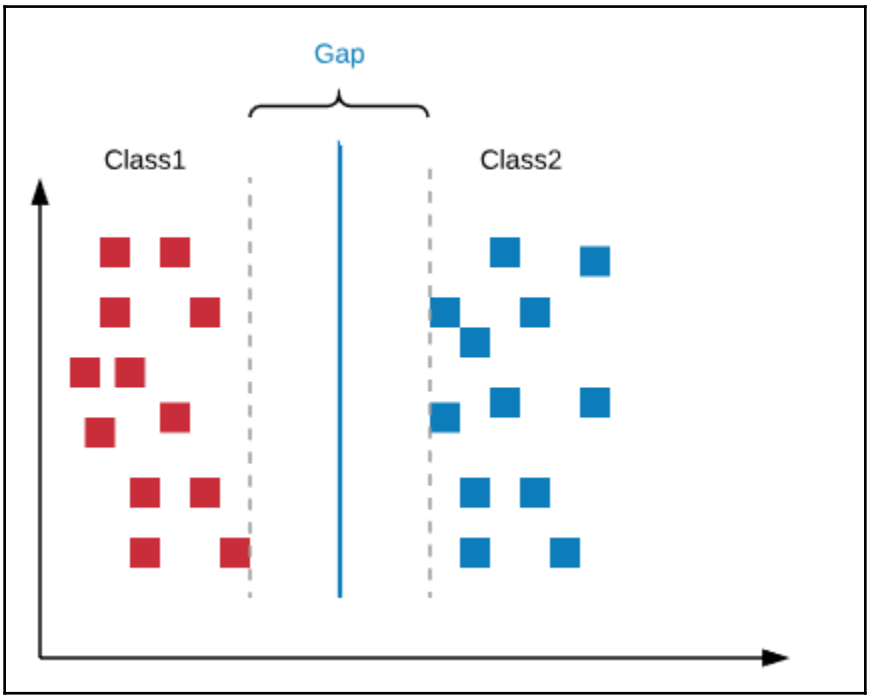
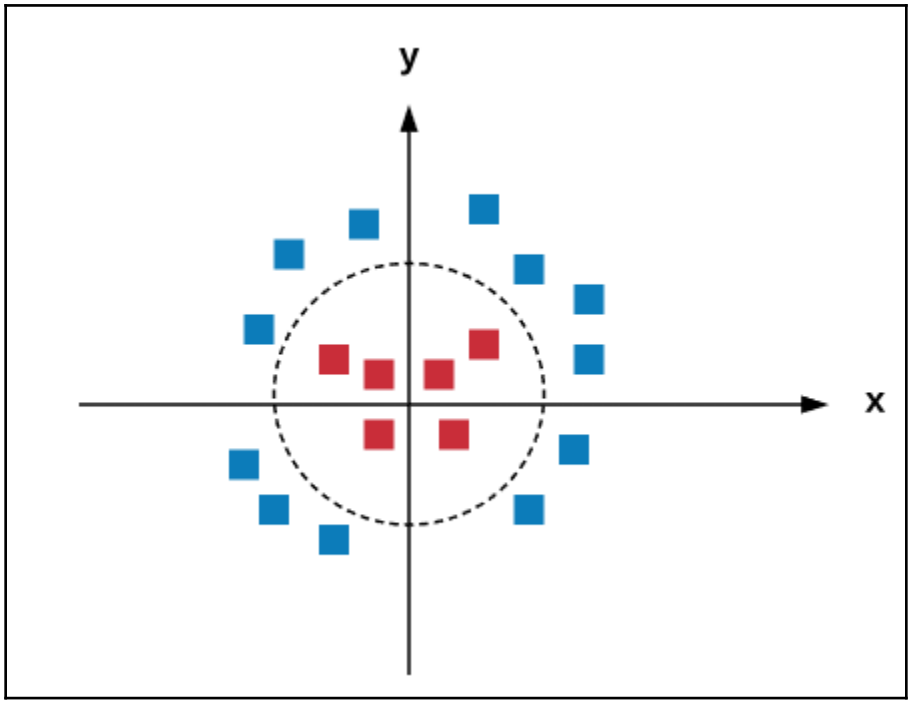
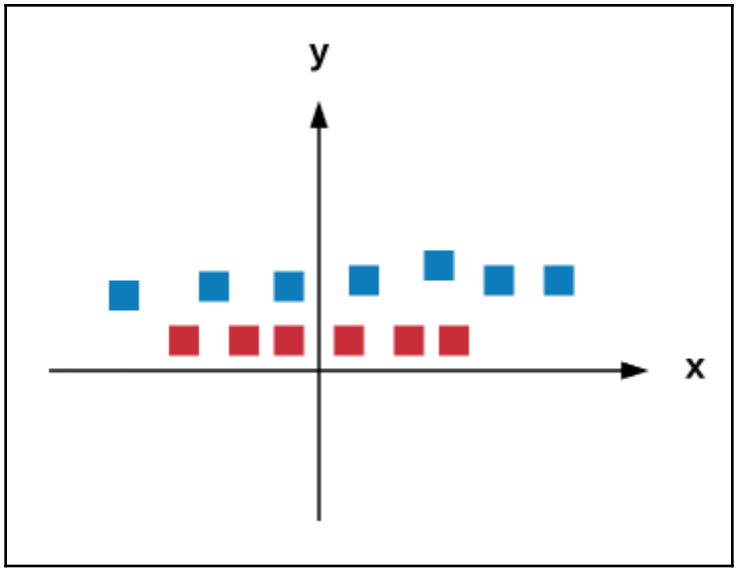
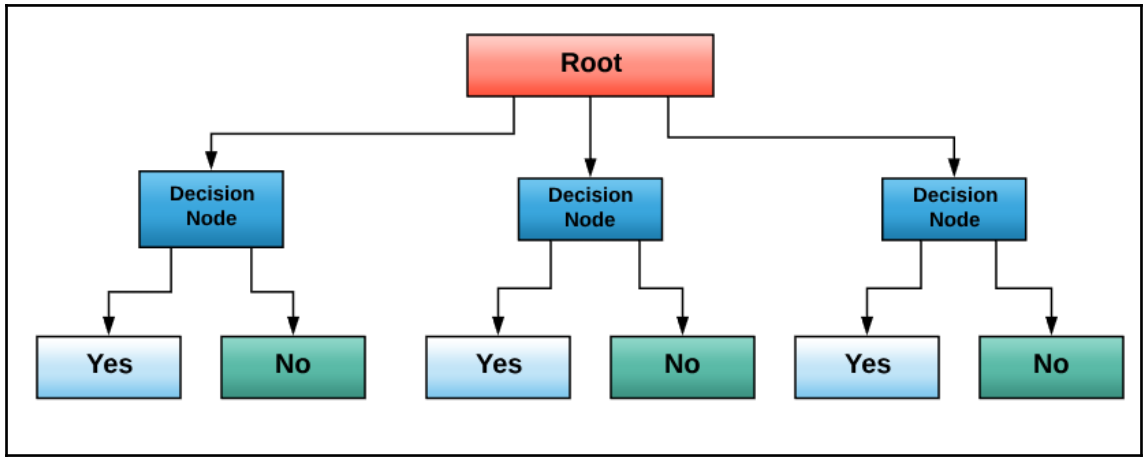
Chapter 1: Introduction to Machine Learning in Pentesting

