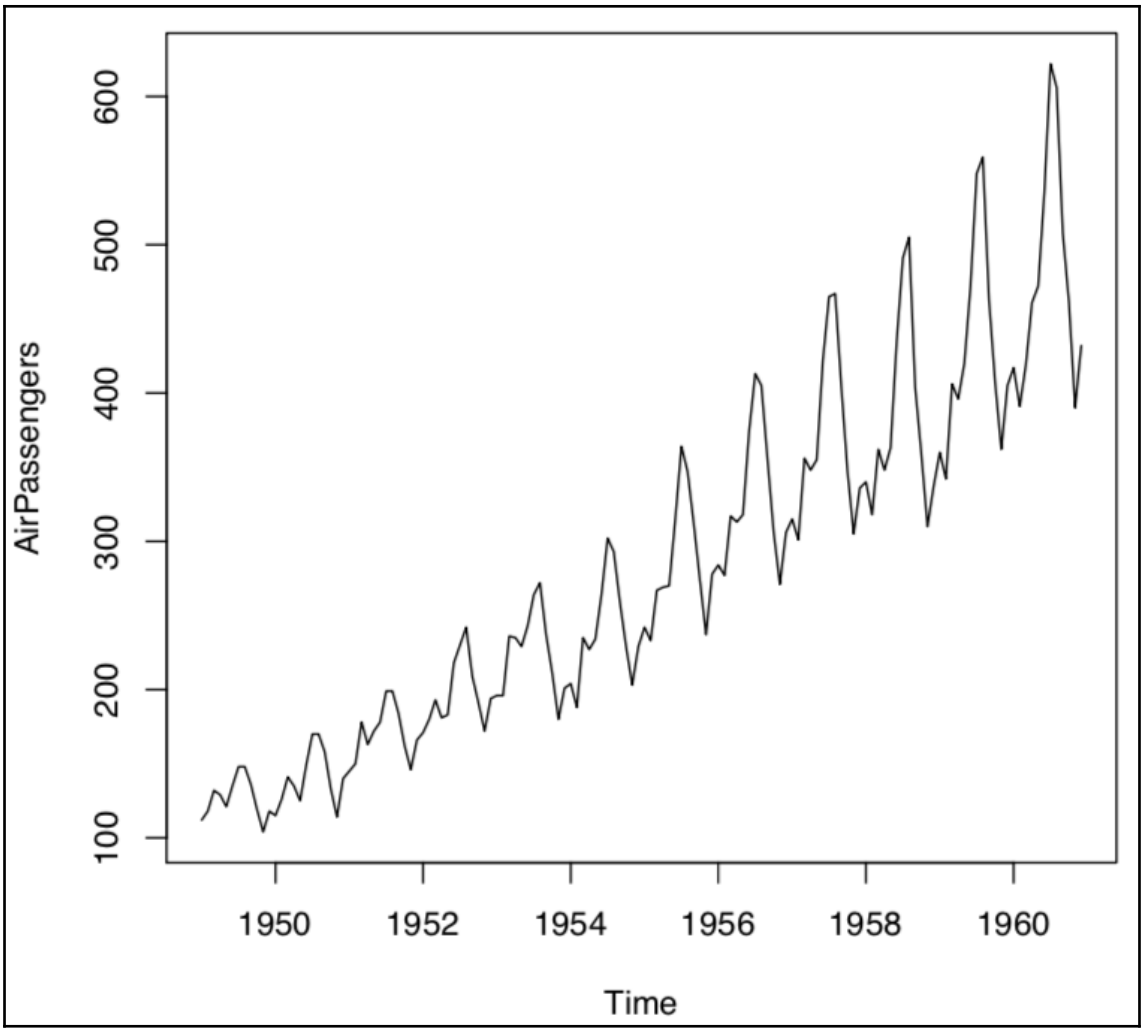


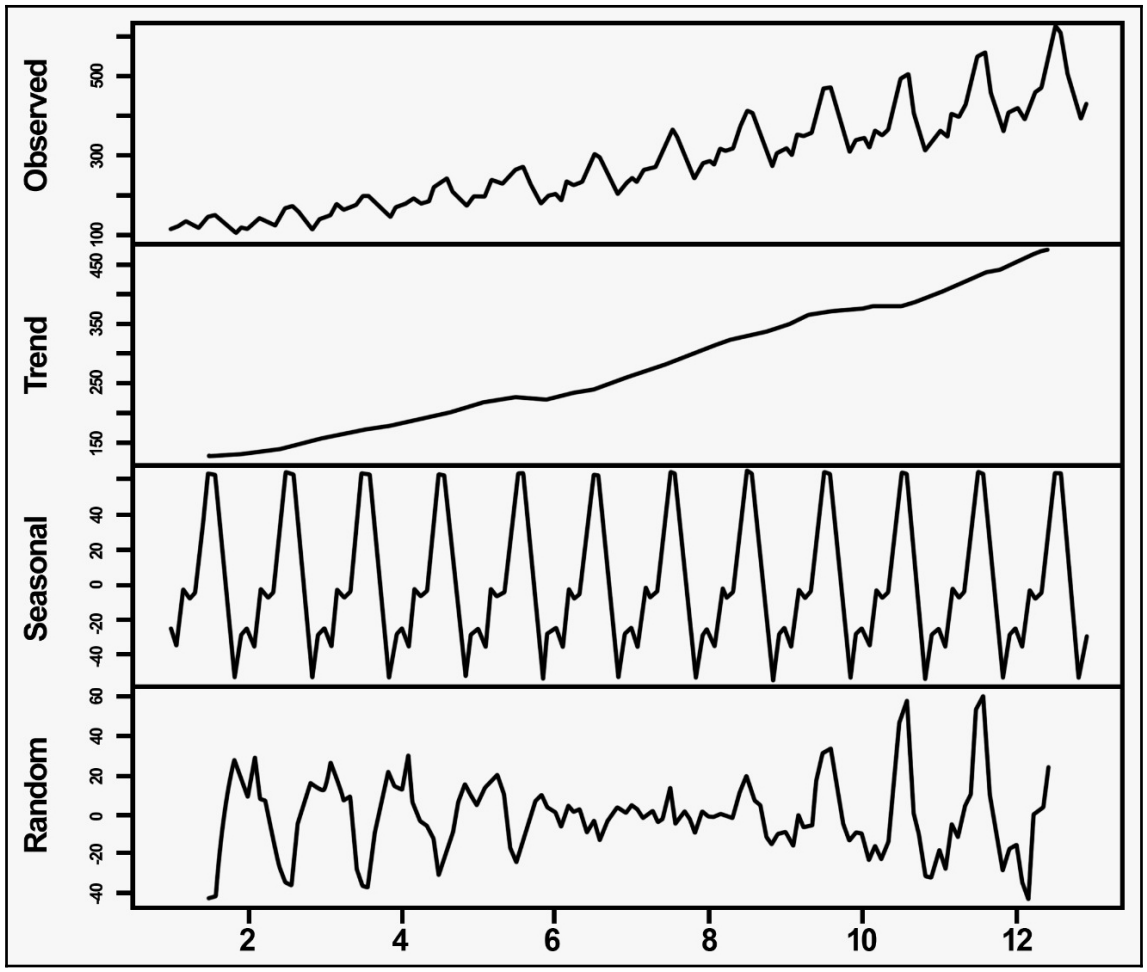

Stationary Time Series
10
5
0
-5
0
200
400
600
800
1000
Non-stationary Time Series
20
0
-20
-40
-60
-100
0
200
400
600
800
1000

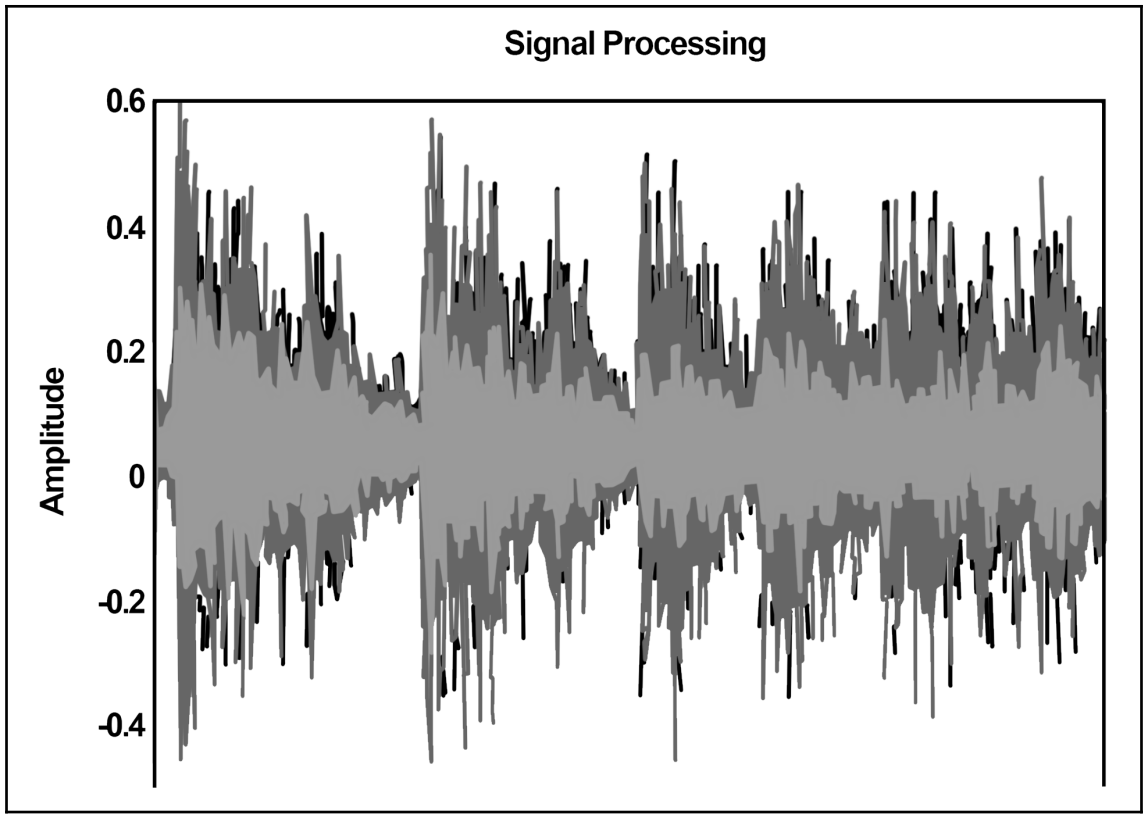

Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
0
200
400
600
800
1000

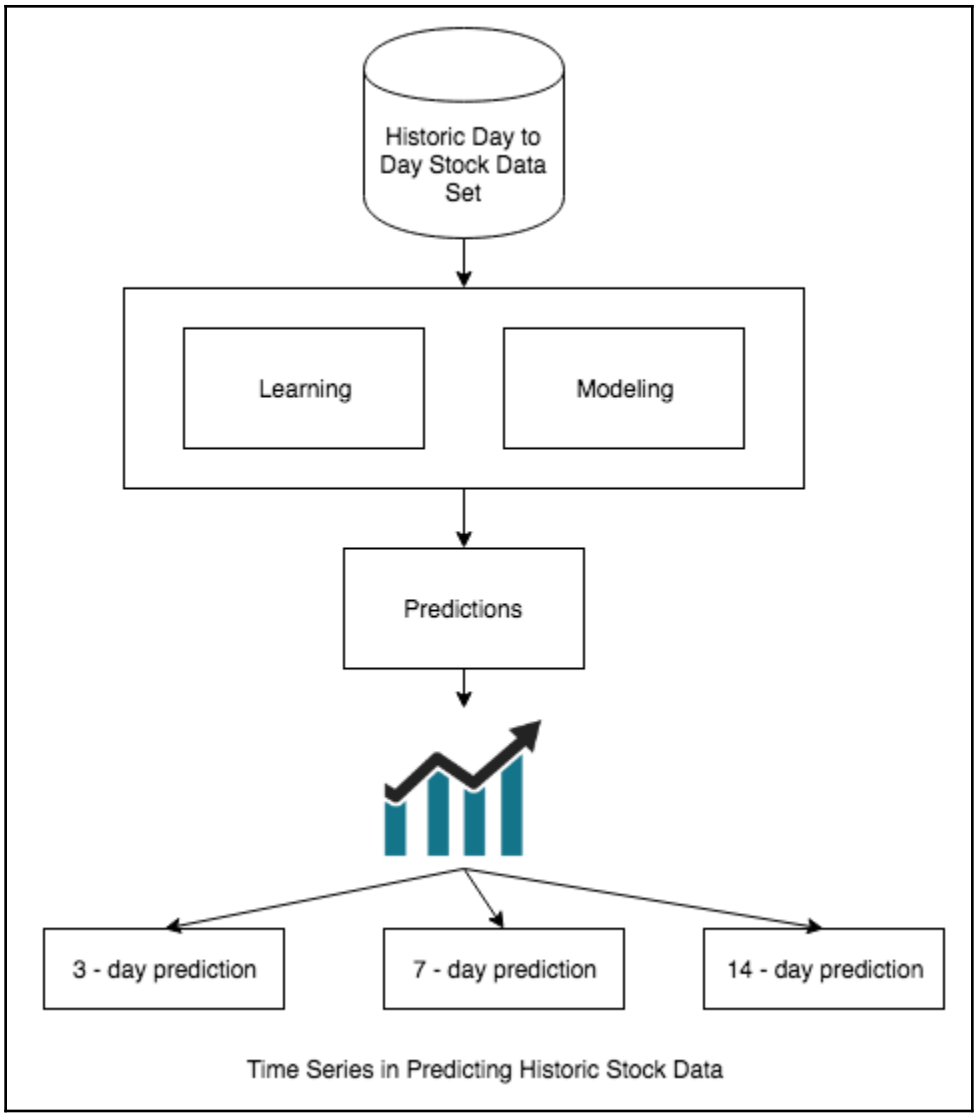

Partial Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50


Observed
Trend
Seasonal
Residual
120
100
80
60
40
20
0
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.06
0
20
40
60
80
Additive Model Decomposition Plot

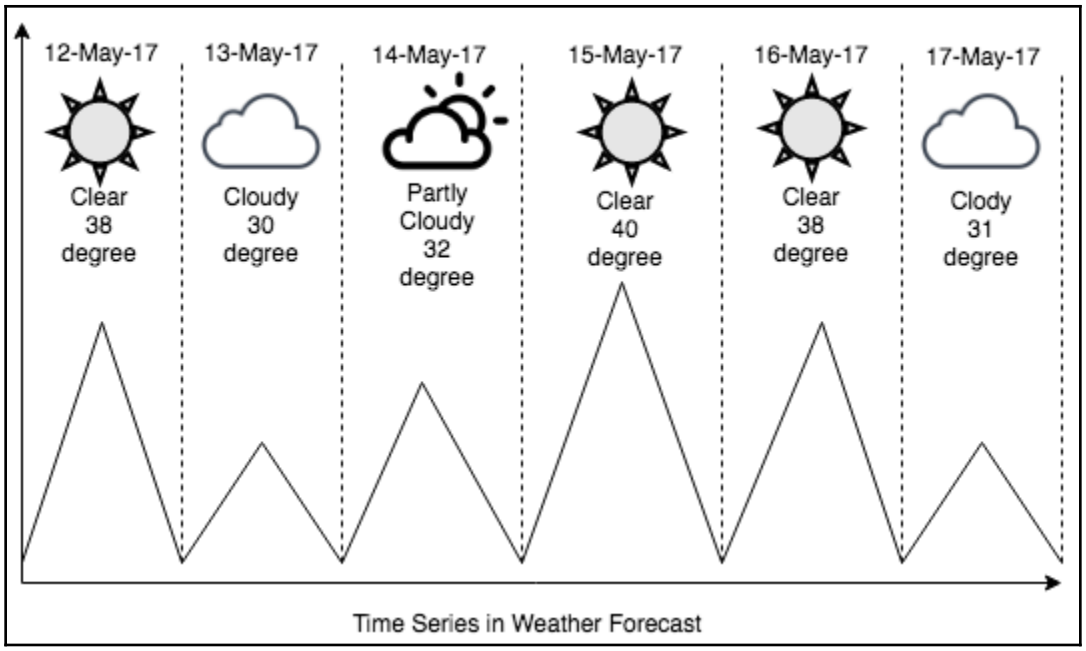

AirPassengers
100
200
300
400
500
600
1950
1952
1954
1956
1958
1960
Time

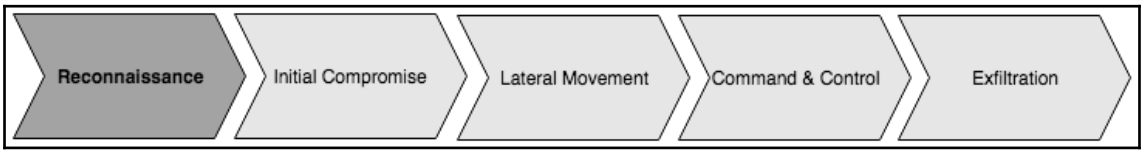

Observed
100
300
500
Trend
150
250
350
450
Seasonal
-40
-20
0
20
40
Random
-40
-20
0
20
40
60
2
4
6
8
10
12