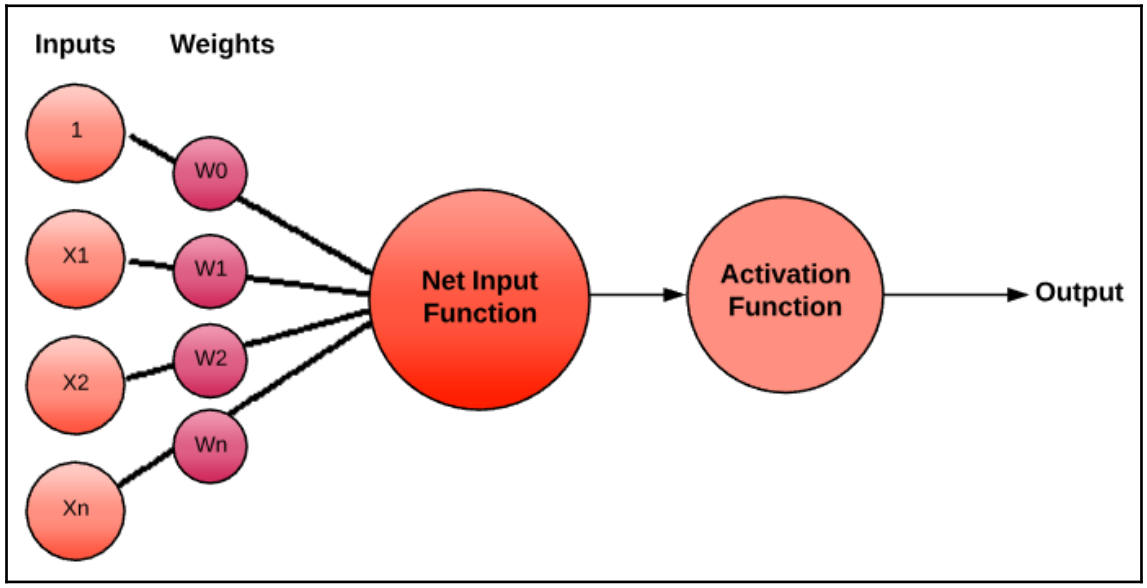
$$P(c|x) = \frac{P(x|c) \times P(c)}{P(x)}$$




y
x


y
x

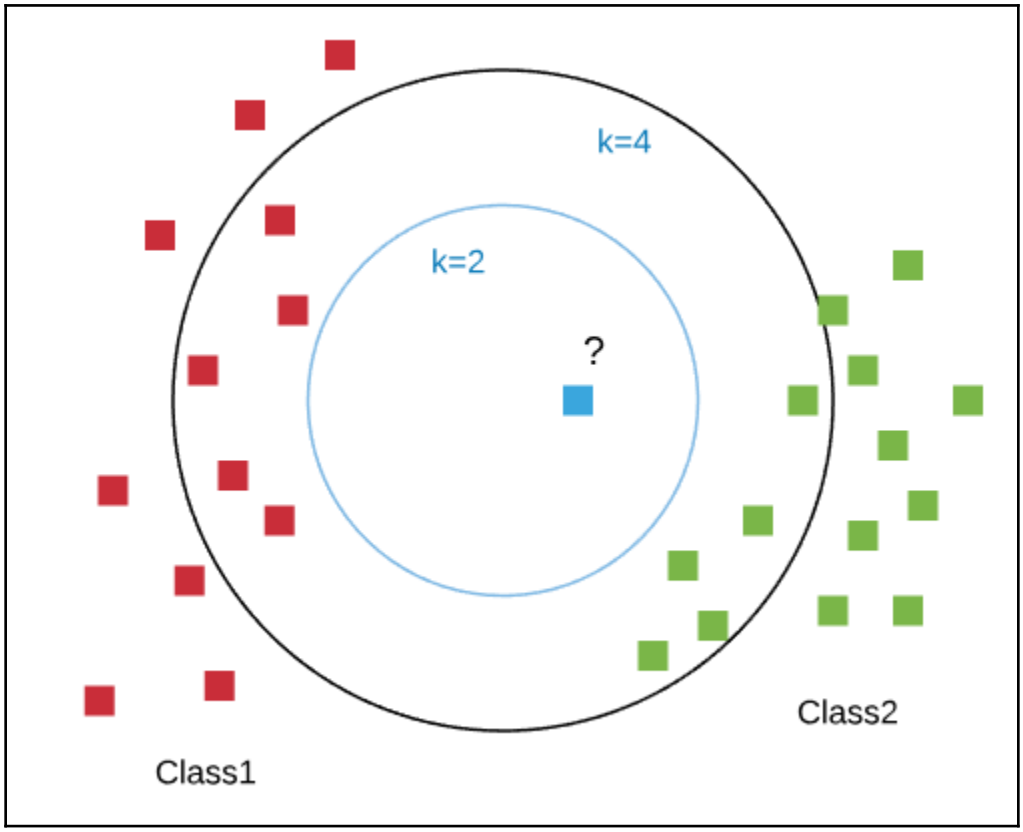

Root
Decision Node
Decision Node
Decision Node
Yes
No
Yes
No
Yes
No


Inputs
Weights
1
X1
X2
Xn
W0
W1
W2
Wn
Net Input Function
Activation Function
Output

$$f(x) = \frac{1}{1 + e^{-x}}$$

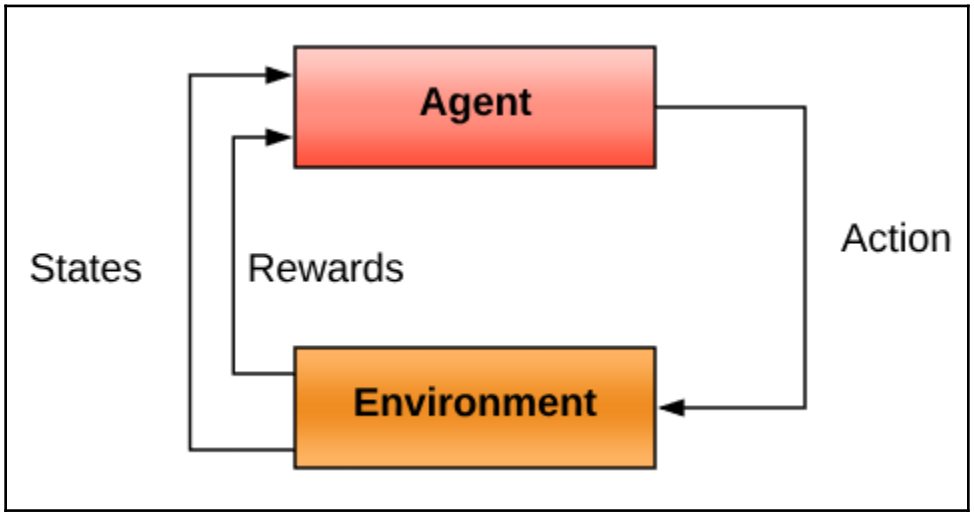
$$f(x) = \frac{2}{1 + e^{-2x}} - 1$$

$$f(x) = \begin{cases} 0 for x < 0 \\ x for x >= 0 \end{cases}$$



$$E(x, y) = \sqrt{\sum_{i=0}^{n} (x_i - y_i)^2}$$

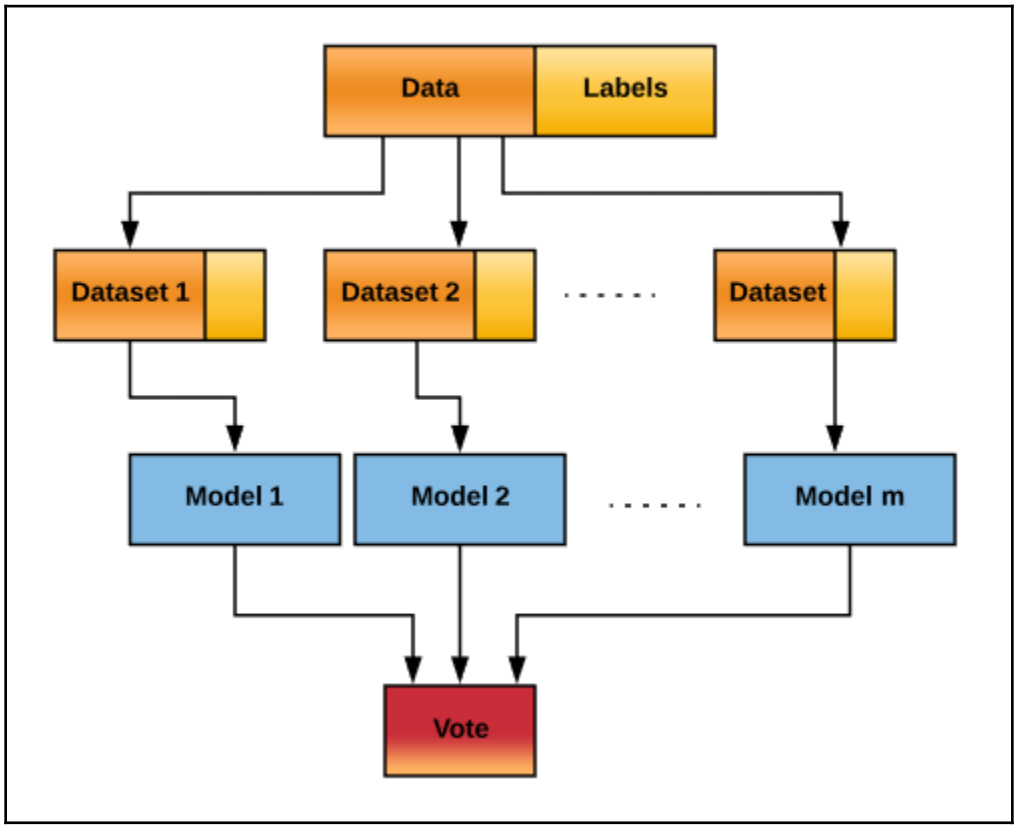
$$Similarity = \cos(\varnothing) = \frac{A.B}{\|A\| \times \|B\|}$$