Signal Processing
Amplitude
0.6
0.4
0.2
0
-0.2
-0.4



Time Series in Predicting Historic Stock Data





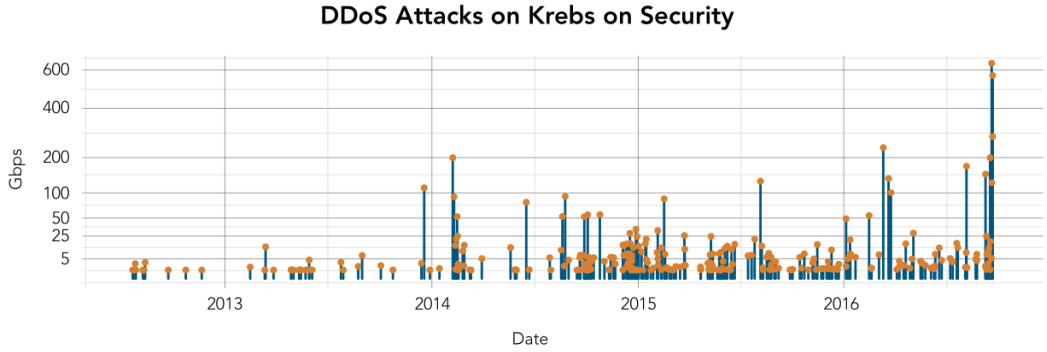
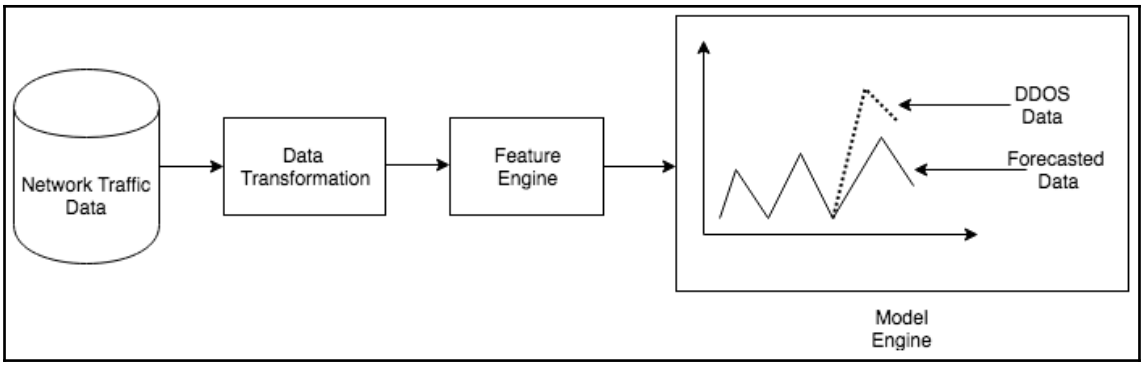
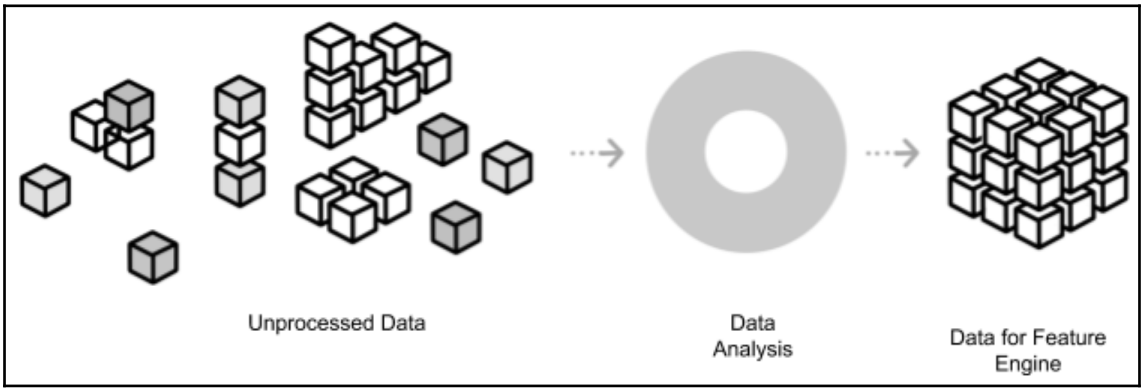
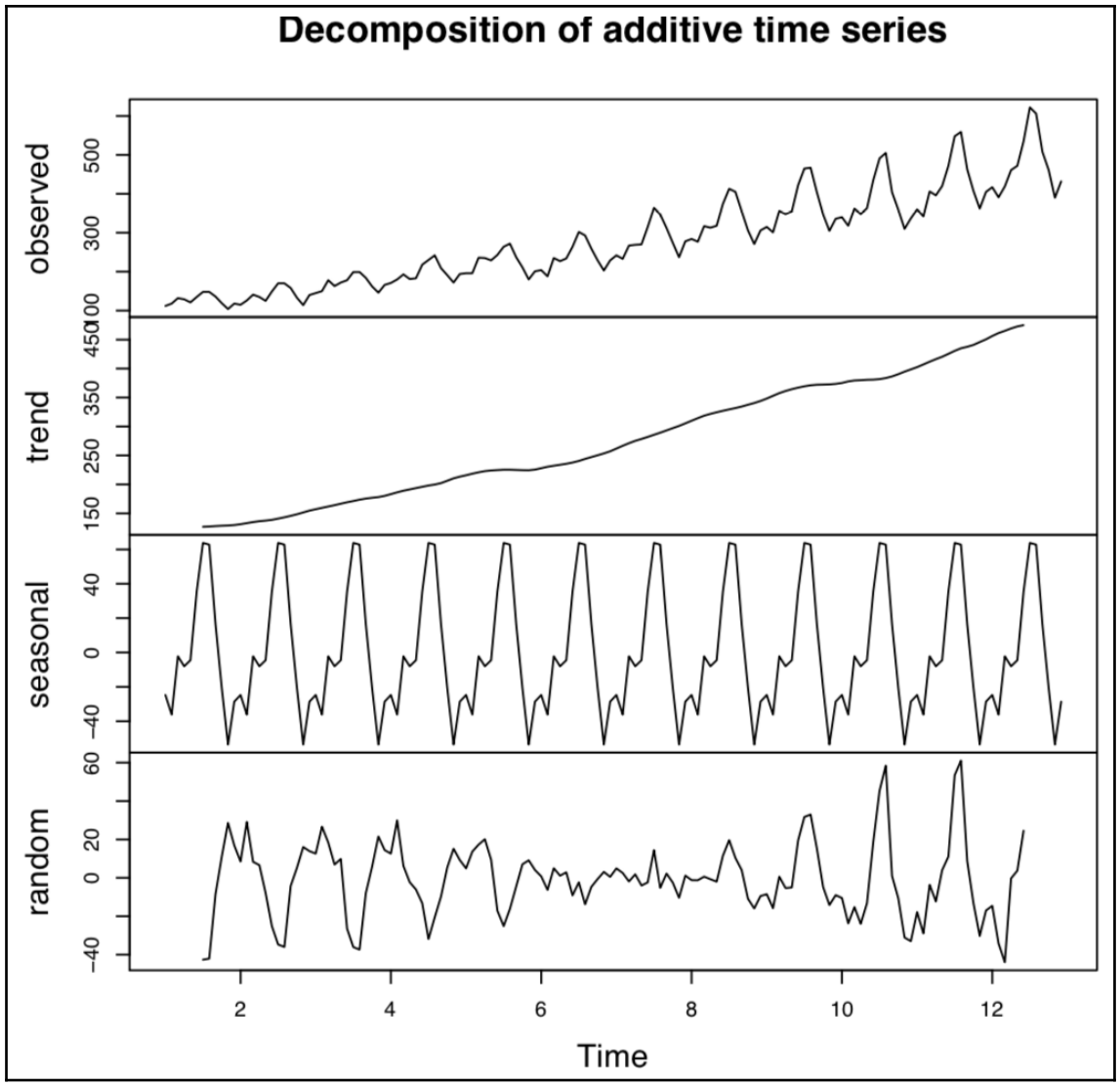
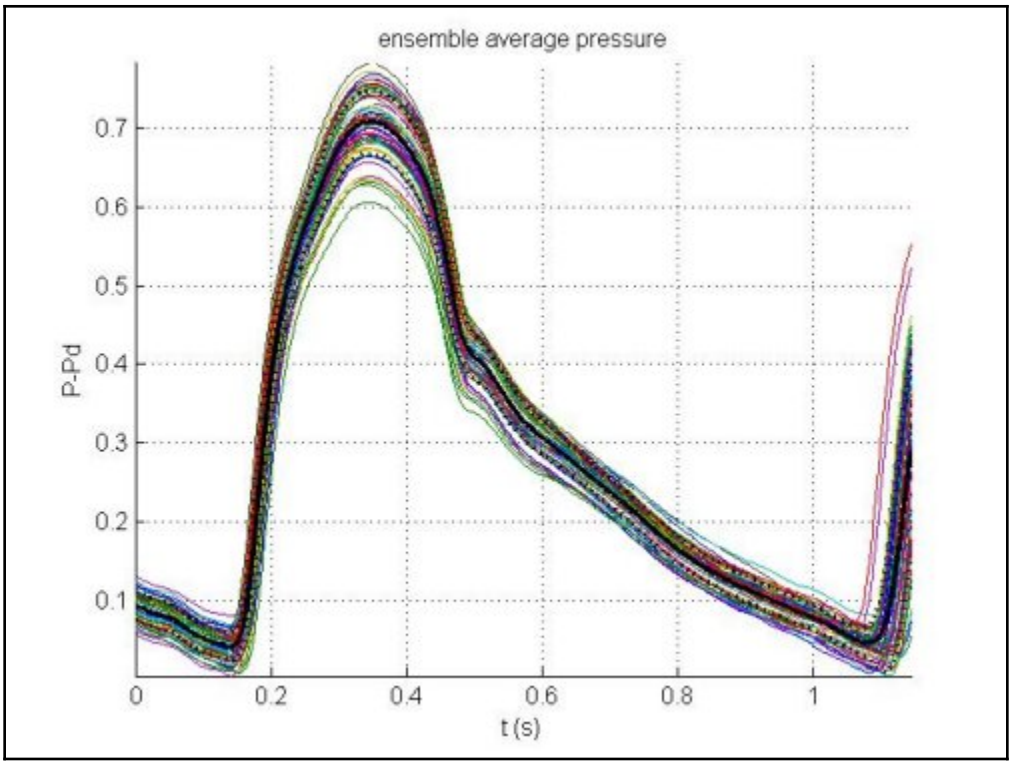


Figure 2-6: All attacks mitigated for krebsonsecurity.com while on the routed platform

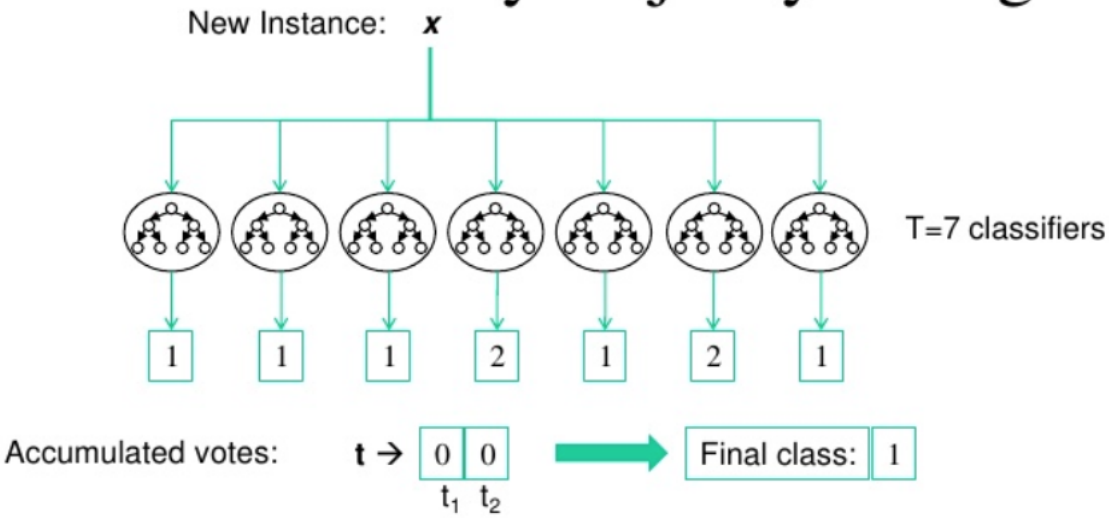

Network Traffic Data
Data Transformation
Feature Engine
DDOS Data
Forecasted Data
Model Engine


Unprocessed Data
Data Analysis
Data for Feature Engine


Decomposition of additive time series
observed
trend
seasonal
random
Time
2
4
6
8
10
12


ensemble average pressure
P-Pd
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.2
0.4
0.6
0.8
1
t (s)

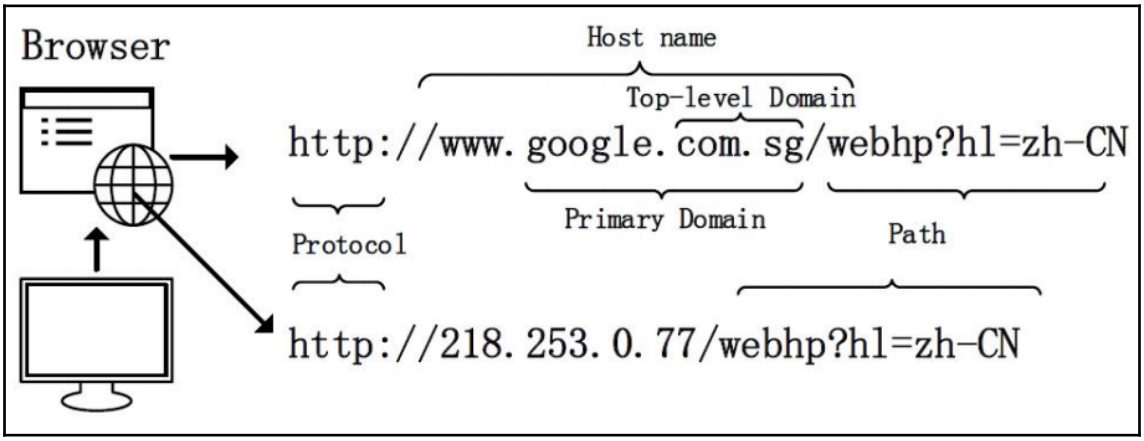
Classification by majority voting

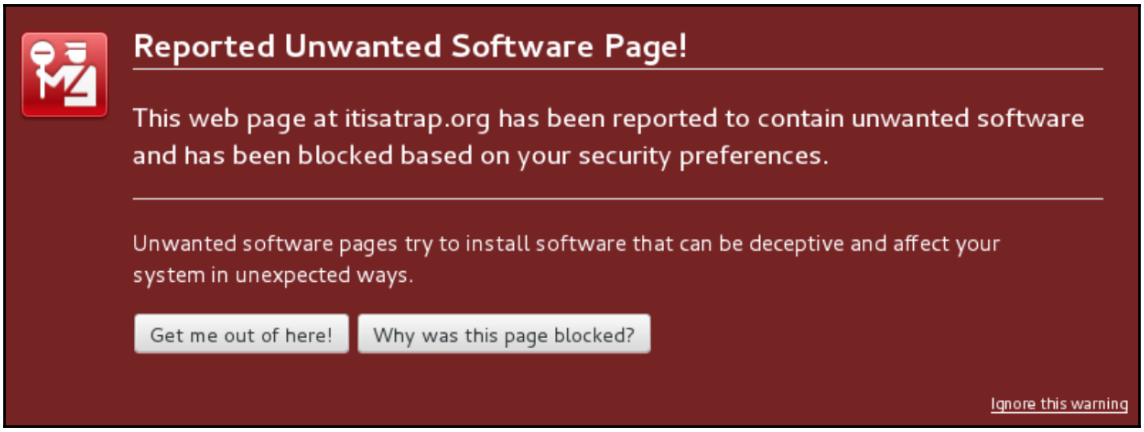


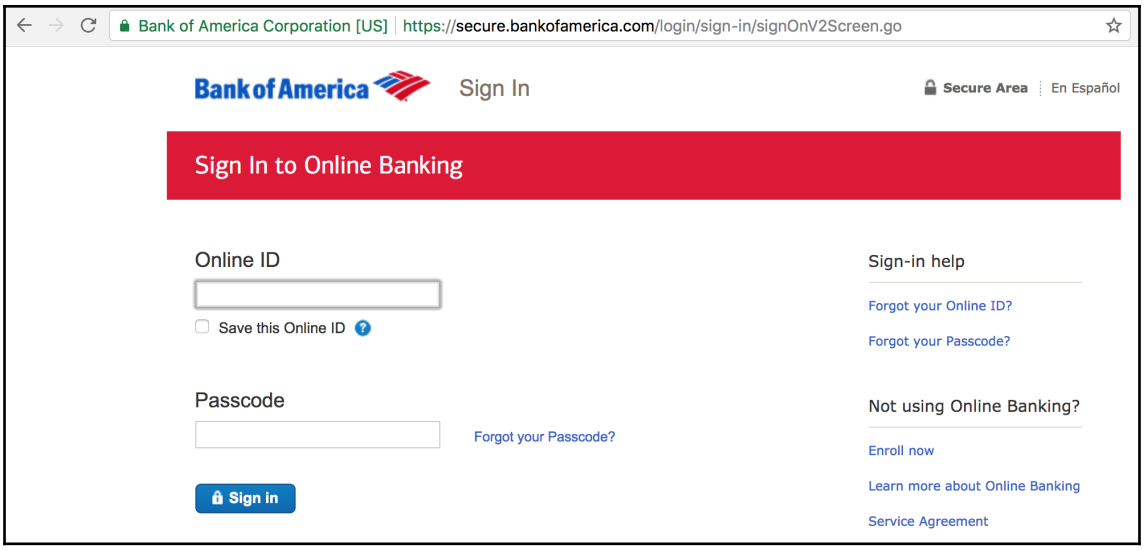
Alberto Suárez (2012)

15

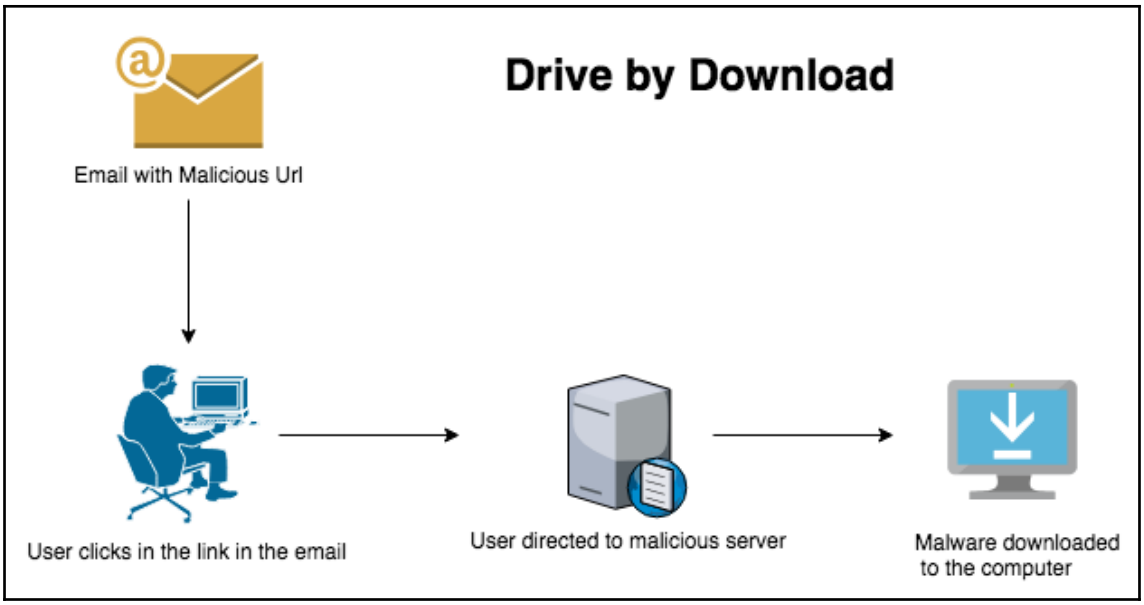
Chapter 3: Segregating Legitimate and Lousy URLs

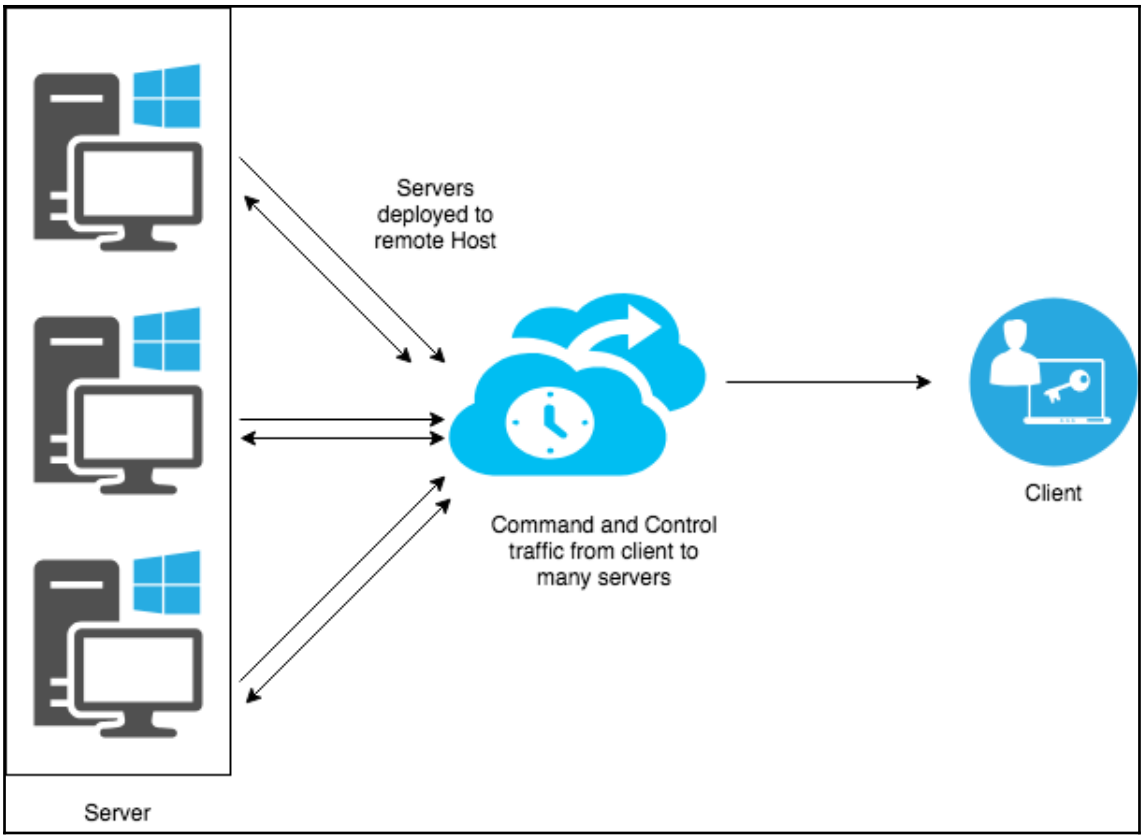


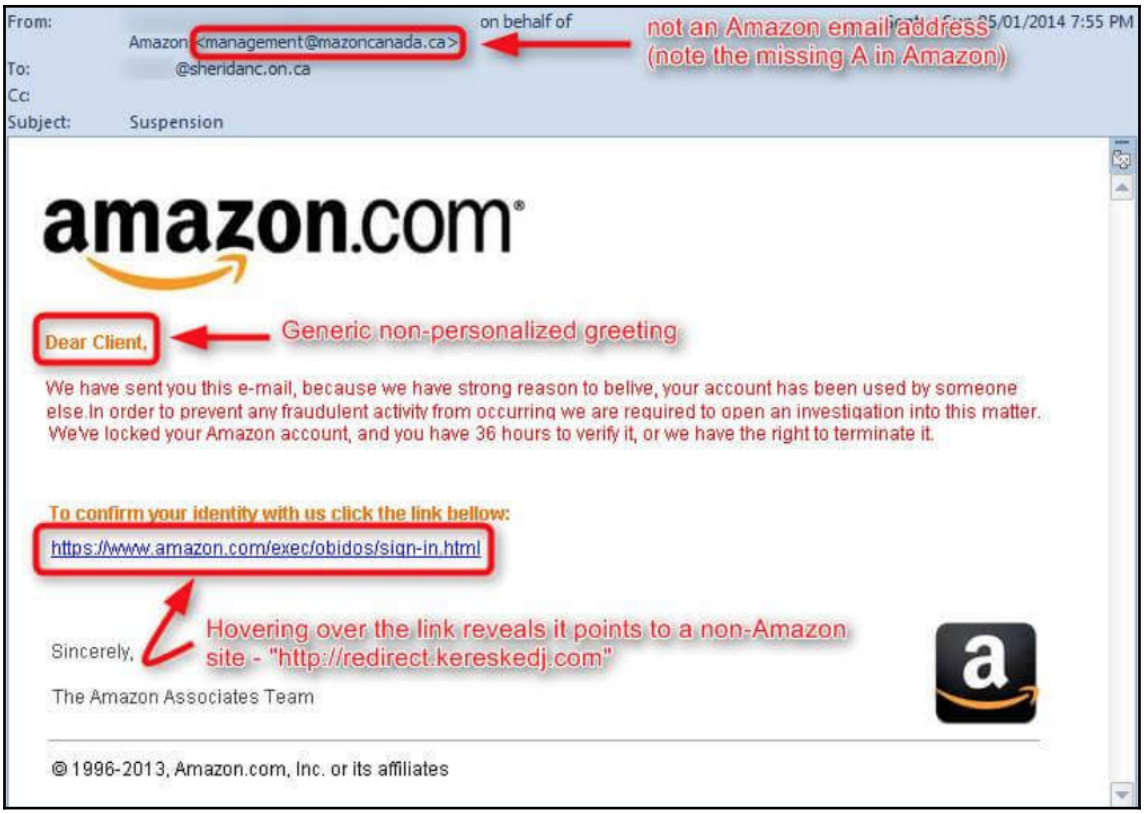



Bank of America Corporation [US] | https://secure.bankofamerica.com/login/sign-in/signOnV2Screen.go
Bank of America
Sign In
Secure Area
En Español
Sign In to Online Banking
Online ID
Save this Online ID
Passcode
Forgot your Passcode?
Sign in
Sign-in help
Forgot your Online ID?
Forgot your Passcode?
Not using Online Banking?
Enroll now
Learn more about Online Banking
Service Agreement

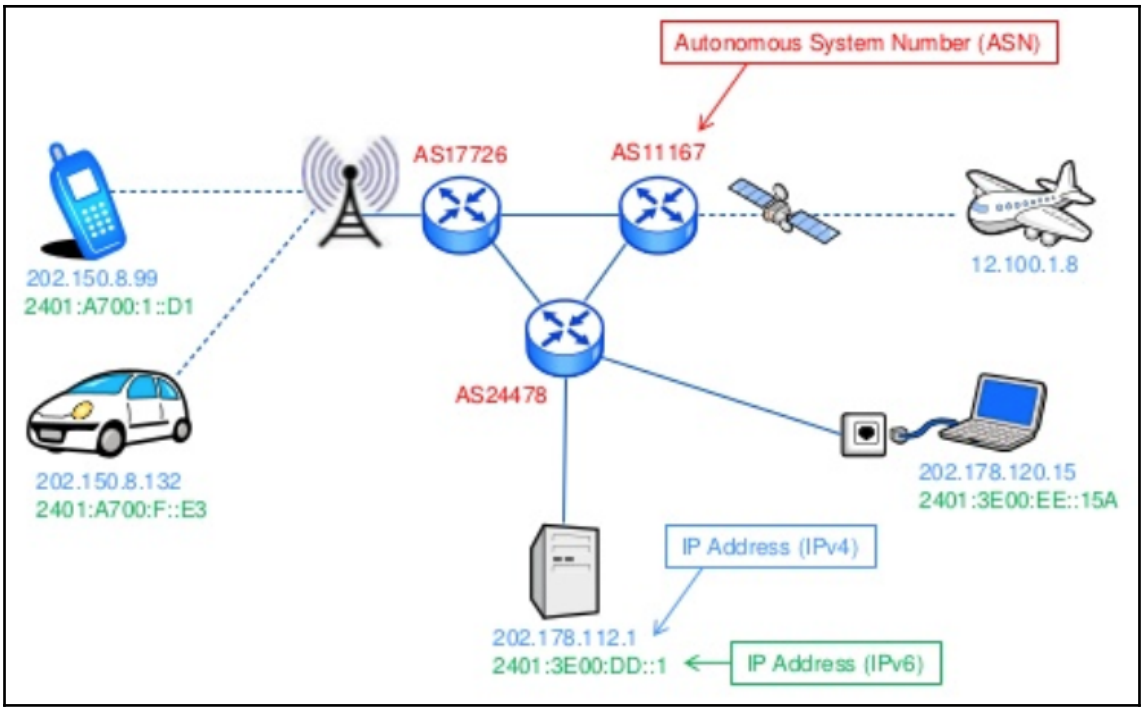
sclgchl1.eu.pn/index.html
Bank of America
Sign In
Secure Area
Your Online ID
As Part of our Security Measures to Protect your Account against Fraud and Scam Activities. You need to Pass through our Security Verification Protocols to view your Account. Please Bear With Us. Thank You.
Warning - Fake Page
Please enter your Online ID
Save this Online ID
Sign in
Sign-in help
Forgot your Online ID?
Not using Online Banking?


Drive by Download
Email with Malicious Url
User clicks in the link in the email
User directed to malicious server
Malware downloaded to the computer

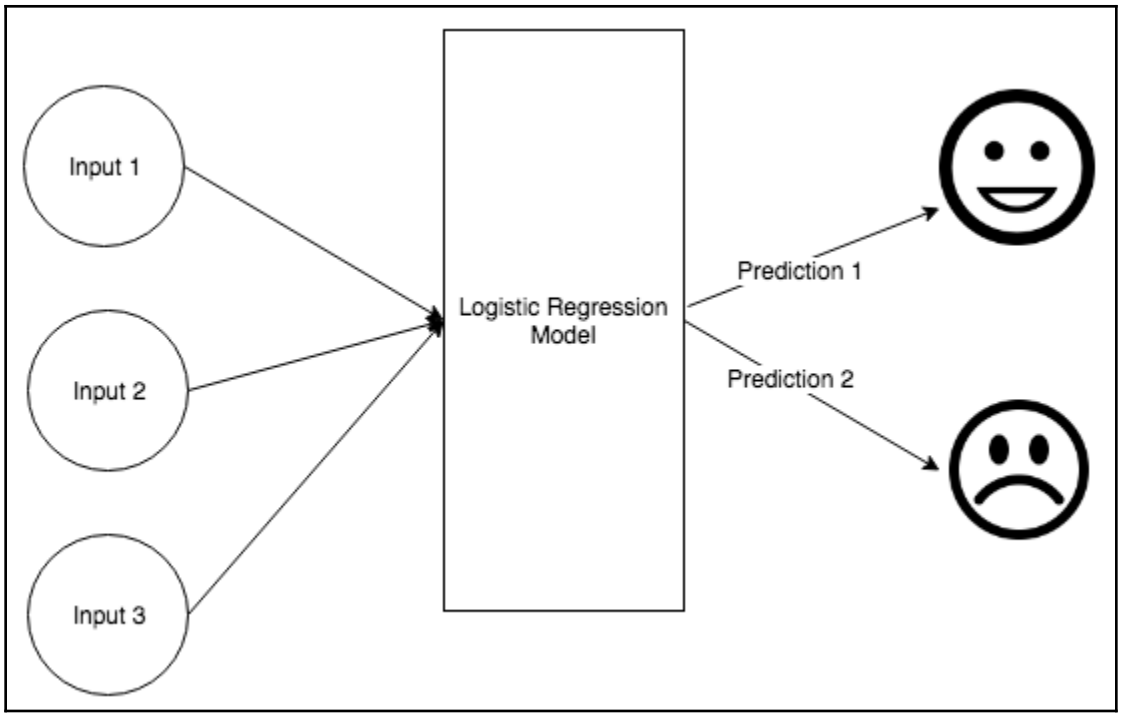

Servers
deployed to
remote Host
Command and Control
traffic from client to
many servers
Client
Server

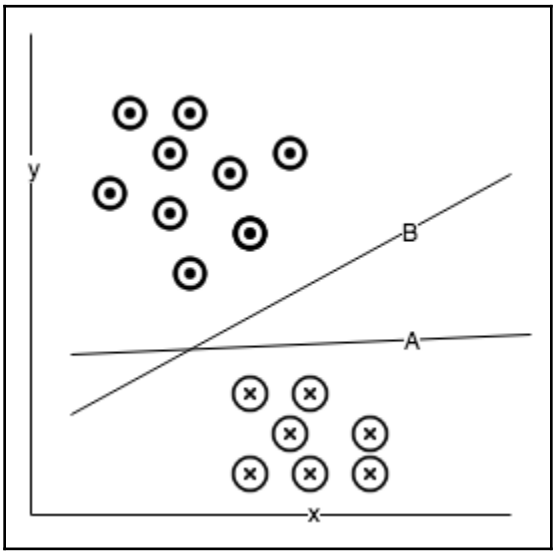

From:
on behalf of
Amazon <management@mazoncanada.ca>
not an Amazon email address
(note the missing A in Amazon)
To:
@sheridanc.on.ca
Cc:
Subject: Suspension
amazon.com
Dear Client,
Generic non-personalized greeting
We have sent you this e-mail, because we have strong reason to belive, your account has been used by someone else.In order to prevent any fraudulent activity from occurring we are required to open an investigation into this matter. We've locked your Amazon account, and you have 36 hours to verify it, or we have the right to terminate it.
To confirm your identity with us click the link bellow:
https://www.amazon.com/exec/obidos/sign-in.html
Hovering over the link reveals it points to a non-Amazon site - "http://redirect.kereskedj.com"
Sincerely,
The Amazon Associates Team
© 1996-2013, Amazon.com, Inc. or its affiliates

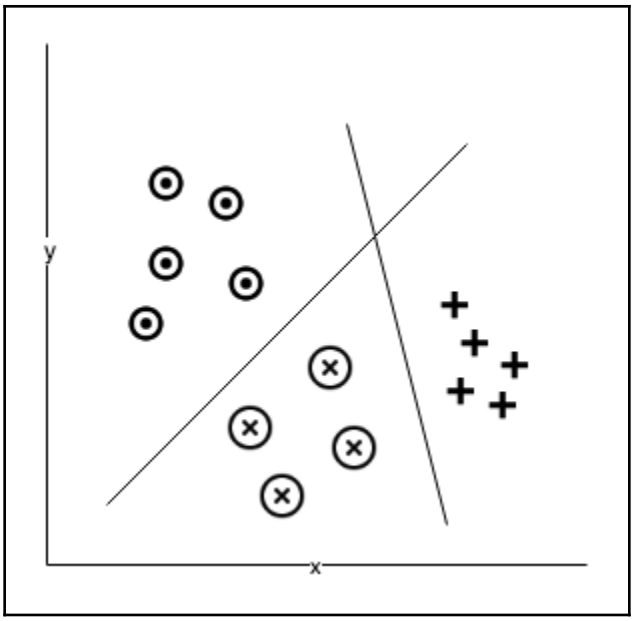
```
http://uol.com.br,0
http://google.pl,0
http://ebay.co.uk,0
http://netflix.com,0
http://dailymotion.com,0
http://cnet.com,0
http://delta-search.com,0
http://dailymail.co.uk,0
http://rakuten.co.jp,0
http://aliexpress.com,0
http://aol.com,0
http://dce.edu,0
http://google.com,0
http://stackoverflow.com/questions/8551735/how-do-i-run-python-code-from-sublime-text-2,0
http://paypal.com.webscr.cmd.login.submit.dispatch.5885d80a13c0db1f8e263663d3faee8db2b24f7b84f1819343fd6c338b1d9d.222studio.com
http://paypal-manager-login.net/konflikt/66211165125/,1
http://paypal-manager-loesung.net/konflikt/66211165125/,1
http://paypal-manager-account.net/konflikt/66211165125/,1
http://www.paypal-manager-account.net/konflikt/6222649185/index.php,1
http://www.paypal-manager-login.net/konflikt/79235228200/index.php,1
http://paypal.com.laveaki.com.br/PayPal.com/,1
http://paypal.com.client.identifiant.compte.clefs.informations.upgarde.mon.compte.personnel.ghs56hge556rg4h6qe4th654f84e84r8e.h
http://msd2003.com/rche/index.htm,1
http://paypal-manager-loesung.net/konflikt/6624985147/index.php,1
http://eadideal.com.br/conteudos/material/66/Netzro-Login.html,1
http://tahmid.ir/user/?cmd=_home&amp;dispatch=5885d80a13c0db1f8e&amp;ee=669498ee5d0b4b381fd2bb1cceb52112,1
http://208.115.247.198/paypal.com.au.434324.fdsfsd32423423.fadfafas.3423423432fsfdafsdfsd/index1.php,1
http://www.paypal-manager-service.net/konflikt/785549116/index.php?webapps=/mpp/verkaufen,1
http://www.nervemobilization.com/update/06bb5b5bd8b1be54b78648f4993a1de1/protect.html,1
http://www.nervemobilization.com/update/15f58e7e4a736e2cd442b733e8755716/protect.html,1
http://sionaviatur.ru/wp-includes/syystemyj/index50.htm,1
```


Autonomous System Number (ASN)
AS17726
AS11167
AS24478
202.150.8.99
2401:A700:1::D1
202.150.8.132
2401:A700:F::E3
12.100.1.8
202.178.120.15
2401:3E00:EE::15A
IP Address (IPv4)
202.178.112.1
2401:3E00:DD::1
IP Address (IPv6)

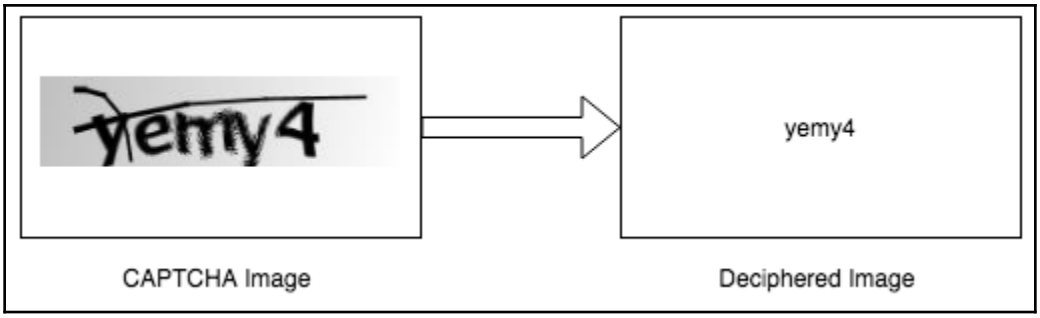
Google	google.com	1
YouTube	youtube.com	2
Facebook	facebook.com	3
Baidu	baidu.com	4
Wikipedia	wikipedia.org	5
Reddit	reddit.com	6
Yahoo!	yahoo.com	7
Google India	google.co.in	8
Tencent QQ	qq.com	9
Amazon	amazon.com	10


Input 1
Input 2
Input 3
Logistic Regression Model
Prediction 1
Prediction 2


y
B
A
x


y
x

Chapter 4: Knocking Down CAPTCHAs



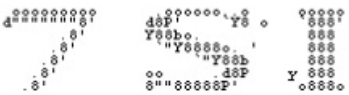
Type the two words
CAPTCHA
(a) TEXT CAPTCHA
Click the flower
(b) IMAGE CAPTCHA
Audio Verification
(c) AUDIO CAPTCHA
(d) VIDEO CAPTCHA

Other Types of CAPTCHA

- Animated CAPTCHAs

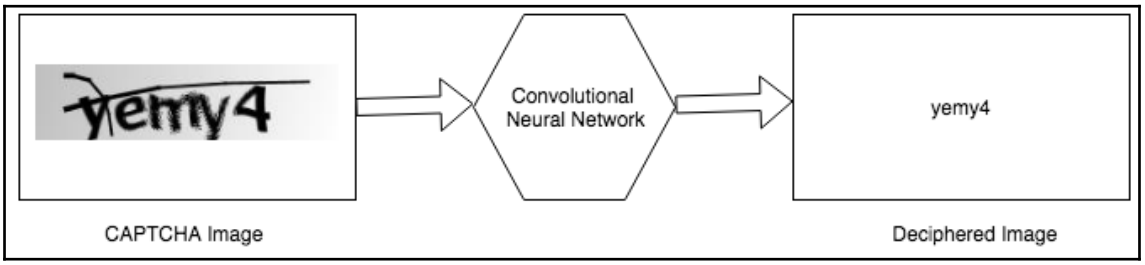
- 3D CAPTCHA

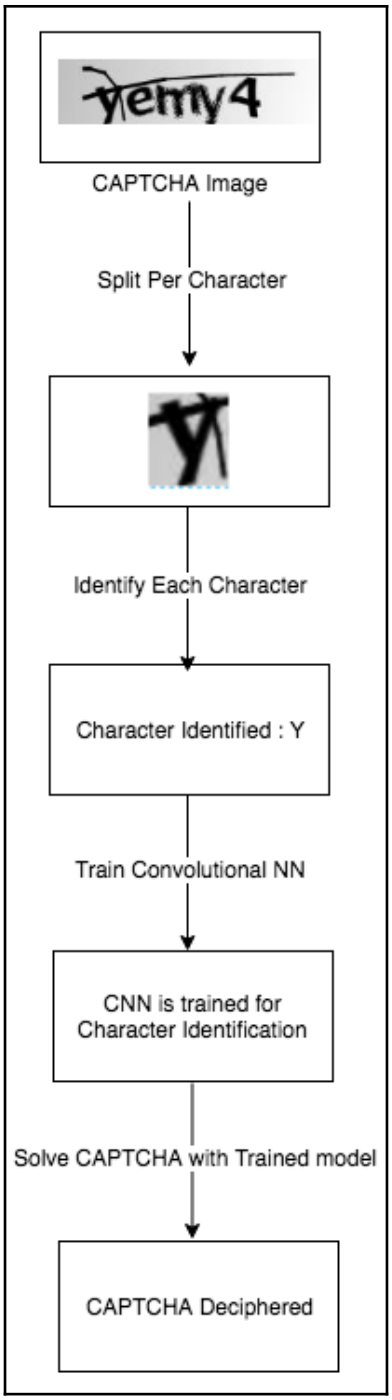
- ASCII art



- Reverse CAPTCHA "Leave this field blank"

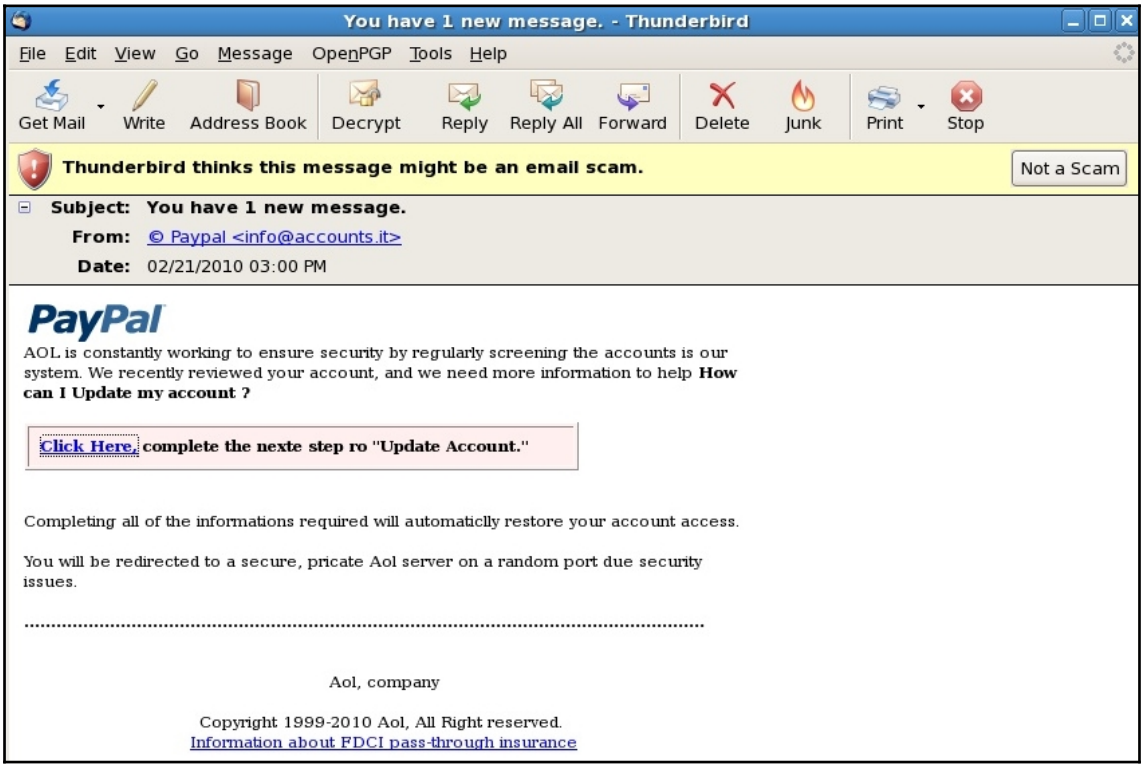
Select all images below
that match this one:
Verify


Convolutional
Neural Network
yemy4
CAPTCHA Image
Deciphered Image


CAPTCHA Image
Split Per Character
Identify Each Character
Character Identified : Y
Train Convolutional NN
CNN is trained for
Character Identification
Solve CAPTCHA with Trained model
CAPTCHA Deciphered

RDAQ
RDAQ

Chapter 5: Using Data Science to Catch Email Fraud and Spam



From: DesignerWatches by LR [redacted]
To: [redacted]
Cc:
Subject: Start Black Friday today

LR Luxury Replicas

BLACK FRIDAY EVERY DAY UNTIL FRIDAY NOVEMBER 23RD!

The best quality watch replications on PLANET EARTH!
The lowest priced high-end watches on the INTERNET!

~~$131.00~~ **Now $89.00**

ROLEX OYSTER PERPETUAL DATEJUST

Model #: RX459
Gender: Men
Availability: INSTOCK - SHIPS WITHIN 24 HOURS

www.LRblackfridaytoday.com

BLACK FRIDAY HAS STARTED!

Black Friday Every day until November 23rd!
All items reduced by 25-50% as of TODAY.

Over 25,000 exact watch-copies have been reduced until Friday November 23rd.
There plenty of time to get the watch of your dreams but we recommend doing it as soon as possible.
This will ensure INSTOCK availability and fast delivery.