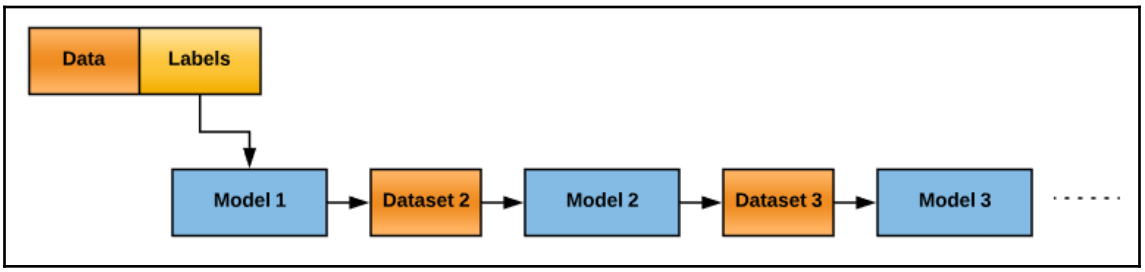


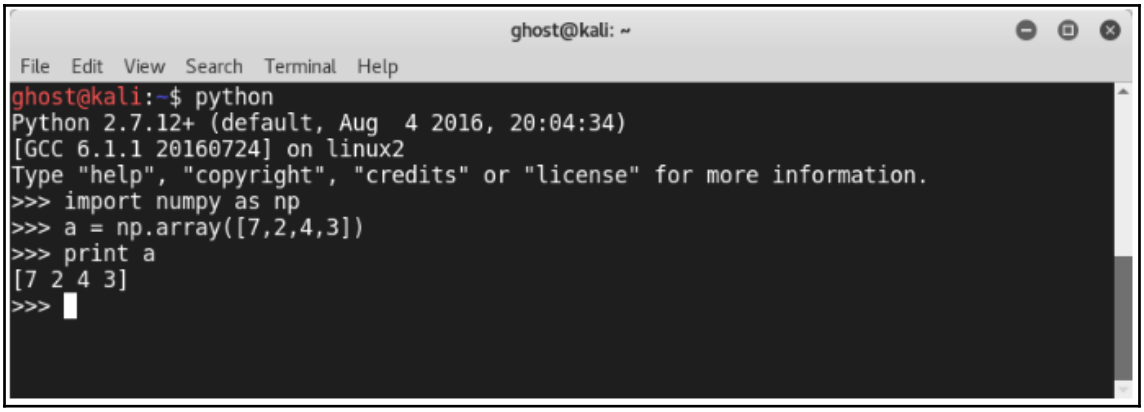
$$precision = \frac{tp}{tp + fp}$$

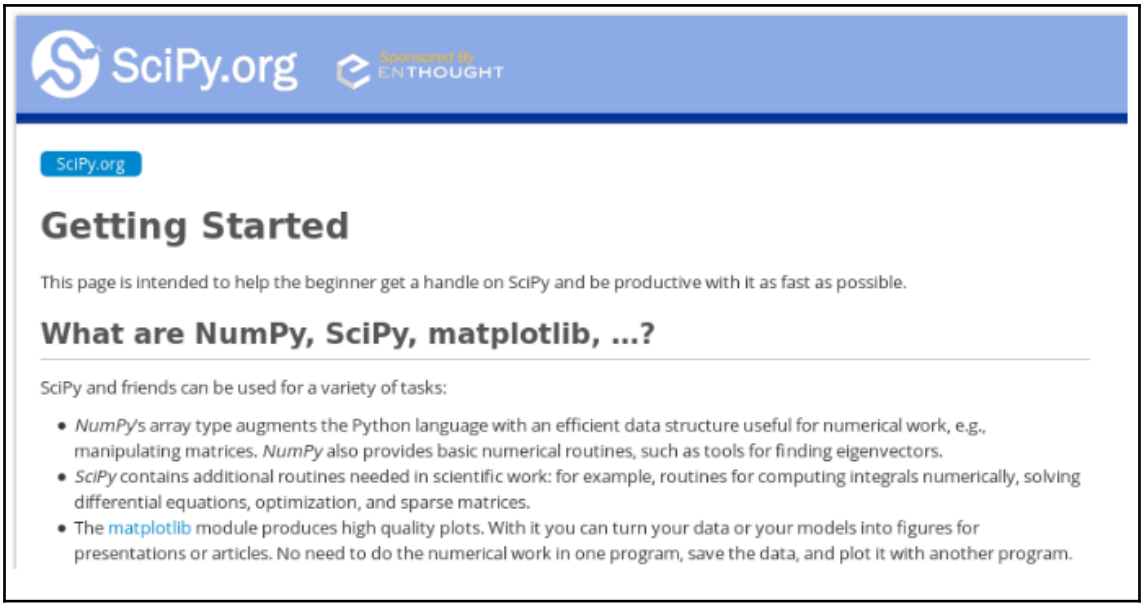
$$Recall = \frac{tp}{tp + fn}$$

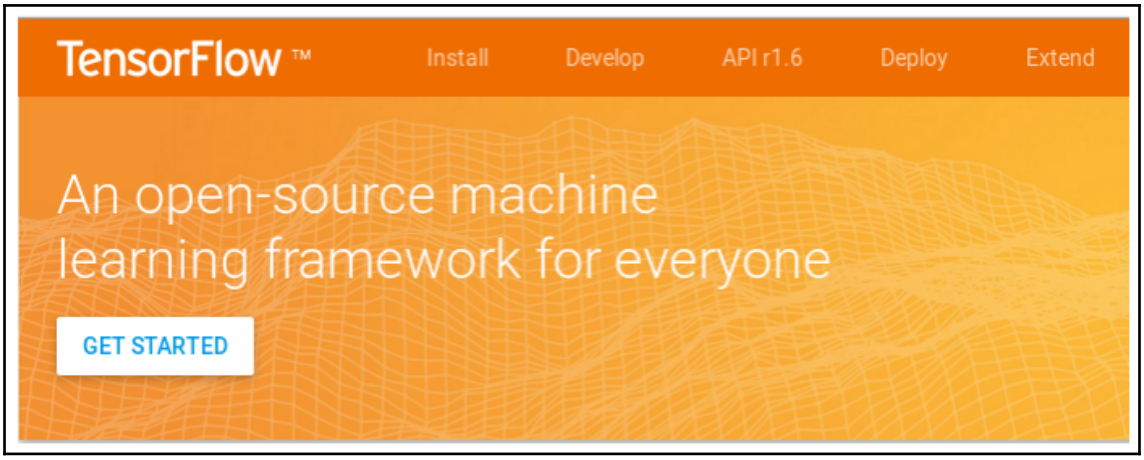
$$F - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