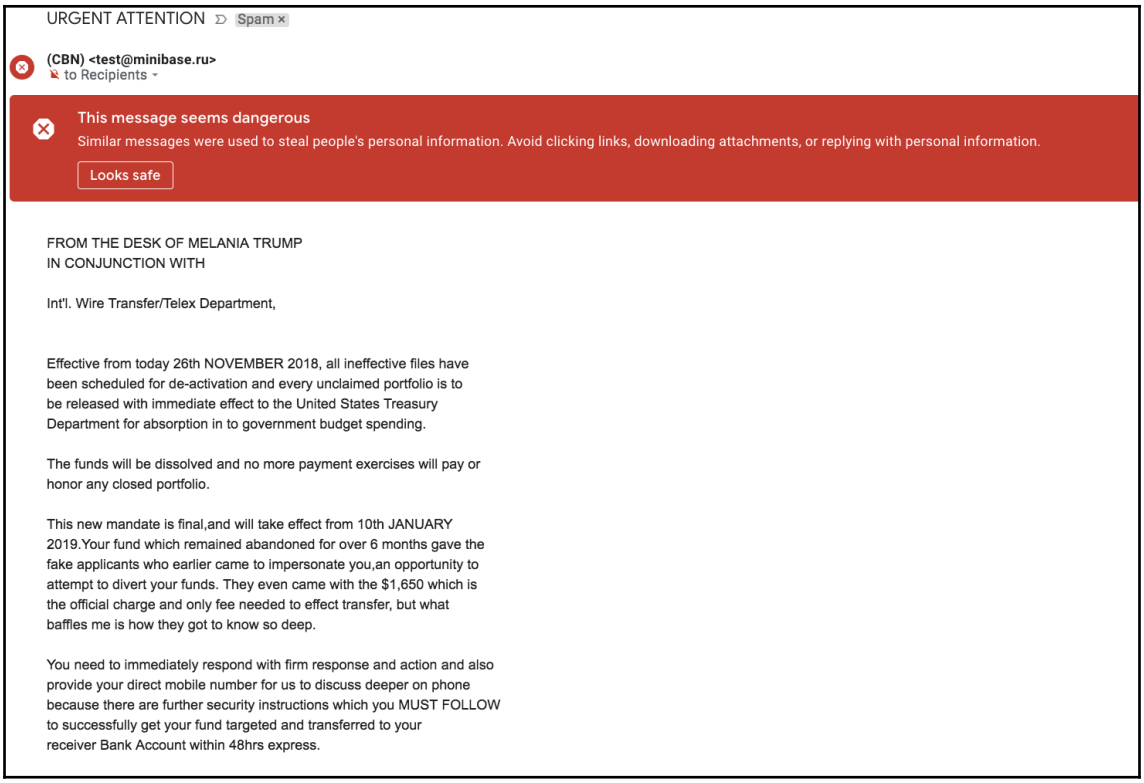
NOTE: **BLACK FRIDAY PRICES ARE AVAILABLE ON INSTOCK ITEMS ONLY!**
Currently every watch-model is INSTOCK and ready to ship within 1 hour.

THESE ARE NOT CHEAP CHINA KNOCK-OFFS:

These are hand crafted high-end watch-copies.
These are made using identical parts and materials.
These are tested inside and out to be identical.
There is no difference between our watch-copies and the originals!

http://lrblackfridaytoday.com

GFI®


URGENT ATTENTION Spam ×

(CBN) <test@minibase.ru>
to Recipients

This message seems dangerous
Similar messages were used to steal people's personal information. Avoid clicking links, downloading attachments, or replying with personal information.
Looks safe

FROM THE DESK OF MELANIA TRUMP
IN CONJUNCTION WITH

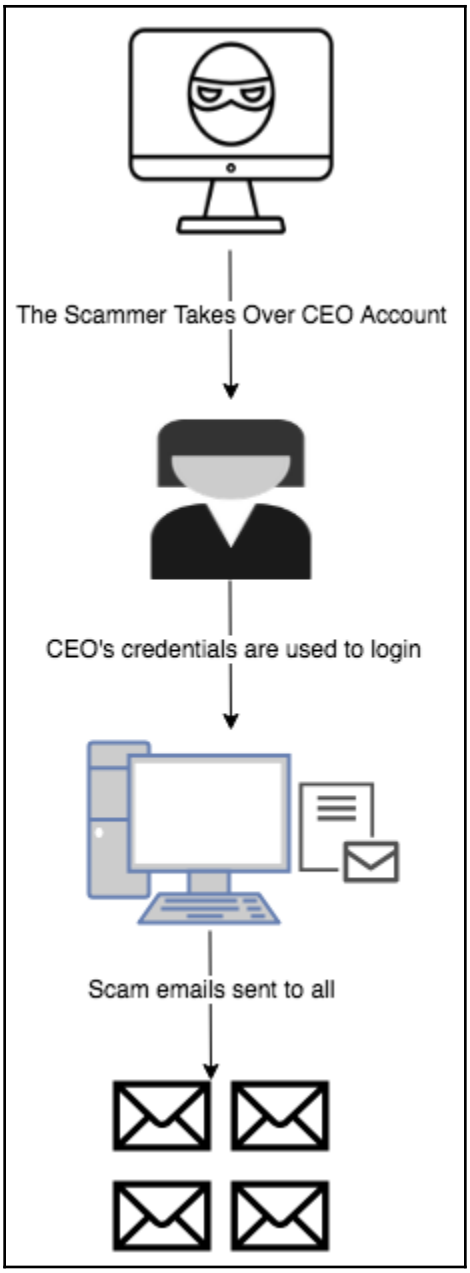
Int'l. Wire Transfer/Telex Department,

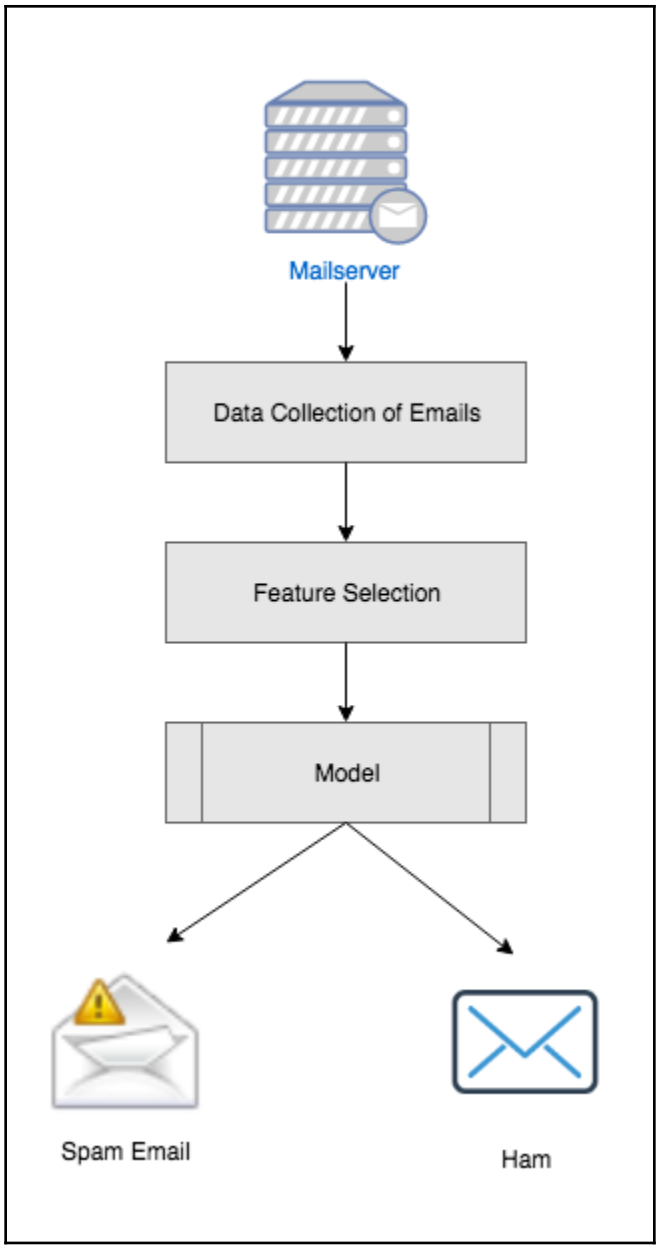
Effective from today 26th NOVEMBER 2018, all ineffective files have been scheduled for de-activation and every unclaimed portfolio is to be released with immediate effect to the United States Treasury Department for absorption in to government budget spending.

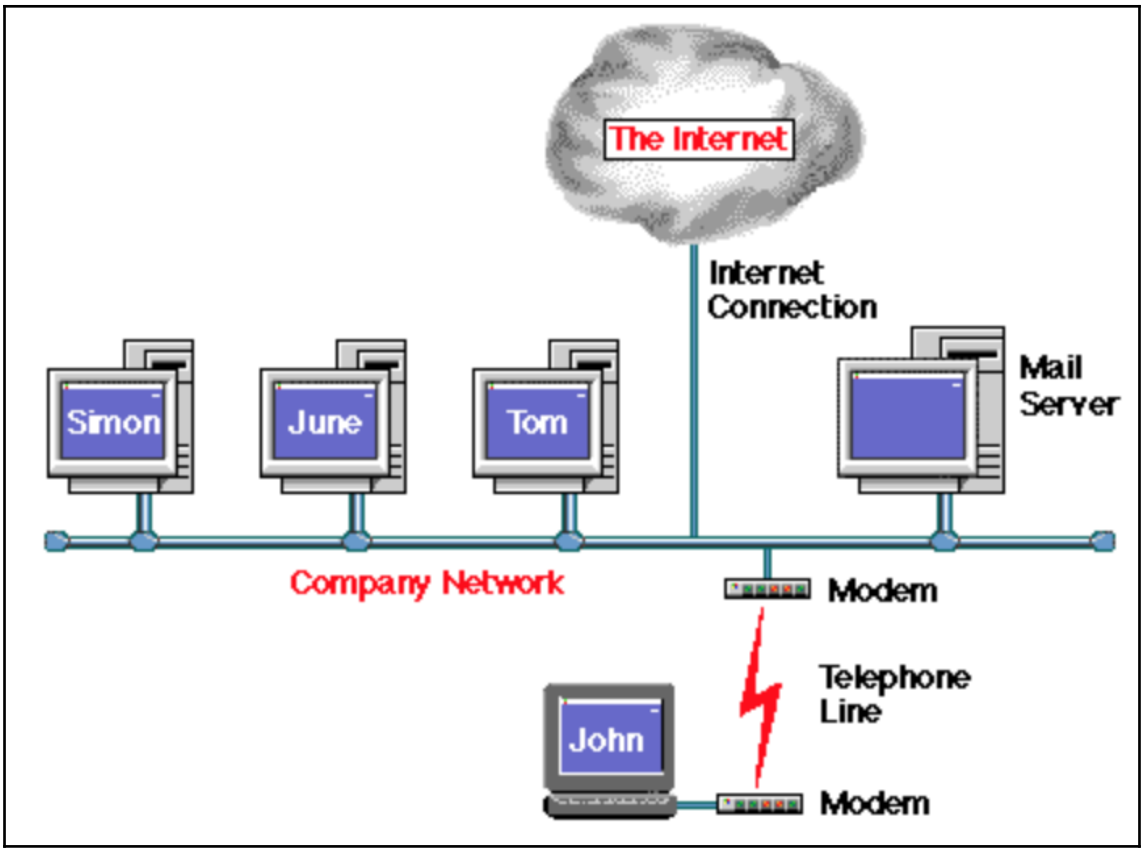
The funds will be dissolved and no more payment exercises will pay or honor any closed portfolio.

This new mandate is final,and will take effect from 10th JANUARY 2019.Your fund which remained abandoned for over 6 months gave the fake applicants who earlier came to impersonate you,an opportunity to attempt to divert your funds. They even came with the $1,650 which is the official charge and only fee needed to effect transfer, but what baffles me is how they got to know so deep.

You need to immediately respond with firm response and action and also provide your direct mobile number for us to discuss deeper on phone because there are further security instructions which you MUST FOLLOW to successfully get your fund targeted and transferred to your receiver Bank Account within 48hrs express.


The Scammer Takes Over CEO Account
CEO's credentials are used to login
Scam emails sent to all


Mailserver
Data Collection of Emails
Feature Selection
Model
Spam Email
Ham

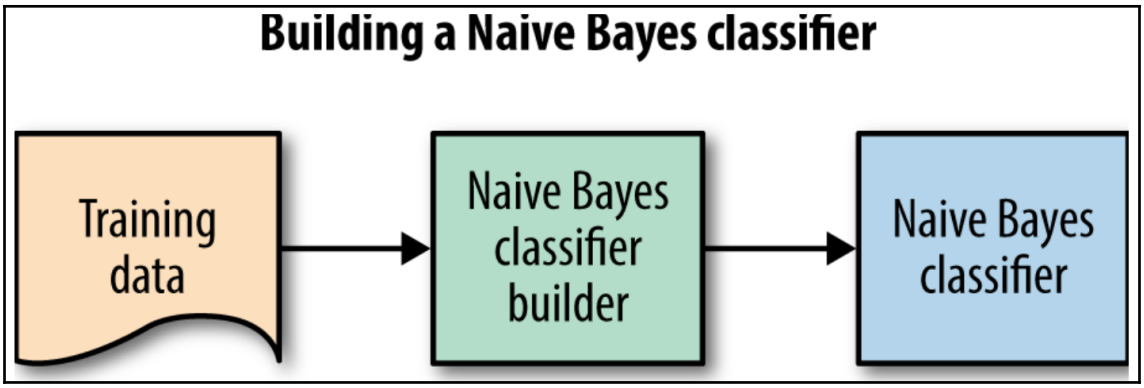

The Internet
Internet
Connection
Simon
June
Tom
Mail
Server
Company Network
Modem
Telephone
Line
John
Modem

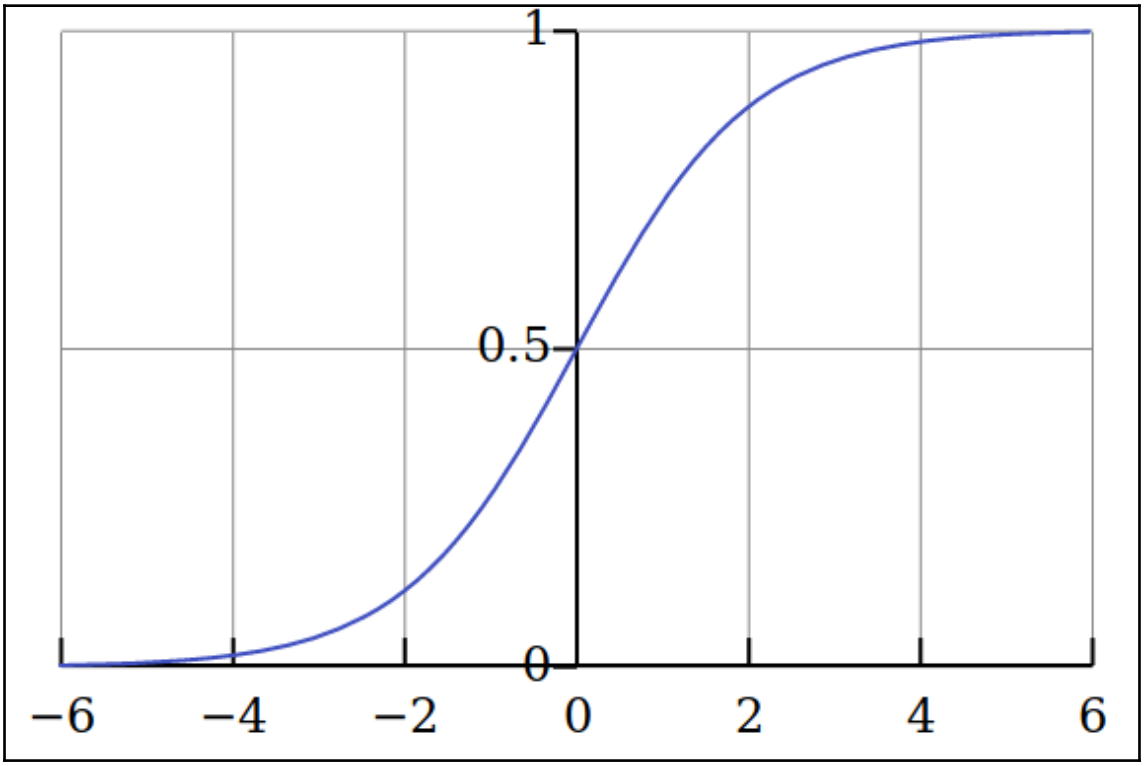
	ham 87% spam 13%	5169 unique values	43 unique values	10 unique values	5 unique values
1	ham	Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...			
2	ham	Ok lar... Joking wif u oni...			
3	spam	Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)T&C's apply 08452810075over18's			
4	ham	U dun say so early hor... U c already then say...			
5	ham	Nah I don't think he goes to usf, he lives around here			

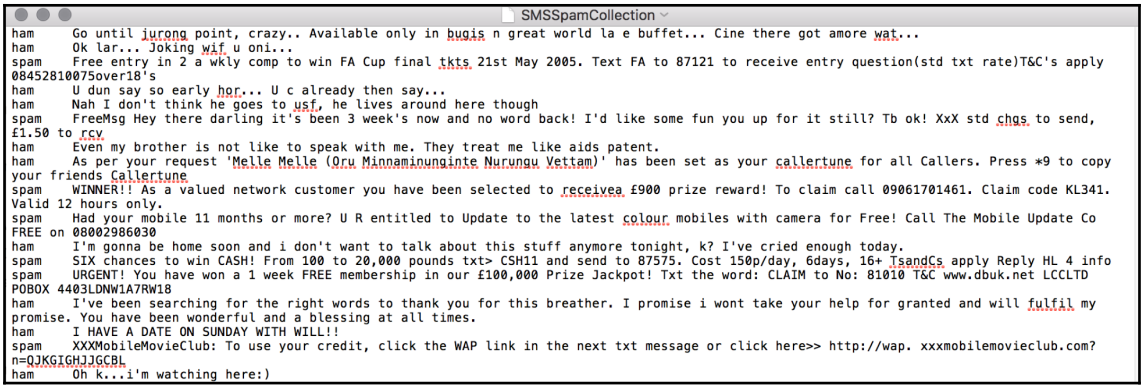
$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)},$$

where A and B are events and $P(B) \neq 0$.

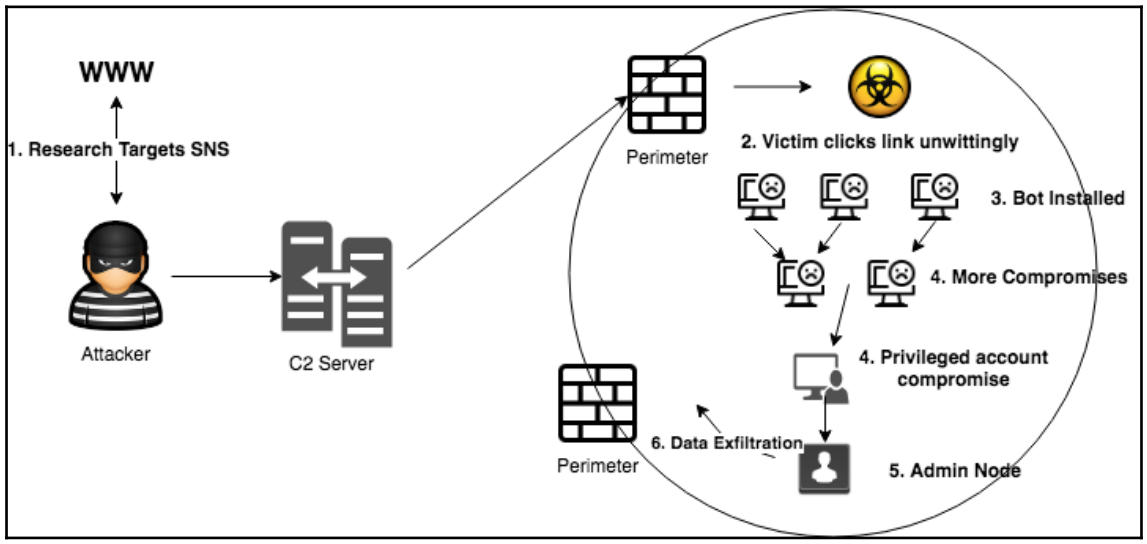
- $P(A)$ and $P(B)$ are the probabilities of observing A and B without regard to each other.
- $P(A \mid B)$, a conditional probability, is the probability of observing event A given that B is true.
- $P(B \mid A)$ is the probability of observing event B given that A is true.