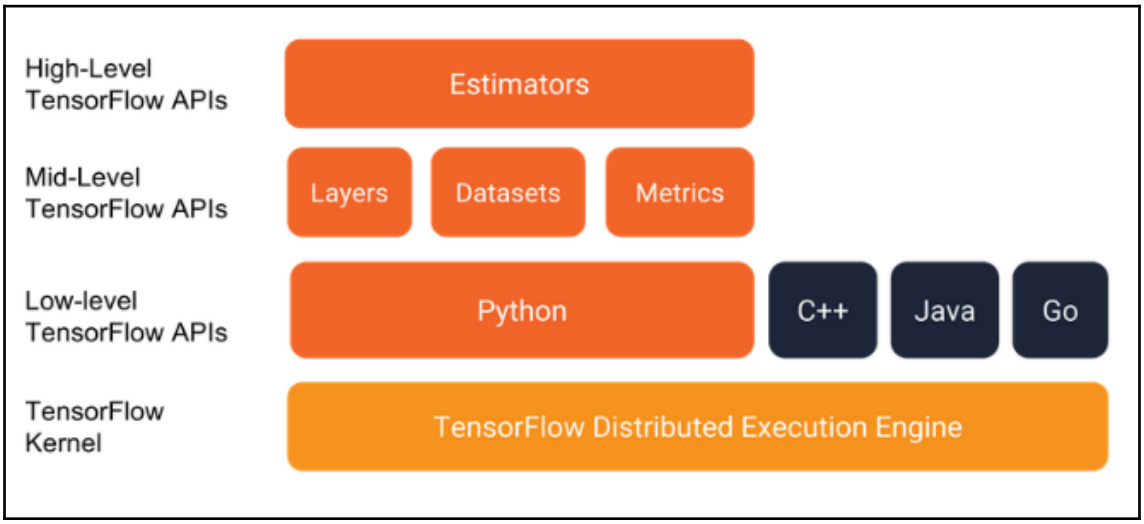


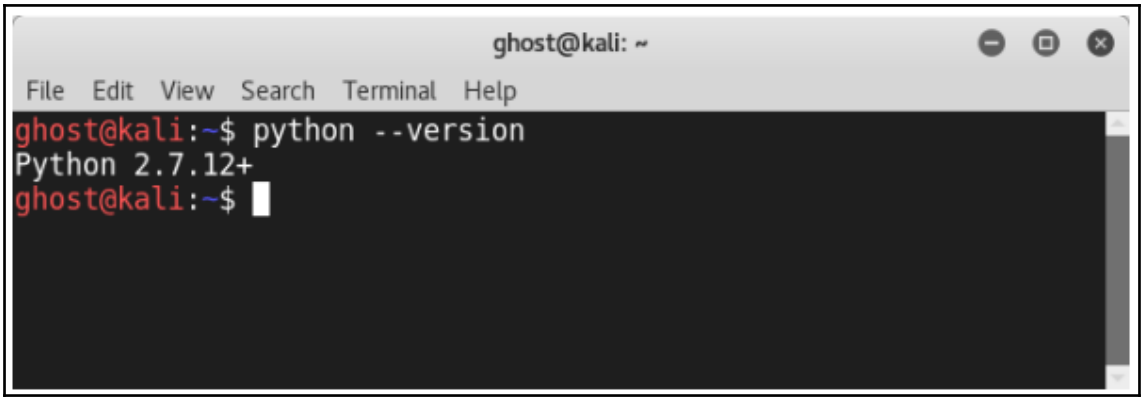


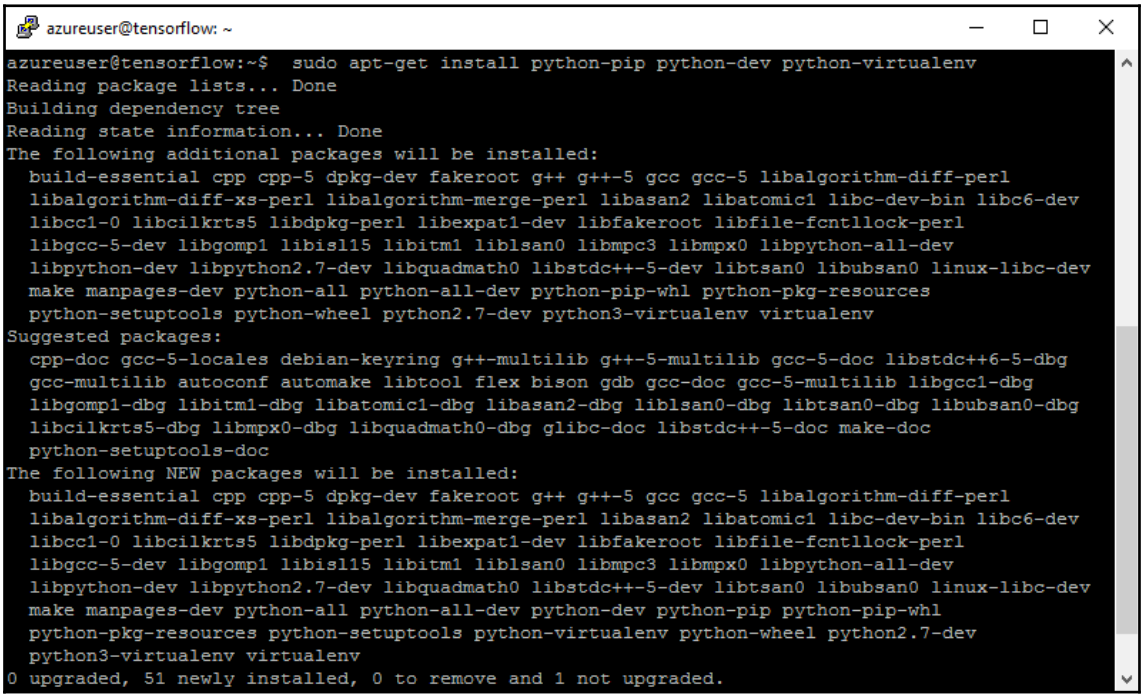

ghost@kali: ~
File Edit View Search Terminal Help
ghost@kali:~$ python
Python 2.7.12+ (default, Aug 4 2016, 20:04:34)
[GCC 6.1.1 20160724] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> import numpy as np
>>> a = np.array([7,2,4,3])
>>> print a
[7 2 4 3]
>>>

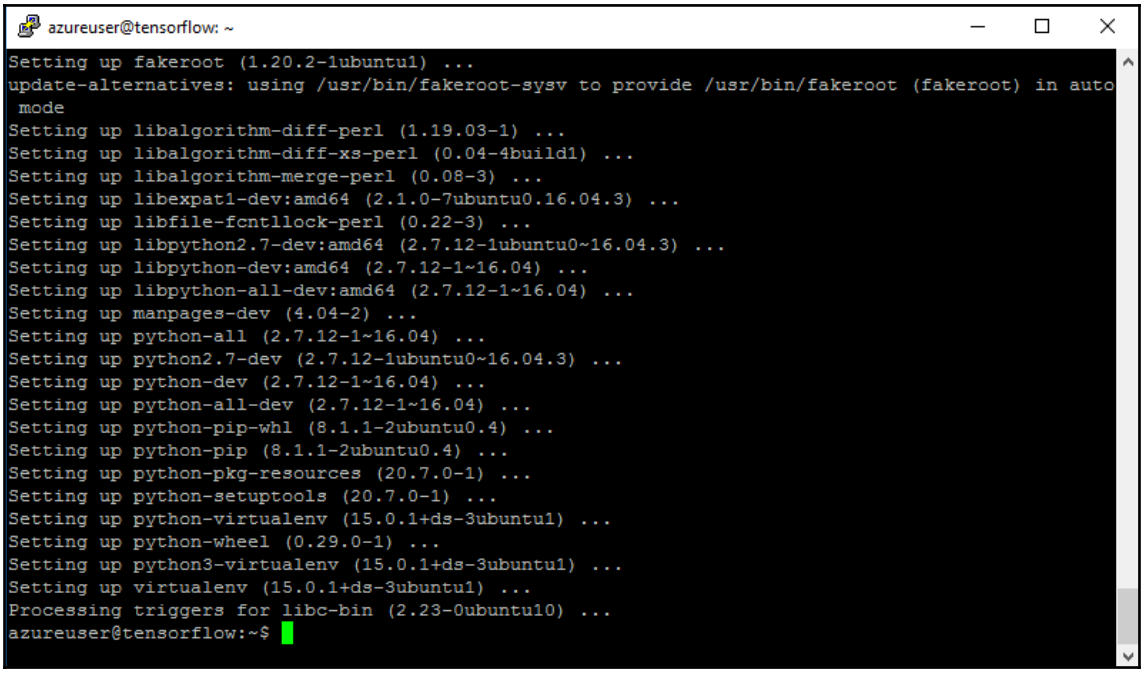

SciPy.org
Sponsored By ENTHOUGHT
SciPy.org
Getting Started
This page is intended to help the beginner get a handle on SciPy and be productive with it as fast as possible.
What are NumPy, SciPy, matplotlib, ...?
SciPy and friends can be used for a variety of tasks:
NumPy's array type augments the Python language with an efficient data structure useful for numerical work, e.g., manipulating matrices. NumPy also provides basic numerical routines, such as tools for finding eigenvectors.
SciPy contains additional routines needed in scientific work: for example, routines for computing integrals numerically, solving differential equations, optimization, and sparse matrices.
The matplotlib module produces high quality plots. With it you can turn your data or your models into figures for presentations or articles. No need to do the numerical work in one program, save the data, and plot it with another program.