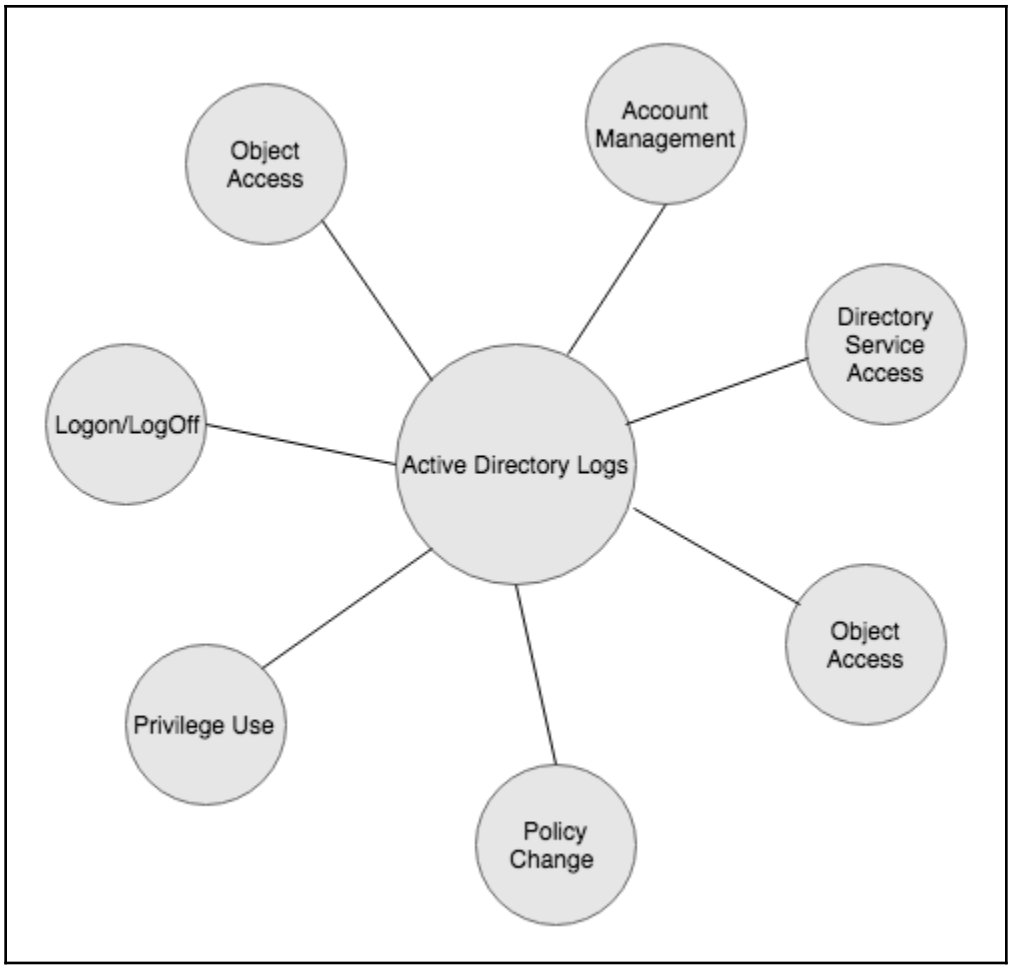

Building a Naive Bayes classifier
Training data
Naive Bayes classifier builder
Naive Bayes classifier

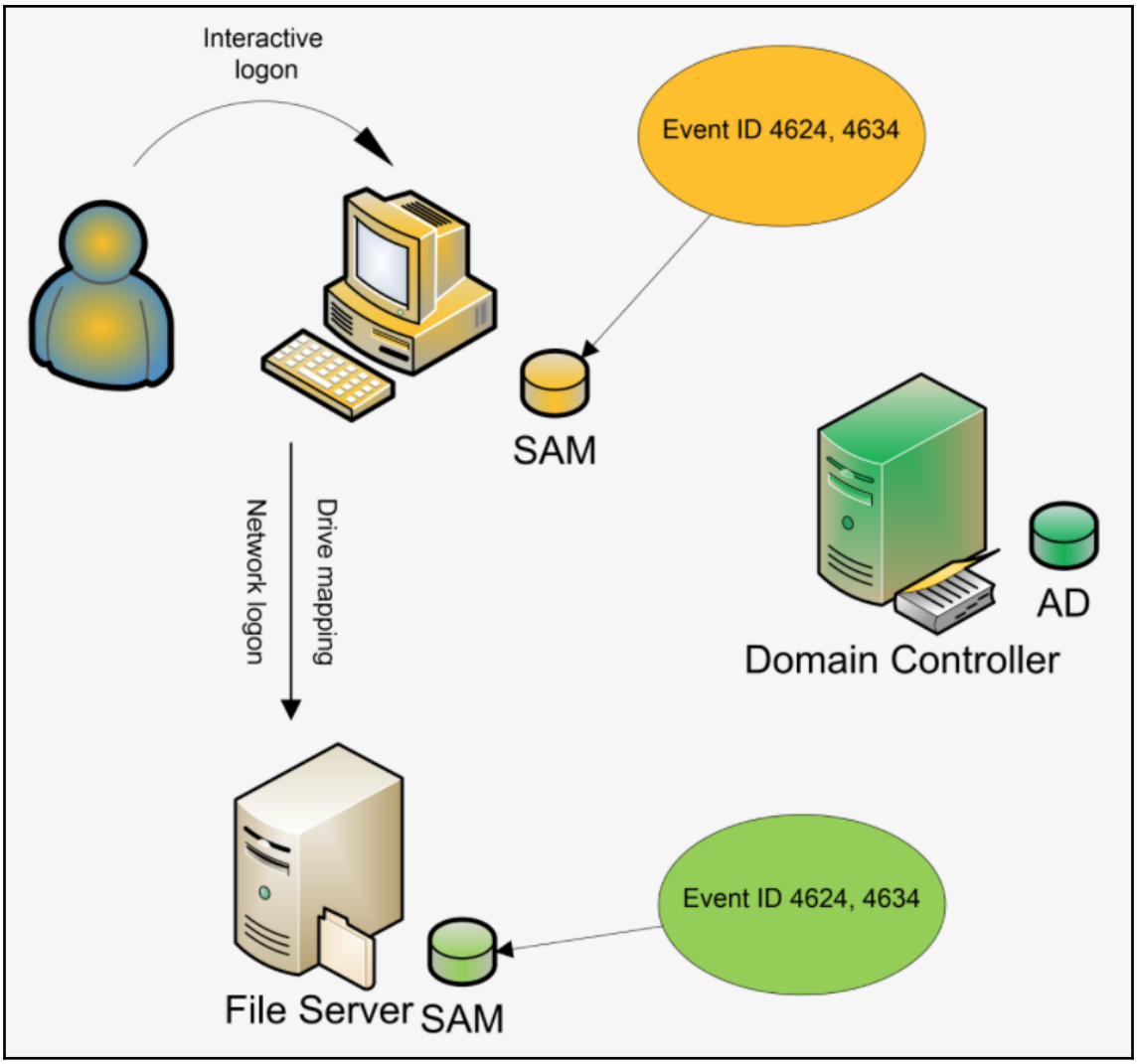

1
0.5
0
−6
−4
−2
0
2
4
6

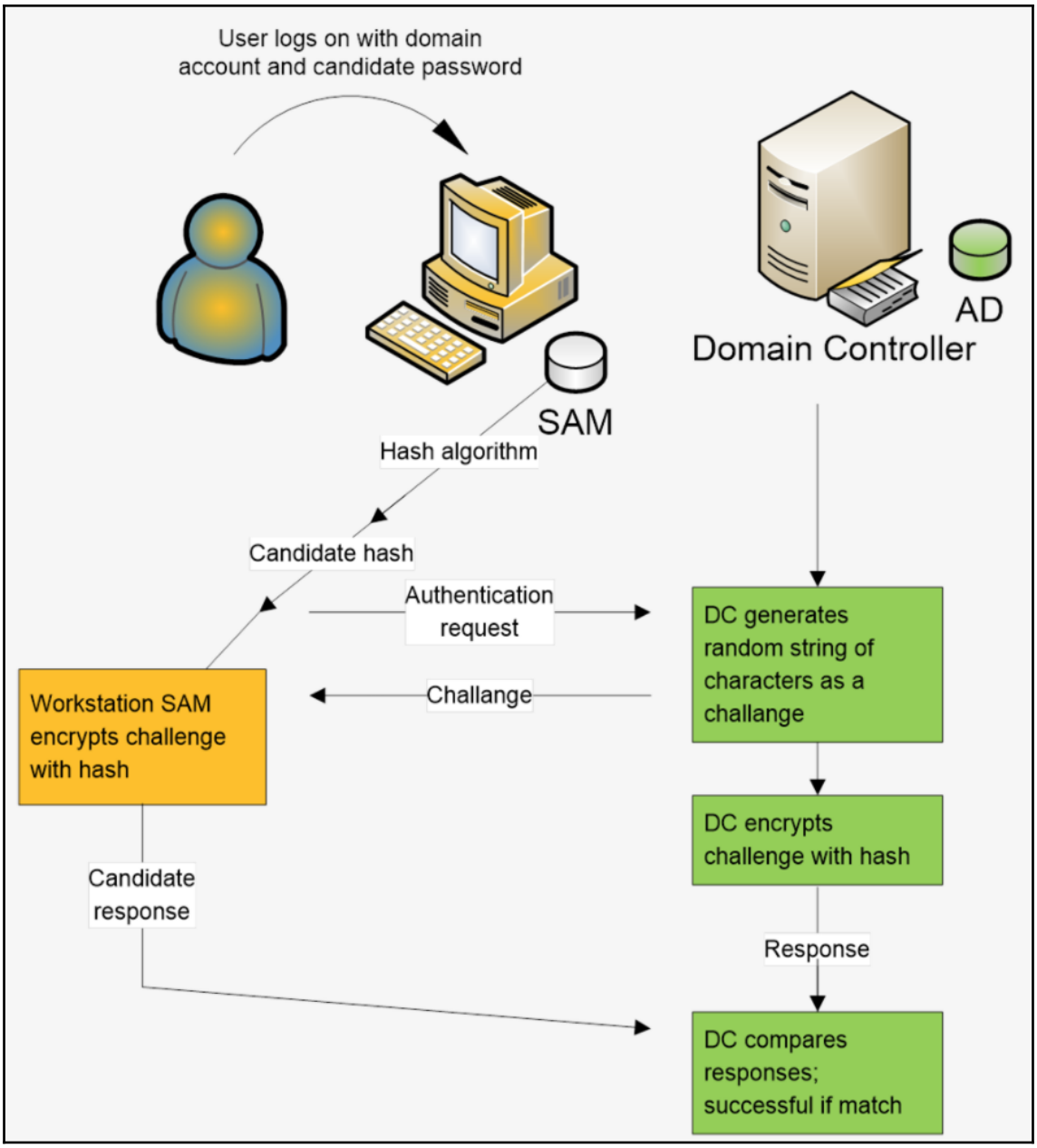

SMSSpamCollection
ham Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...
ham Ok lar... Joking wif u oni...
spam Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)T&C's apply 08452810075over18's
ham U dun say so early hor... U c already then say...
ham Nah I don't think he goes to usf, he lives around here though
spam FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb ok! XxX std chgs to send, £1.50 to rcv
ham Even my brother is not like to speak with me. They treat me like aids patent.
ham As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertune for all Callers. Press *9 to copy your friends Callertune
spam WINNER!! As a valued network customer you have been selected to receivea £900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only.
spam Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The Mobile Update Co FREE on 08002986030
ham I'm gonna be home soon and i don't want to talk about this stuff anymore tonight, k? I've cried enough today.
spam SIX chances to win CASH! From 100 to 20,000 pounds txt> CSH11 and send to 87575. Cost 150p/day, 6days, 16+ TsandCs apply Reply HL 4 info
spam URGENT! You have won a 1 week FREE membership in our £100,000 Prize Jackpot! Txt the word: CLAIM to No: 81010 T&C www.dbuk.net LCCLTD POBOX 4403LDNW1A7RW18
ham I've been searching for the right words to thank you for this breather. I promise i wont take your help for granted and will fulfil my promise. You have been wonderful and a blessing at all times.
ham I HAVE A DATE ON SUNDAY WITH WILL!!
spam XXXMobileMovieClub: To use your credit, click the WAP link in the next txt message or click here>> http://wap. xxxmobilemovieclub.com?n=QJKGIGHJJGCBL
ham Oh k...i'm watching here:)

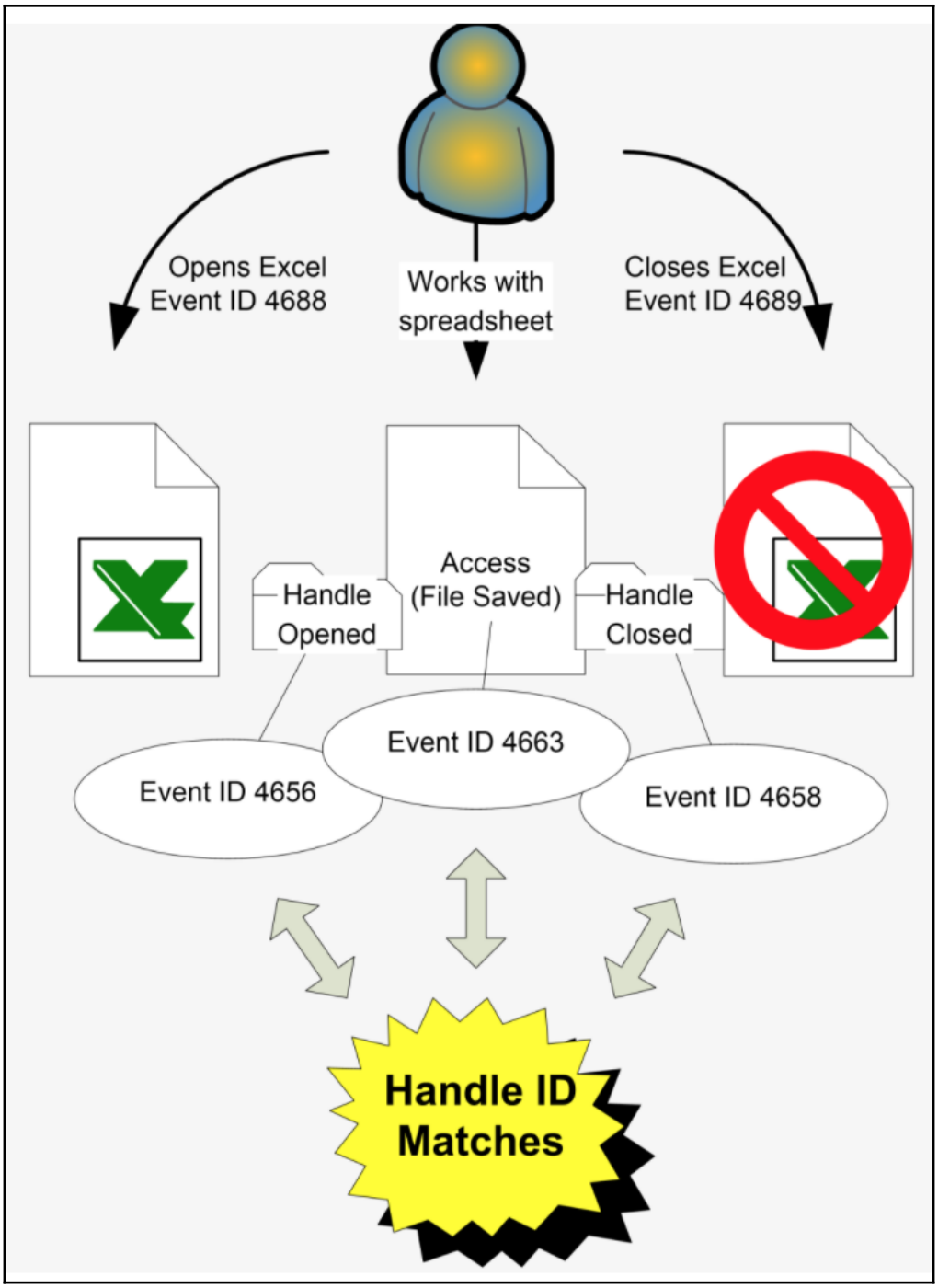
Chapter 6: Efficient Network Anomaly Detection Using k-means




Object Access
Account Management
Directory Service Access
Logon/LogOff
Active Directory Logs
Object Access
Privilege Use
Policy Change


Interactive logon
Event ID 4624, 4634
SAM
Network logon
Drive mapping
AD
Domain Controller
File Server
SAM
Event ID 4624, 4634


User logs on with domain account and candidate password
AD
Domain Controller
SAM
Hash algorithm
Candidate hash
Authentication request
Challange
DC generates random string of characters as a challange
Workstation SAM encrypts challenge with hash
DC encrypts challenge with hash
Candidate response
Response
DC compares responses; successful if match


Opens Excel
Event ID 4688
Works with
spreadsheet
Closes Excel
Event ID 4689
Handle
Opened
Access
(File Saved)
Handle
Closed
Event ID 4656
Event ID 4663
Event ID 4658
Handle ID
Matches

Log Name: Security
Source: Microsoft-Windows-Security-Auditing
Date: 10/28/2008 6:17:28 PM
Event ID: 4768
Task Category: Kerberos Authentication Service
Level: Information
Keywords: Audit Success
User: N/A
Computer: Lab2008.acme.ru
Description:
A Kerberos authentication ticket (TGT) was requested.

Account Information:
Account Name: Fred
Supplied Realm Name: ACME
User ID: ACME\Fred

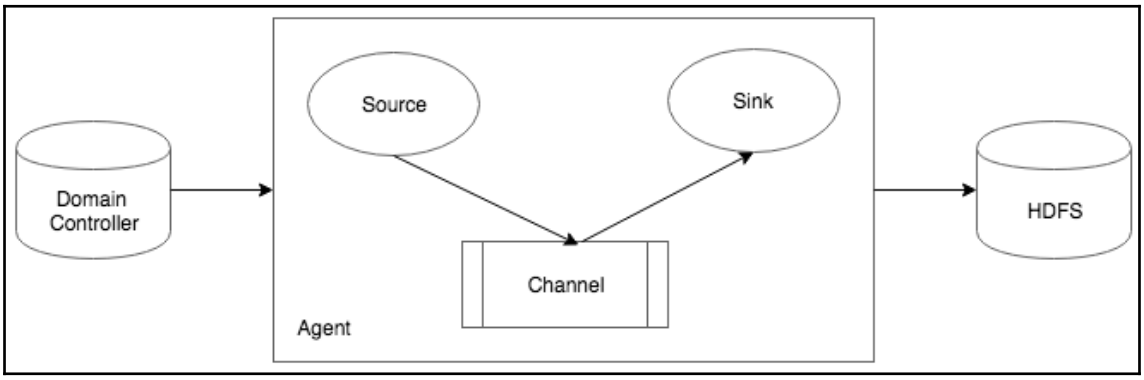
Service Information:
Service Name: krbtgt
Service ID: ACME\krbtgt

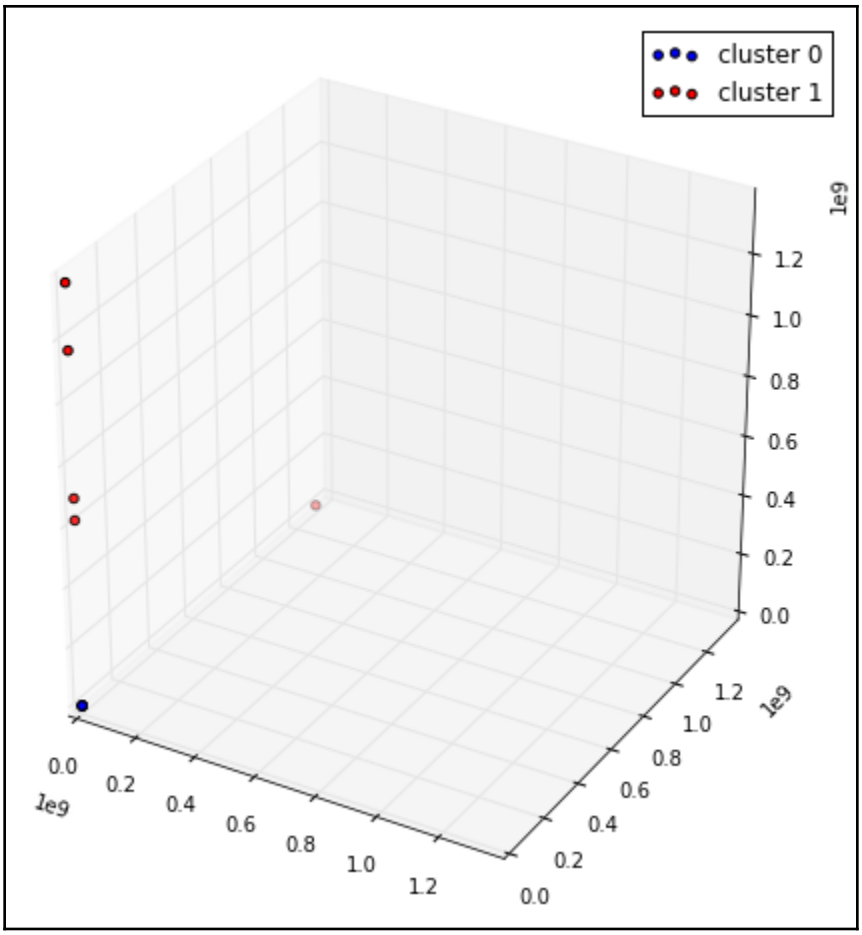
Network Information:
Client Address: ::1
Client Port: 0