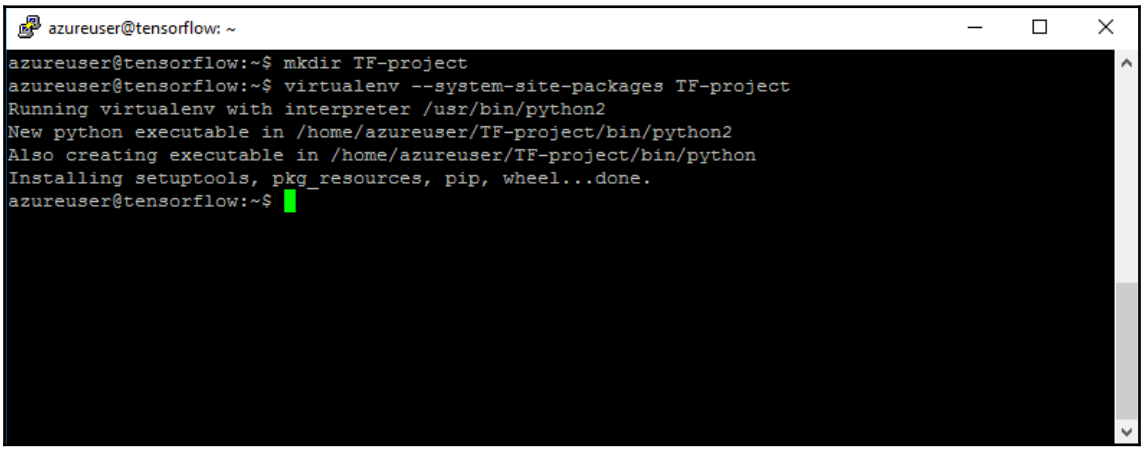

TensorFlow ™
Install
Develop
API r1.6
Deploy
Extend
An open-source machine learning framework for everyone
GET STARTED

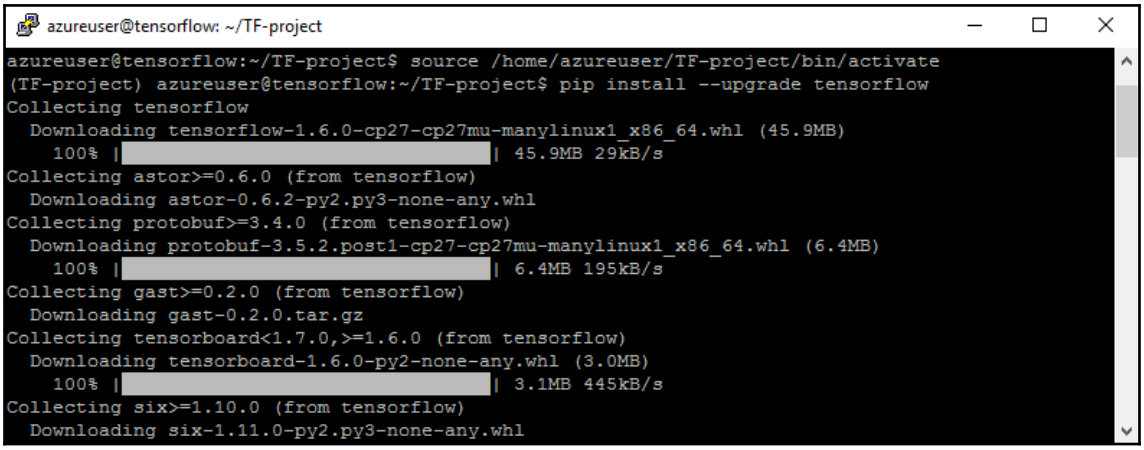

High-Level TensorFlow APIs
Estimators
Mid-Level TensorFlow APIs
Layers
Datasets
Metrics
Low-level TensorFlow APIs
Python
C++
Java
Go
TensorFlow Kernel
TensorFlow Distributed Execution Engine