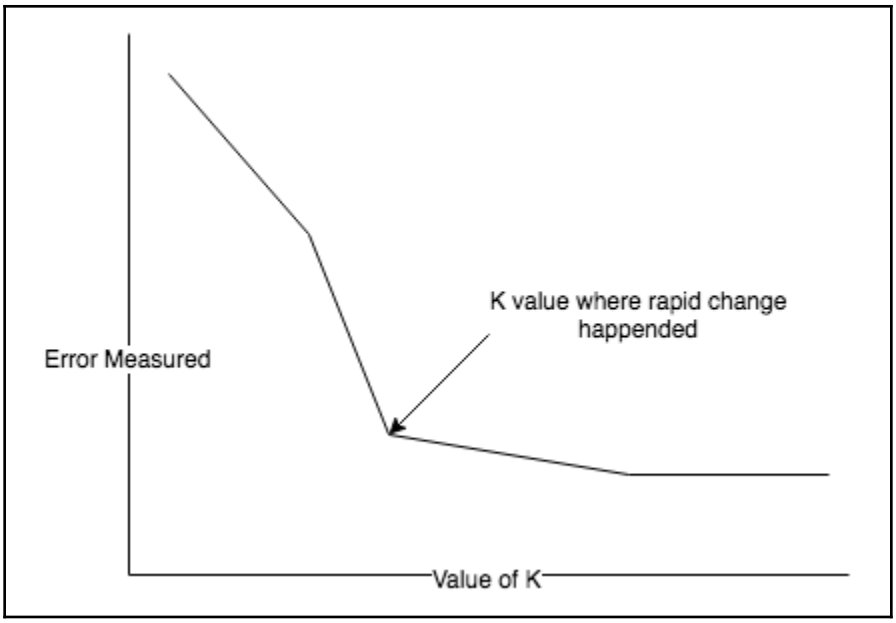
Additional Information:
Ticket Options: 0x40810010
Result Code: 0x0
Ticket Encryption Type: 0x17
Pre-Authentication Type: 2

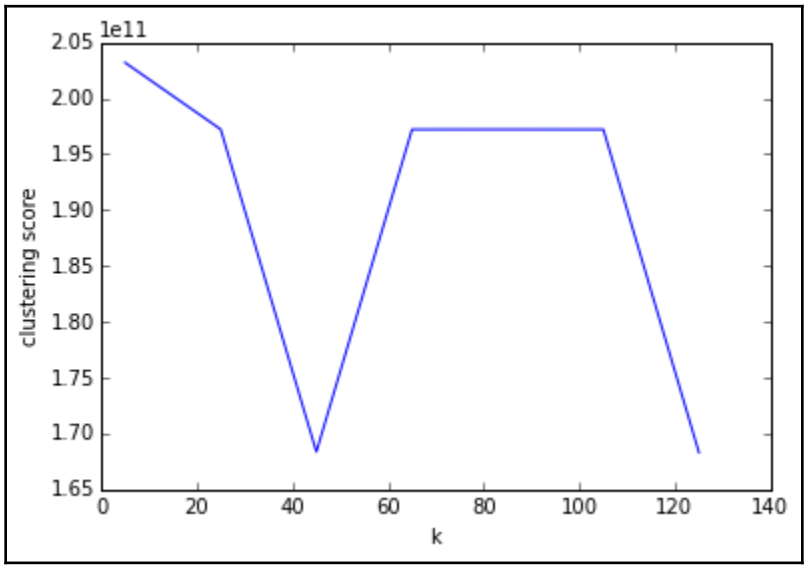
Certificate Information:
Certificate Issuer Name:
Certificate Serial Number:
Certificate Thumbprint:

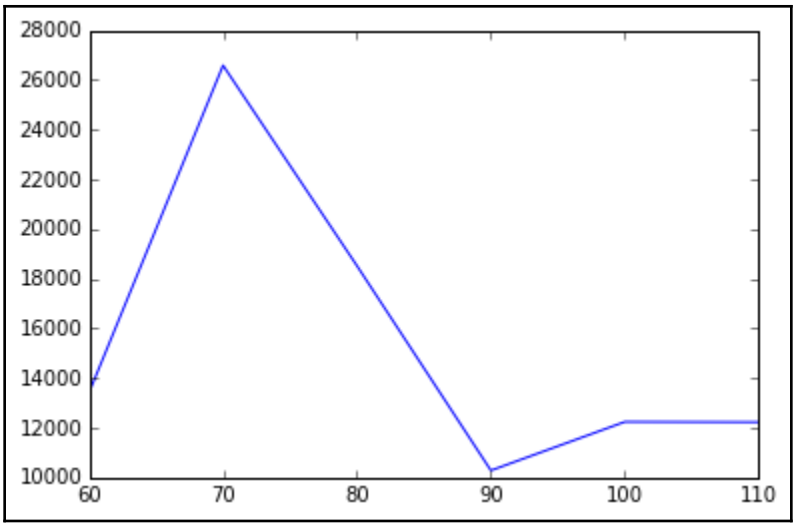
Certificate information is only provided if a certificate was used for pre-authentication.
Pre-
authentication types, ticket options, encryption types and result codes are defined in RFC 4120.

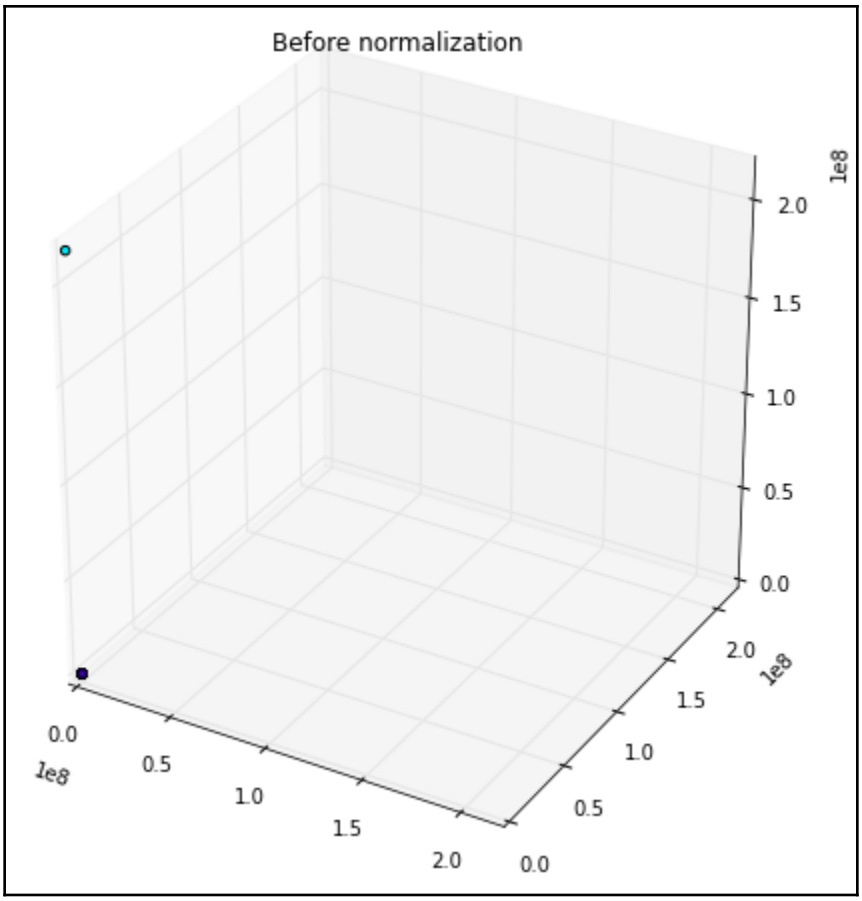


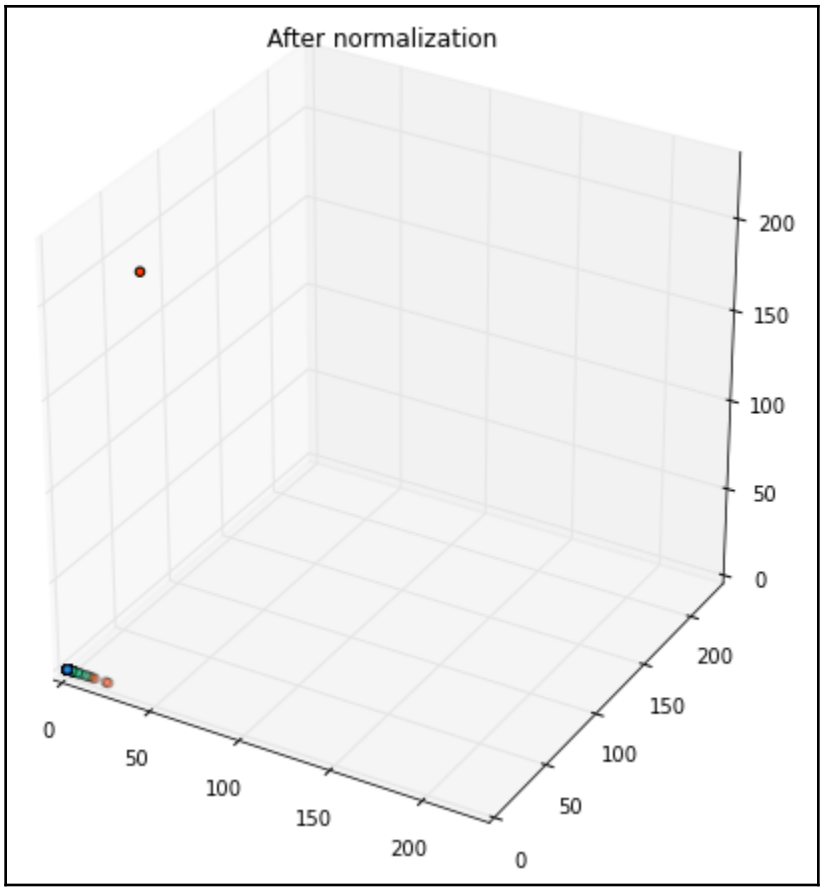

cluster 0
cluster 1
1e9
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1e9
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1e9
0.0
0.2
0.4
0.6
0.8
1.0
1.2

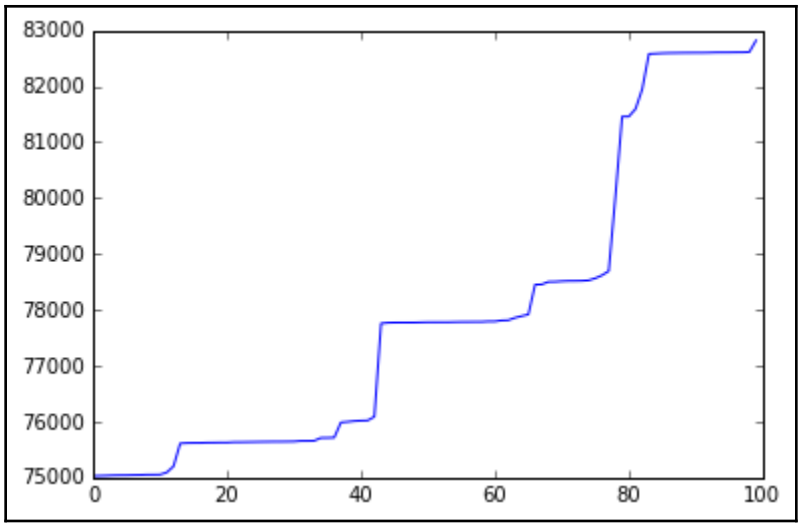

K value where rapid change happended
Error Measured
Value of K


1e11
2.05
2.00
1.95
1.90
1.85
1.80
1.75
1.70
1.65
clustering score
0
20
40
60
80
100
120
140
k


28000
26000
24000
22000
20000
18000
16000
14000
12000
10000
60
70
80
90
100
110

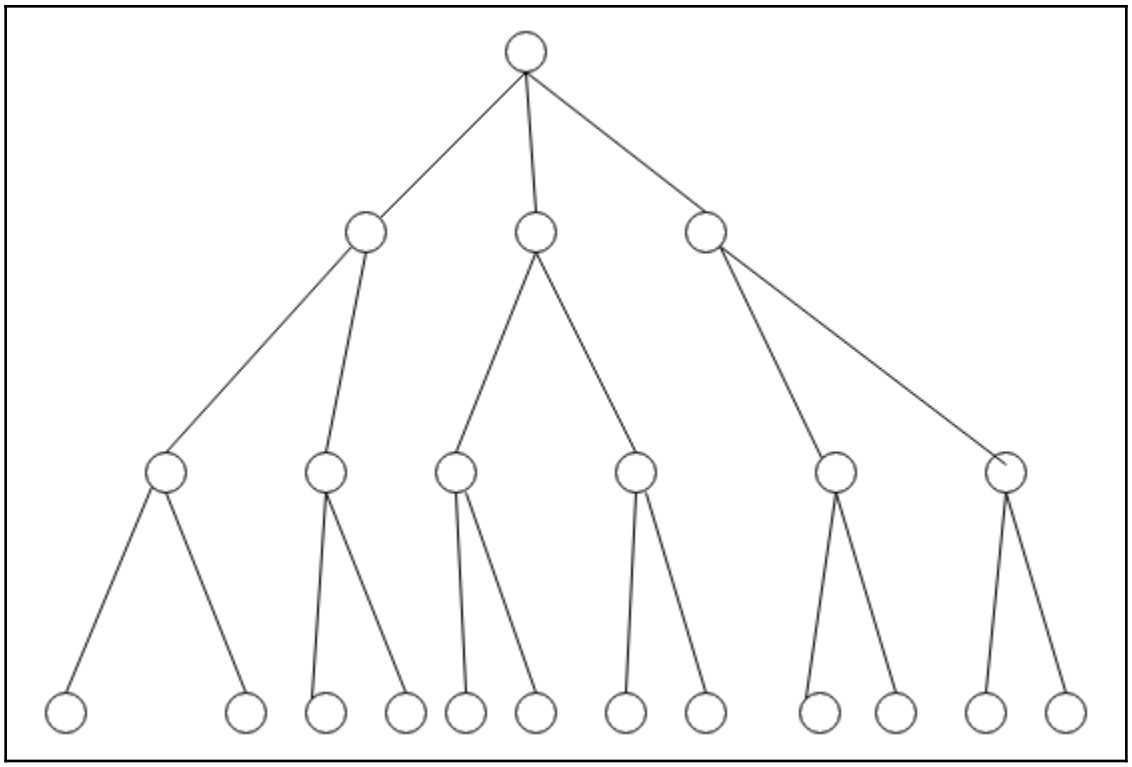

Before normalization
1e8
2.0
1.5
1.0
0.5
0.0
1e8
0.0
0.5
1.0
1.5
2.0
1e8
0.0
0.5
1.0
1.5
2.0

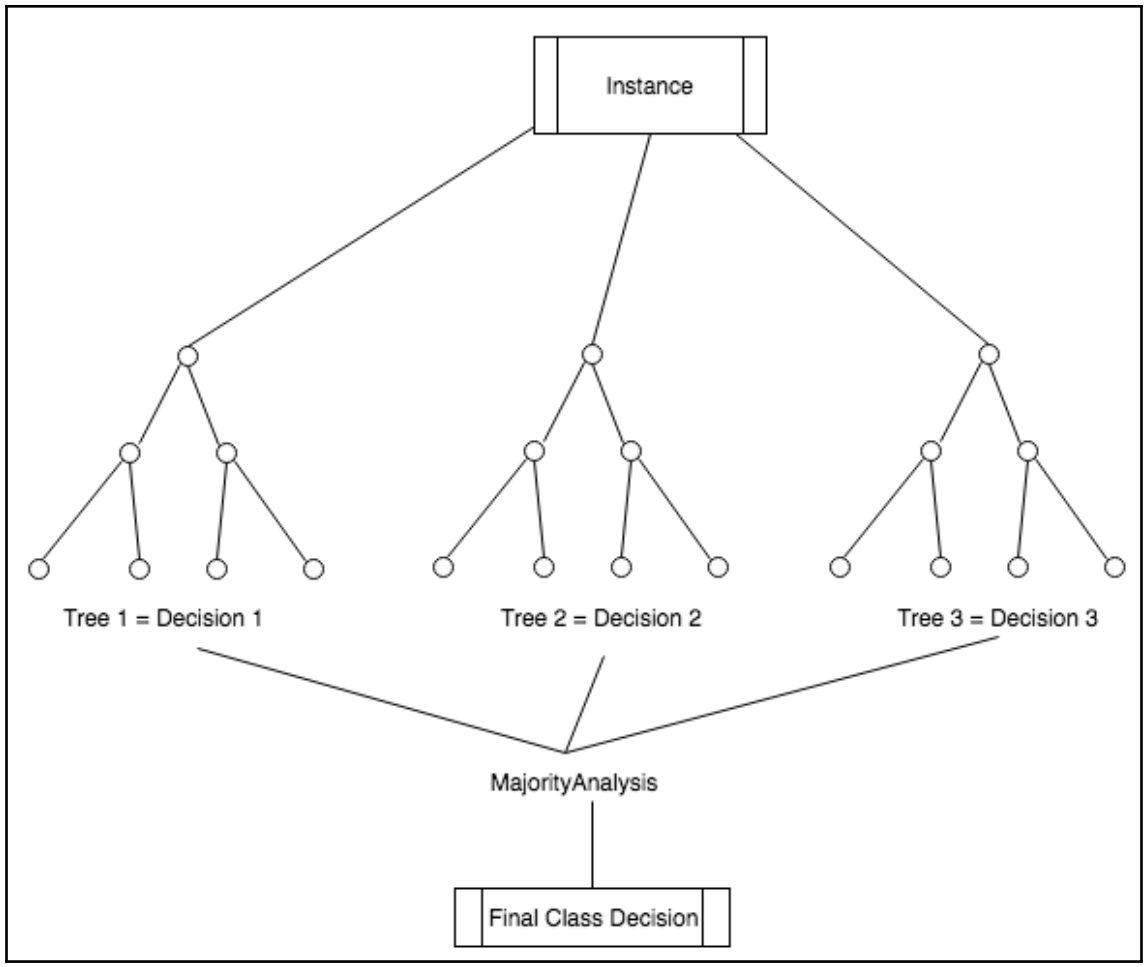

After normalization
200
150
100
50
0
200
150
100
50
0
0
50
100
150
200

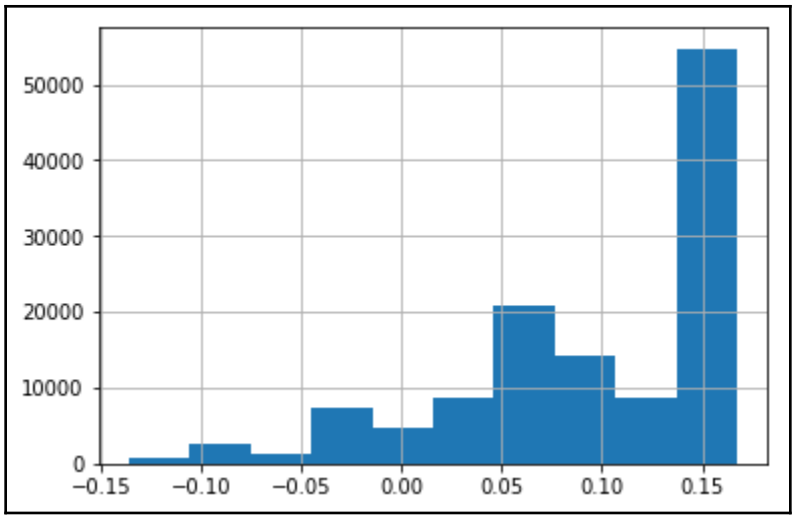

83000
82000
81000
80000
79000
78000
77000
76000
75000
0
20
40
60
80
100

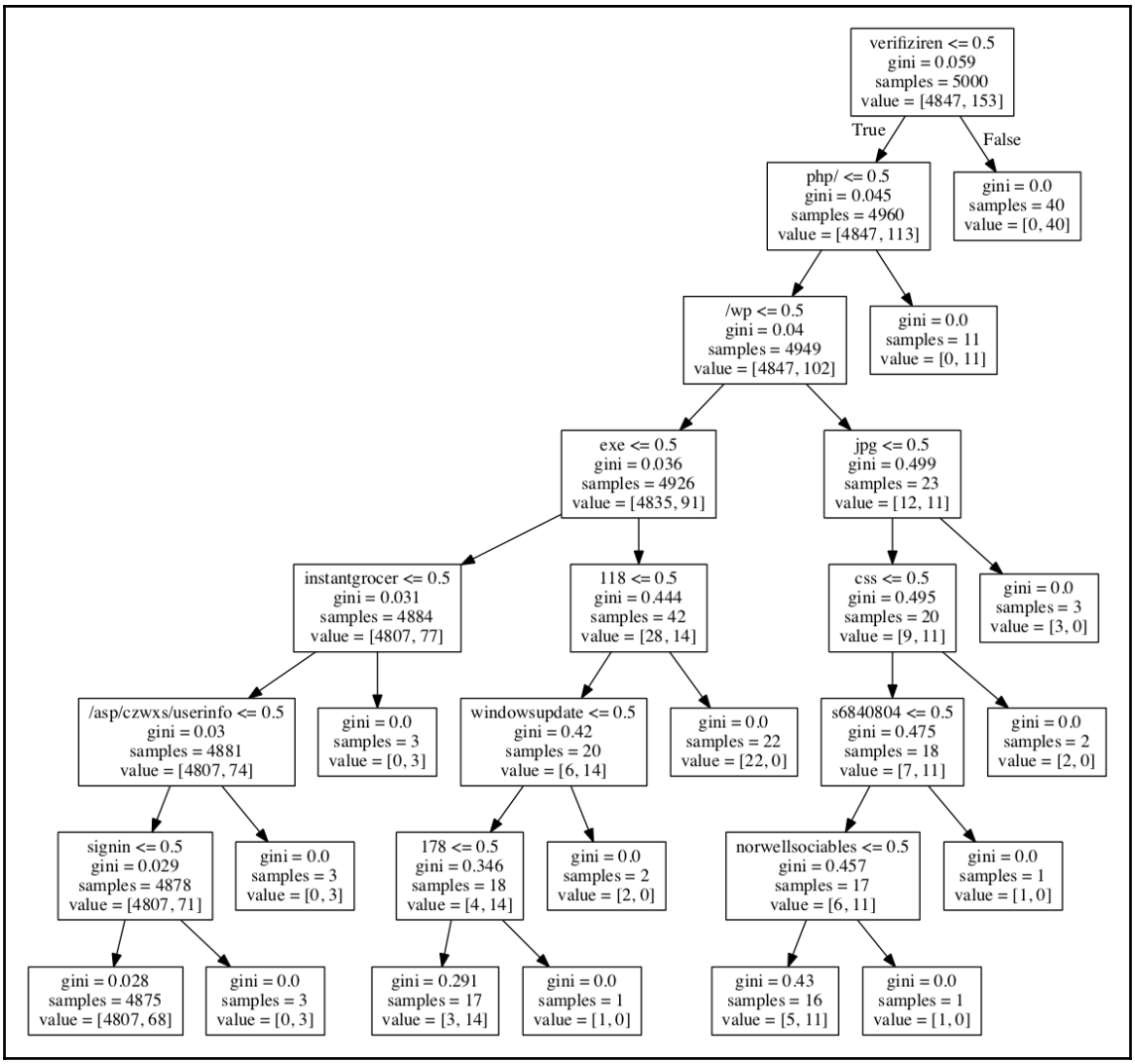
Chapter 7: Decision Tree and Context-Based Malicious Event Detection

```
13:08:05.737768 ppp0 > slip139-92-26-177.ist.tr.ibm.net.1221 > dsl-usw-cust-110.inetarena.com.www: . 342:342(0) ack 1449 win 31856 <nop
,nop,timestamp 1247771 114849487> (DF)
13:08:07.467571 ppp0 < dsl-usw-cust-110.inetarena.com.www > slip139-92-26-177.ist.tr.ibm.net.1221: . 1449:2897(1448) ack 342 win 31856
<nop,nop,timestamp 114849637 1247771> (DF)
13:08:07.707634 ppp0 < dsl-usw-cust-110.inetarena.com.www > slip139-92-26-177.ist.tr.ibm.net.1221: . 2897:4345(1448) ack 342 win 31856
<nop,nop,timestamp 114849637 1247771> (DF)
13:08:07.707922 ppp0 > slip139-92-26-177.ist.tr.ibm.net.1221 > dsl-usw-cust-110.inetarena.com.www: . 342:342(0) ack 4345 win 31856 <nop
,nop,timestamp 1247968 114849637> (DF)
13:08:08.057841 ppp0 > slip139-92-26-177.ist.tr.ibm.net.1045 > ns.de.ibm.net.domain: 8928+ PTR? 110.107.102.209.in-addr.arpa. (46)
13:08:08.747598 ppp0 < dsl-usw-cust-110.inetarena.com.www > slip139-92-26-177.ist.tr.ibm.net.1221: P 4345:5793(1448) ack 342 win 31856
<nop,nop,timestamp 114849813 1247968> (DF)
13:08:08.847870 ppp0 < dsl-usw-cust-110.inetarena.com.www > slip139-92-26-177.ist.tr.ibm.net.1221: FP 5793:6297(504) ack 342 win 31856
<nop,nop,timestamp 114849813 1247968> (DF)
13:08:08.848063 ppp0 > slip139-92-26-177.ist.tr.ibm.net.1221 > dsl-usw-cust-110.inetarena.com.www: . 342:342(0) ack 6298 win 31856 <nop
,nop,timestamp 1248082 114849813> (DF)
13:08:08.907566 ppp0 < ns.de.ibm.net.domain > slip139-92-26-177.ist.tr.ibm.net.1045: 8928* 3/1/1 PTR dsl-usw-cust-110.inetarena.com., P
TR fingerless.or (199)
13:08:09.151742 ppp0 > slip139-92-26-177.ist.tr.ibm.net.1221 > dsl-usw-cust-110.inetarena.com.www: F 342:342(0) ack 6298 win 31856 <nop
,nop,timestamp 1248112 114849813> (DF)
13:08:10.137603 ppp0 < dsl-usw-cust-110.inetarena.com.www > slip139-92-26-177.ist.tr.ibm.net.1221: . 6298:6298(0) ack 343 win 31856 <no
p,nop,timestamp 114849967 1248112> (DF)
13:09:01.984210 ppp0 > slip139-92-26-177.ist.tr.ibm.net.1222 > dsl-usw-cust-110.inetarena.com.www: S 920197285:920197285(0) win 32120 <
mss 1460,sackOK,timestamp 1253395 0,nop,wscale 0> (DF)
13:09:03.097569 ppp0 < dsl-usw-cust-110.inetarena.com.www > slip139-92-26-177.ist.tr.ibm.net.1222: S 1222277738:1222277738(0) ack 92019
7286 win 32120 <mss 1460,sackOK,timestamp 114855252 1253395,nop,wscale 0> (DF)
13:09:03.098197 ppp0 > slip139-92-26-177.ist.tr.ibm.net.1222 > dsl-usw-cust-110.inetarena.com.www: . 1:1(0) ack 1 win 32120 <nop,nop,ti
mestamp 1253507 114855252> (DF)
13:09:03.102171 ppp0 > slip139-92-26-177.ist.tr.ibm.net.1222 > dsl-usw-cust-110.inetarena.com.www: P 1:322(321) ack 1 win 32120 <nop,no
p,timestamp 1253507 114855252> (DF)
13:09:04.147613 ppp0 < dsl-usw-cust-110.inetarena.com.www > slip139-92-26-177.ist.tr.ibm.net.1222: . 1:1(0) ack 322 win 31856 <nop,nop,
timestamp 114855369 1253507> (DF)
13:09:04.507608 ppp0 < dsl-usw-cust-110.inetarena.com.www > slip139-92-26-177.ist.tr.ibm.net.1222: . 1:1449(1448) ack 322 win 31856 <no
p,nop,timestamp 114855369 1253507> (DF)
13:09:04.507934 ppp0 > slip139-92-26-177.ist.tr.ibm.net.1222 > dsl-usw-cust-110.inetarena.com.www: . 322:322(0) ack 1449 win 31856 <nop
,nop,timestamp 1253648 114855369> (DF)
13:09:05.627604 ppp0 < dsl-usw-cust-110.inetarena.com.www > slip139-92-26-177.ist.tr.ibm.net.1222: . 1449:2897(1448) ack 322 win 31856
<nop,nop,timestamp 114855491 1253648> (DF)
13:09:05.857649 ppp0 < dsl-usw-cust-110.inetarena.com.www > slip139-92-26-177.ist.tr.ibm.net.1222: . 2897:4345(1448) ack 322 win 31856
<nop,nop,timestamp 114855491 1253648> (DF)
13:09:05.857918 ppp0 > slip139-92-26-177.ist.tr.ibm.net.1222 > dsl-usw-cust-110.inetarena.com.www: . 322:322(0) ack 4345 win 31856 <nop
,nop,timestamp 1253783 114855491> (DF)
13:09:06.907557 ppp0 < dsl-usw-cust-110.inetarena.com.www > slip139-92-26-177.ist.tr.ibm.net.1222: FP 4345:5792(1447) ack 322 win 31856
 <nop,nop,timestamp 114855627 1253783> (DF)
13:09:06.907887 ppp0 > slip139-92-26-177.ist.tr.ibm.net.1222 > dsl-usw-cust-110.inetarena.com.www: . 322:322(0) ack 5793 win 31856 <nop
,nop,timestamp 1253888 114855627> (DF)
13:09:07.401205 ppp0 > slip139-92-26-177.ist.tr.ibm.net.1222 > dsl-usw-cust-110.inetarena.com.www: F 322:322(0) ack 5793 win 31856 <nop
,nop,timestamp 1253937 114855627> (DF)
13:09:08.317623 ppp0 < dsl-usw-cust-110.inetarena.com.www > slip139-92-26-177.ist.tr.ibm.net.1222: . 5793:5793(0) ack 323 win 31856 <no
p,nop,timestamp 114855780 1253937> (DF)
```

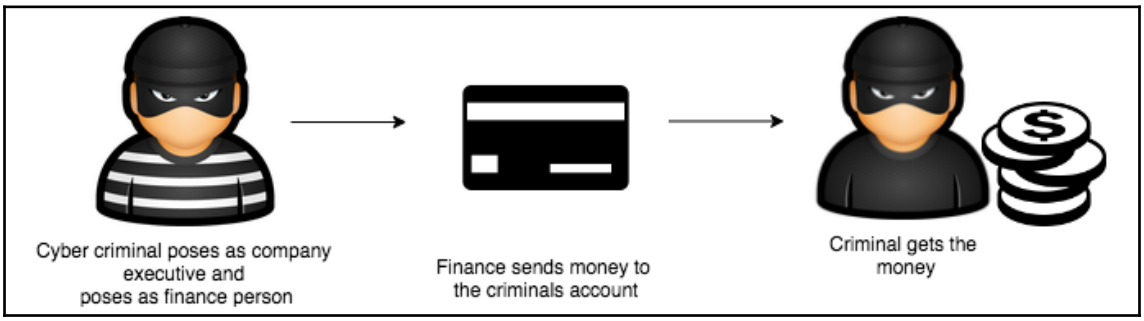



Instance
Tree 1 = Decision 1
Tree 2 = Decision 2
Tree 3 = Decision 3
MajorityAnalysis
Final Class Decision


50000
40000
30000
20000
10000
0
−0.15
−0.10
−0.05
0.00
0.05
0.10
0.15

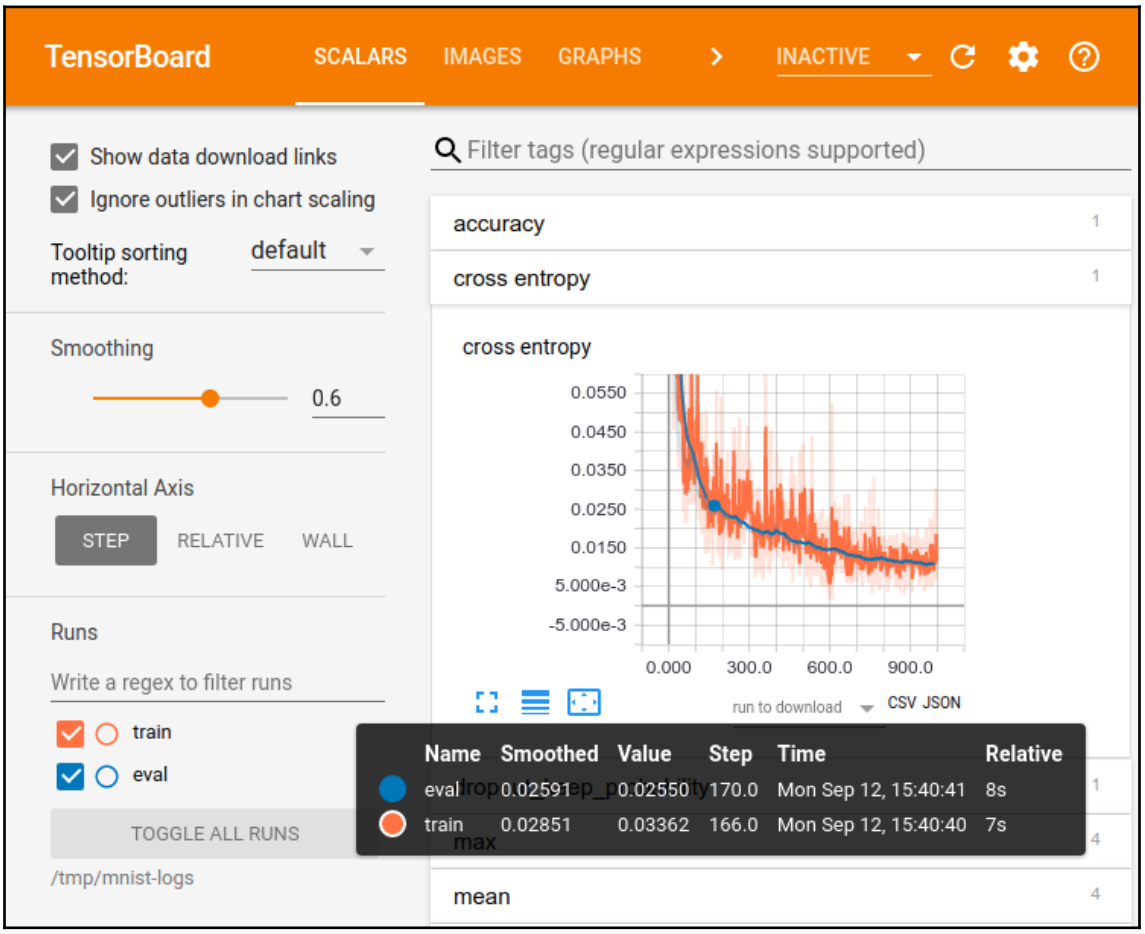

verifiziren <= 0.5
gini = 0.059
samples = 5000
value = [4847, 153]
True
False
php/ <= 0.5
gini = 0.045
samples = 4960
value = [4847, 113]
gini = 0.0
samples = 40
value = [0, 40]
/wp <= 0.5
gini = 0.04
samples = 4949
value = [4847, 102]
gini = 0.0
samples = 11
value = [0, 11]
exe <= 0.5
gini = 0.036
samples = 4926
value = [4835, 91]
jpg <= 0.5
gini = 0.499
samples = 23
value = [12, 11]
instantgrocer <= 0.5
gini = 0.031
samples = 4884
value = [4807, 77]
118 <= 0.5
gini = 0.444
samples = 42
value = [28, 14]
css <= 0.5
gini = 0.495
samples = 20
value = [9, 11]
gini = 0.0
samples = 3
value = [3, 0]
/asp/czwxs/userinfo <= 0.5
gini = 0.03
samples = 4881
value = [4807, 74]
gini = 0.0
samples = 3
value = [0, 3]
windowsupdate <= 0.5
gini = 0.42
samples = 20
value = [6, 14]
gini = 0.0
samples = 22
value = [22, 0]
s6840804 <= 0.5
gini = 0.475
samples = 18
value = [7, 11]
gini = 0.0
samples = 2
value = [2, 0]
signin <= 0.5
gini = 0.029
samples = 4878
value = [4807, 71]
gini = 0.0
samples = 3
value = [0, 3]
178 <= 0.5
gini = 0.346
samples = 18
value = [4, 14]
gini = 0.0
samples = 2
value = [2, 0]
norwellsociables <= 0.5
gini = 0.457
samples = 17
value = [6, 11]
gini = 0.0
samples = 1
value = [1, 0]
gini = 0.028
samples = 4875
value = [4807, 68]
gini = 0.0
samples = 3
value = [0, 3]
gini = 0.291
samples = 17
value = [3, 14]
gini = 0.0
samples = 1
value = [1, 0]
gini = 0.43
samples = 16
value = [5, 11]
gini = 0.0
samples = 1
value = [1, 0]

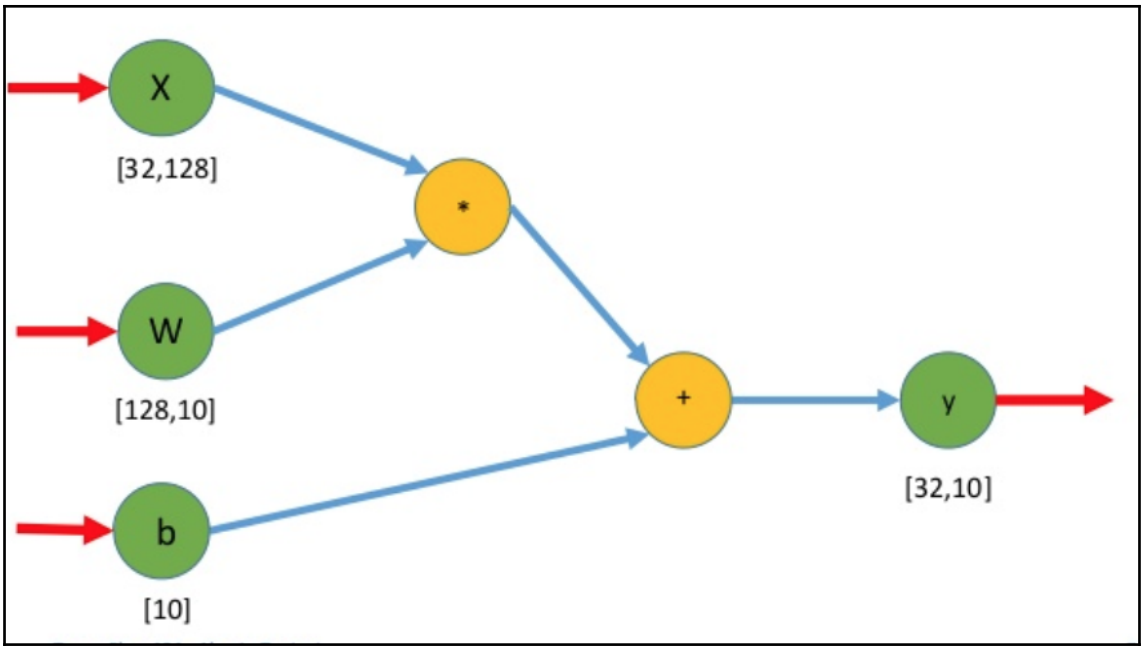
Chapter 8: Catching Impersonators and Hackers Red Handed



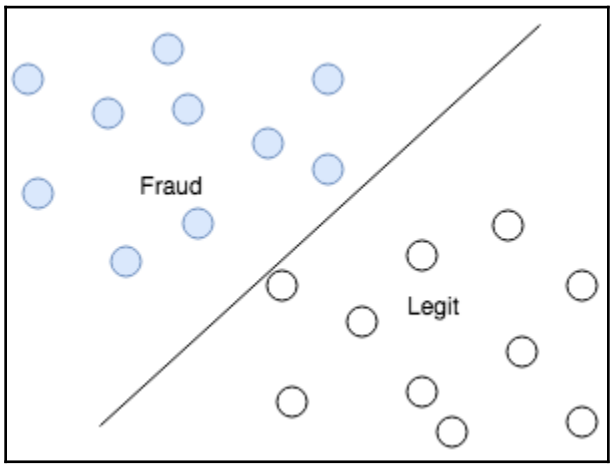
$$\operatorname{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \operatorname{lev}_{a,b}(i-1,j) + 1 \\ \operatorname{lev}_{a,b}(i,j-1) + 1 \\ \operatorname{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

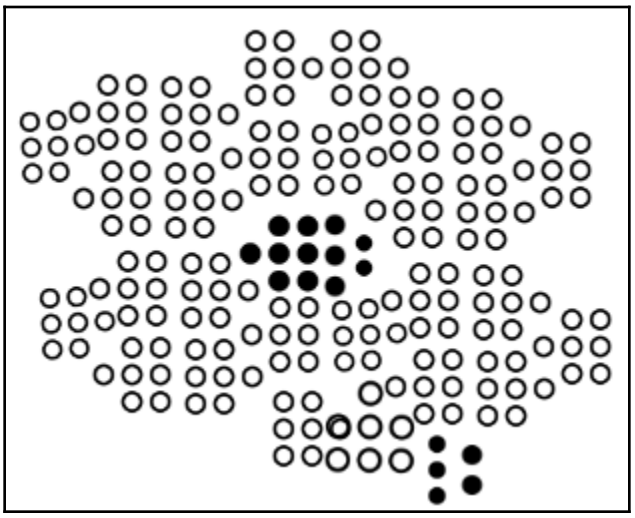
Chapter 9: Changing the Game with TensorFlow

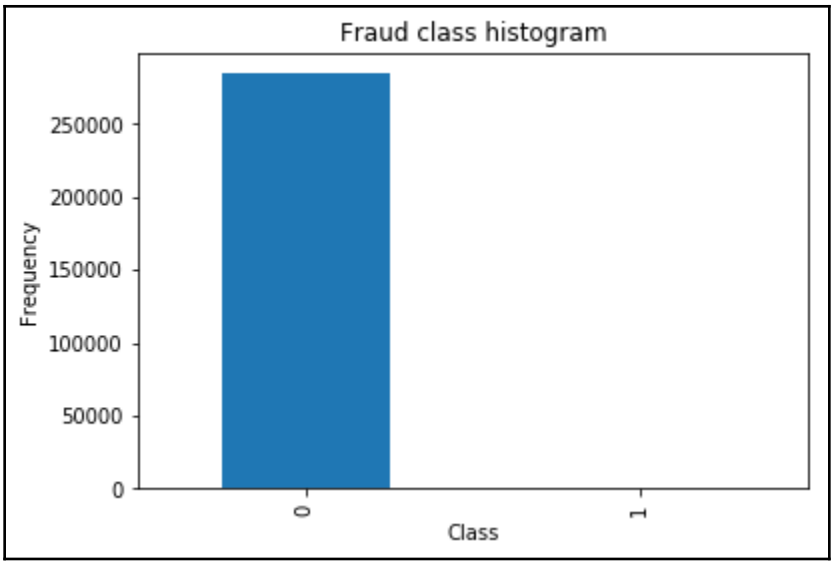


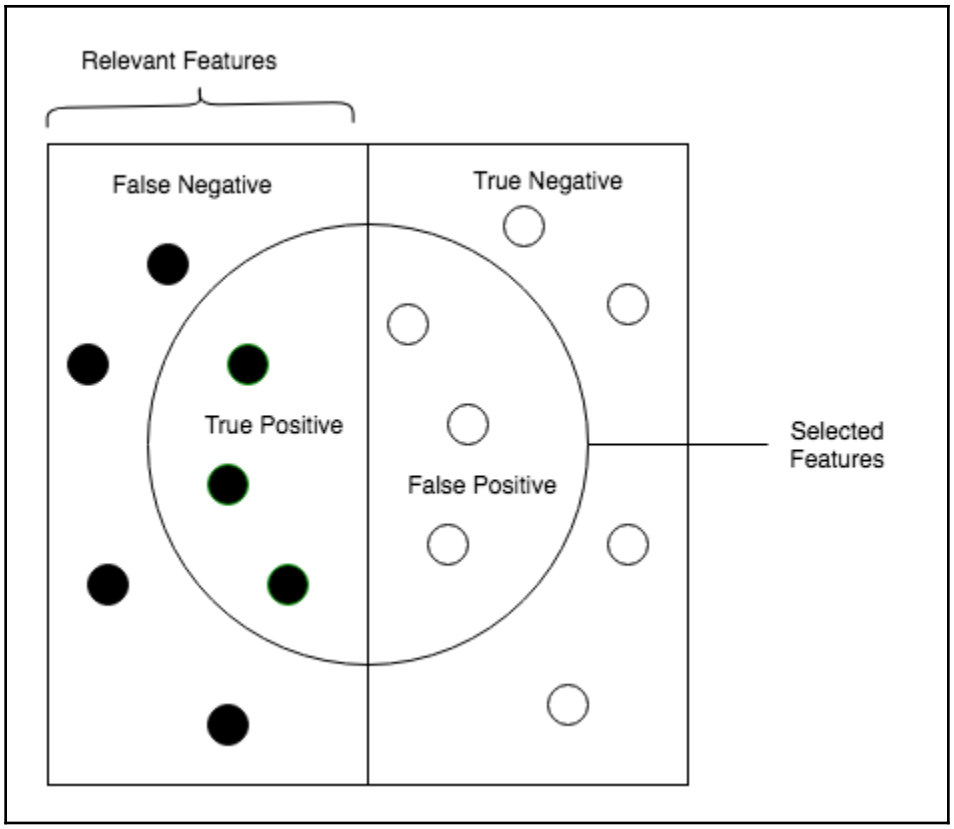

X
[32,128]
*
W
[128,10]
+
y
[32,10]
b
[10]

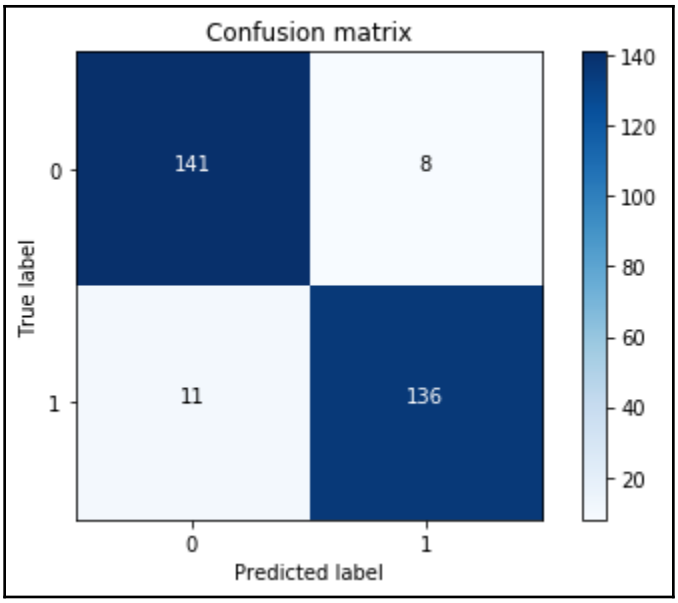
Chapter 10: Financial Fraud and How Deep Learning Can Mitigate It

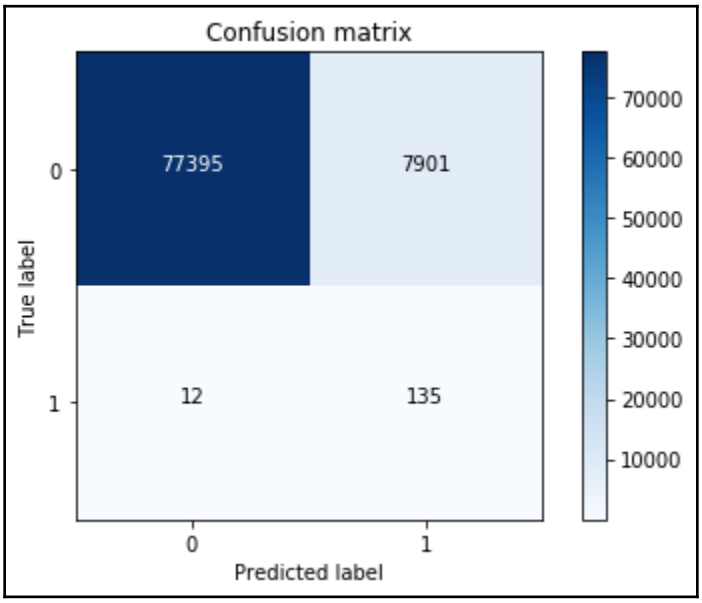


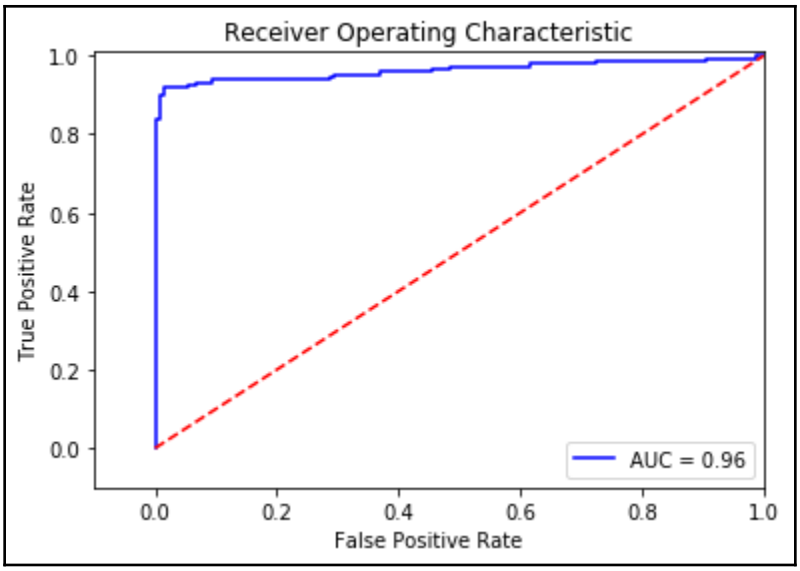


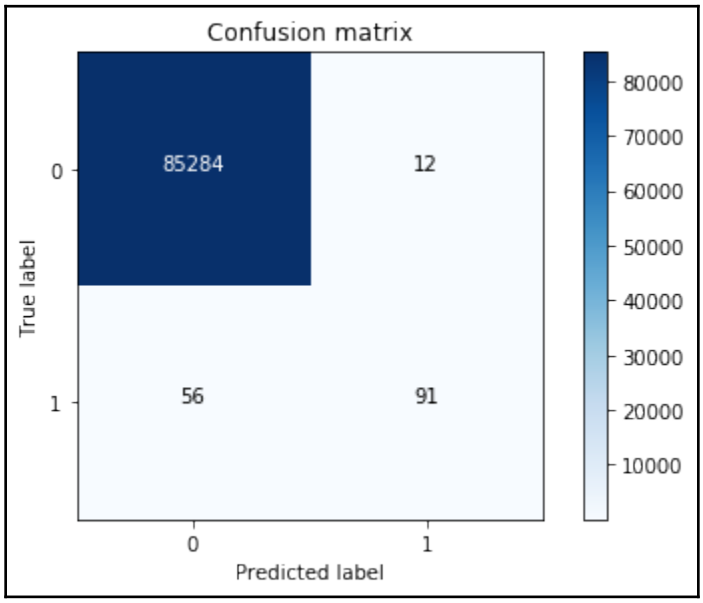

Fraud class histogram
250000
200000
150000
100000
50000
0
Frequency
0
1
Class

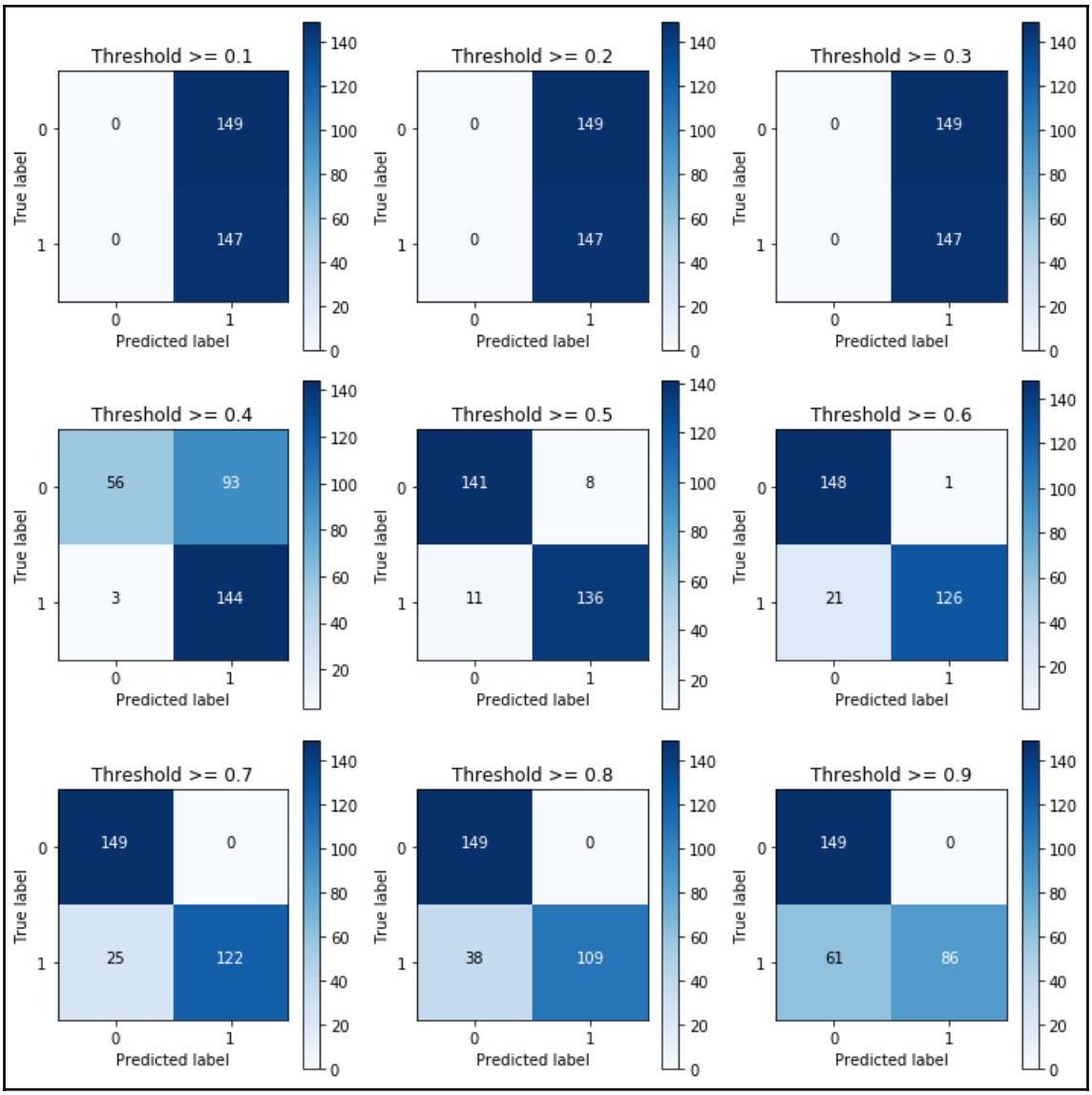

Relevant Features
False Negative
True Negative
True Positive
False Positive
Selected
Features

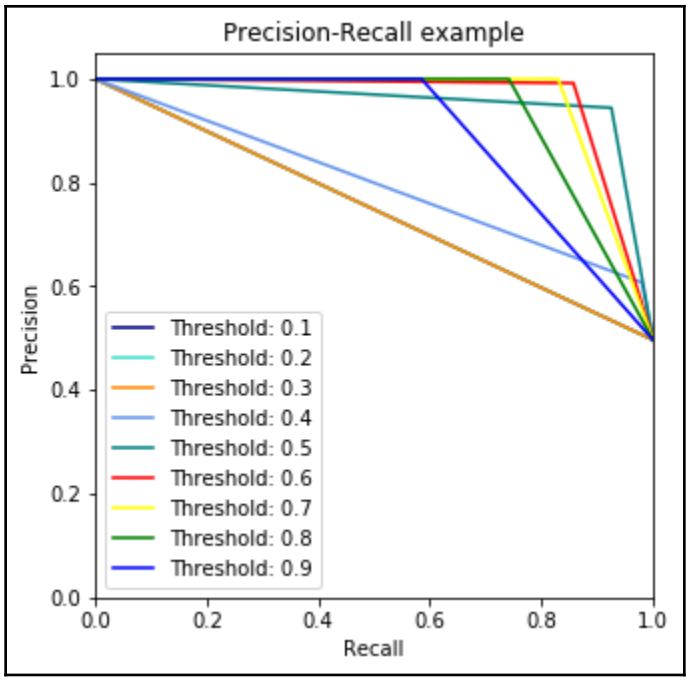

Confusion matrix
141
8
11
136
0
1
True label
Predicted label
140
120
100
80
60
40
20

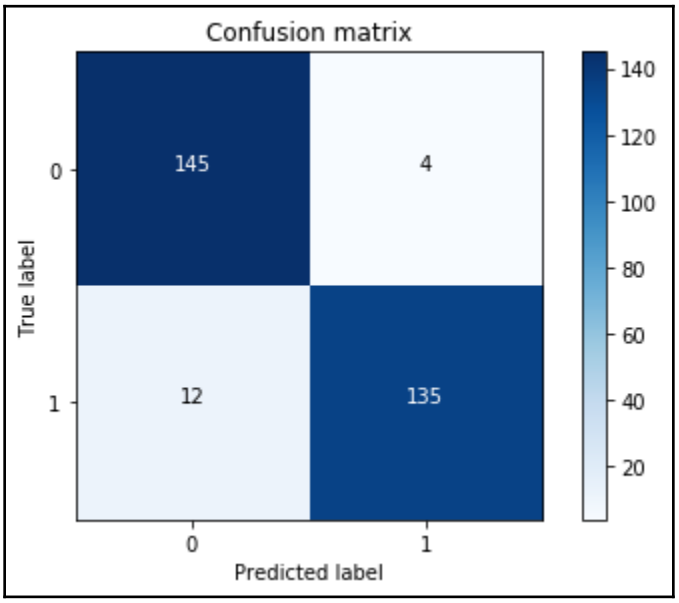

Confusion matrix
77395
7901
12
135
0
1
True label
Predicted label
70000
60000
50000
40000
30000
20000
10000

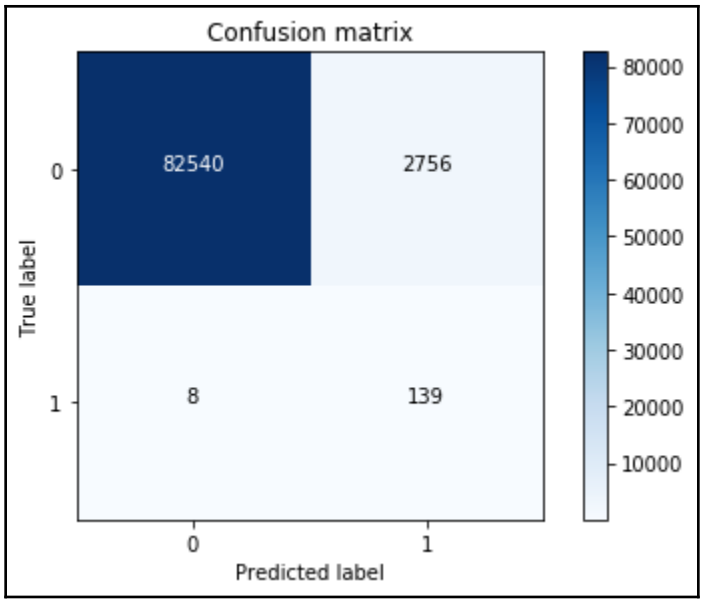
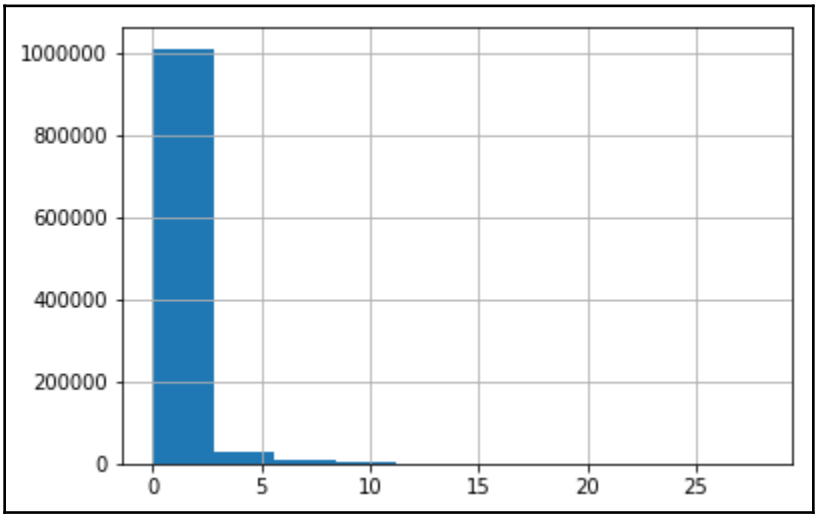
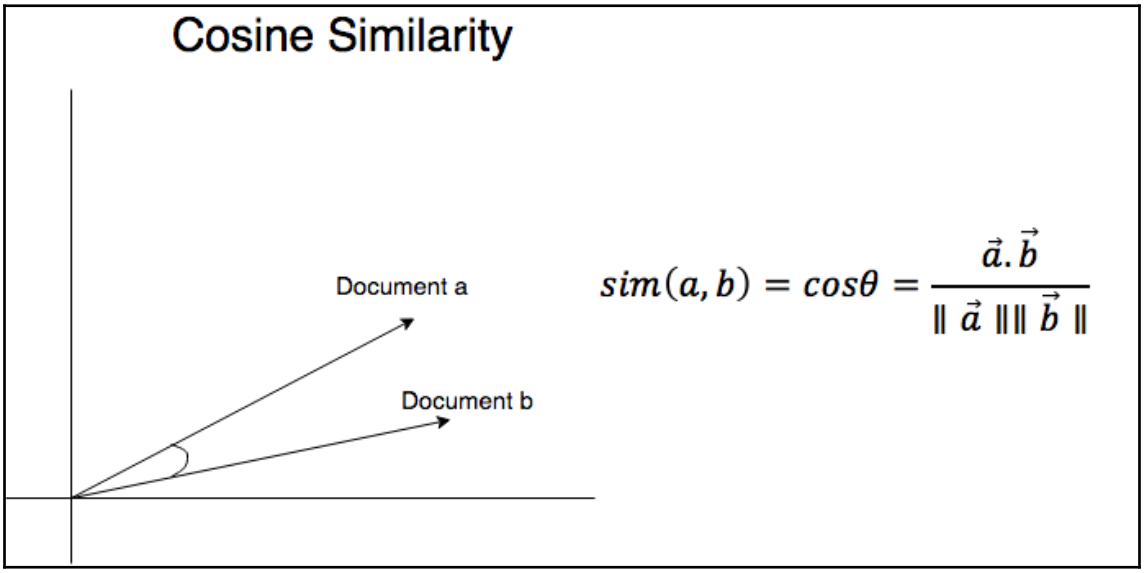

Receiver Operating Characteristic
True Positive Rate
False Positive Rate
AUC = 0.96
0.0
0.2
0.4
0.6
0.8
1.0


Confusion matrix
85284
12
56
91
True label
Predicted label
0
1
80000
70000
60000
50000
40000
30000
20000
10000


Threshold >= 0.1
0
149
0
147
True label
Predicted label
Threshold >= 0.2
0
149
0
147
Threshold >= 0.3
0
149
0
147
Threshold >= 0.4
56
93
3
144
Threshold >= 0.5
141
8
11
136
Threshold >= 0.6
148
1
21
126
Threshold >= 0.7
149
0
25
122
Threshold >= 0.8
149
0
38
109
Threshold >= 0.9
149
0
61
86


Precision-Recall example
Precision
Recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Threshold: 0.1
Threshold: 0.2
Threshold: 0.3
Threshold: 0.4
Threshold: 0.5
Threshold: 0.6
Threshold: 0.7
Threshold: 0.8
Threshold: 0.9


Confusion matrix
True label
Predicted label
145
4
12
135
0
1
0
1
140
120
100
80
60
40
20


Confusion matrix
82540
2756
8
139
0
1
True label
0
1
Predicted label
80000
70000
60000
50000
40000
30000
20000
10000

Chapter 11: Case Studies