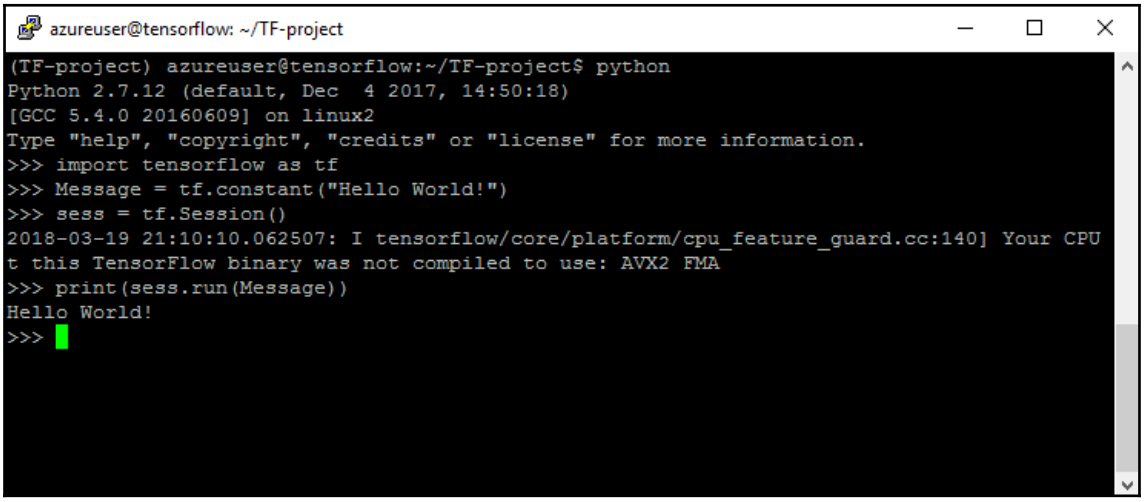

ghost@kali: ~
File Edit View Search Terminal Help
ghost@kali:~$ python --version
Python 2.7.12+
ghost@kali:~$

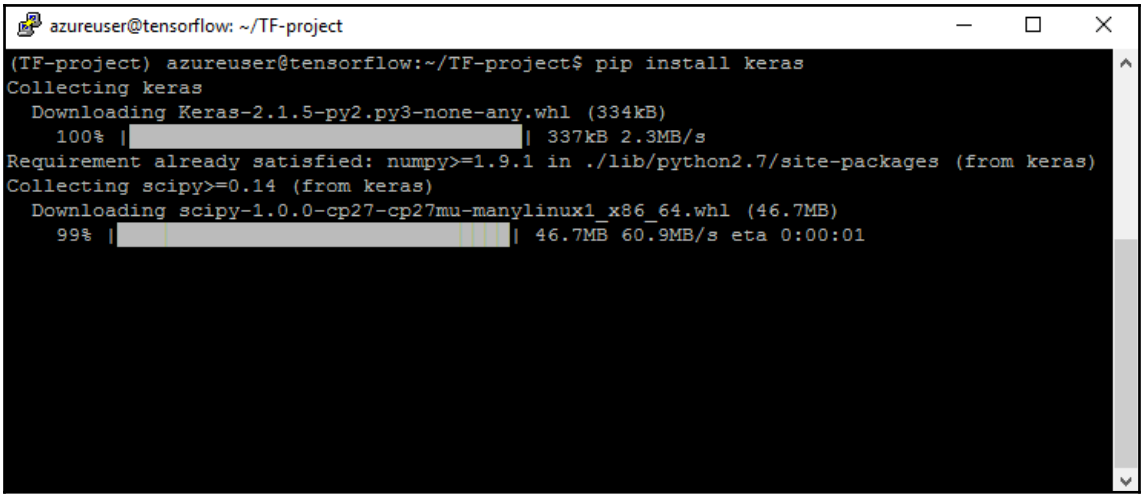

azureuser@tensorflow: ~
azureuser@tensorflow:~$ sudo apt-get install python-pip python-dev python-virtualenv
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
build-essential cpp cpp-5 dpkg-dev fakeroot g++ g++-5 gcc gcc-5 libalgorithm-diff-perl
libalgorithm-diff-xs-perl libalgorithm-merge-perl libasan2 libatomic1 libc-dev-bin libc6-dev
libcc1-0 libcilkrts5 libdpkg-perl libexpat1-dev libfakeroot libfile-fcntllock-perl
libgcc-5-dev libgomp1 libisl15 libitm1 liblsan0 libmpc3 libmpx0 libpython-all-dev
libpython-dev libpython2.7-dev libquadmath0 libstdc++-5-dev libtsan0 libubsan0 linux-libc-dev
make manpages-dev python-all python-all-dev python-pip-whl python-pkg-resources
python-setuptools python-wheel python2.7-dev python3-virtualenv virtualenv
Suggested packages:
cpp-doc gcc-5-locales debian-keyring g++-multilib g++-5-multilib gcc-5-doc libstdc++6-5-dbg
gcc-multilib autoconf automake libtool flex bison gdb gcc-doc gcc-5-multilib libgcc1-dbg
libgomp1-dbg libitm1-dbg libatomic1-dbg libasan2-dbg liblsan0-dbg libtsan0-dbg libubsan0-dbg
libcilkrts5-dbg libmpx0-dbg libquadmath0-dbg glibc-doc libstdc++-5-doc make-doc
python-setuptools-doc
The following NEW packages will be installed:
build-essential cpp cpp-5 dpkg-dev fakeroot g++ g++-5 gcc gcc-5 libalgorithm-diff-perl
libalgorithm-diff-xs-perl libalgorithm-merge-perl libasan2 libatomic1 libc-dev-bin libc6-dev
libcc1-0 libcilkrts5 libdpkg-perl libexpat1-dev libfakeroot libfile-fcntllock-perl
libgcc-5-dev libgomp1 libisl15 libitm1 liblsan0 libmpc3 libmpx0 libpython-all-dev
libpython-dev libpython2.7-dev libquadmath0 libstdc++-5-dev libtsan0 libubsan0 linux-libc-dev
make manpages-dev python-all python-all-dev python-dev python-pip python-pip-whl
python-pkg-resources python-setuptools python-virtualenv python-wheel python2.7-dev
python3-virtualenv virtualenv
0 upgraded, 51 newly installed, 0 to remove and 1 not upgraded.

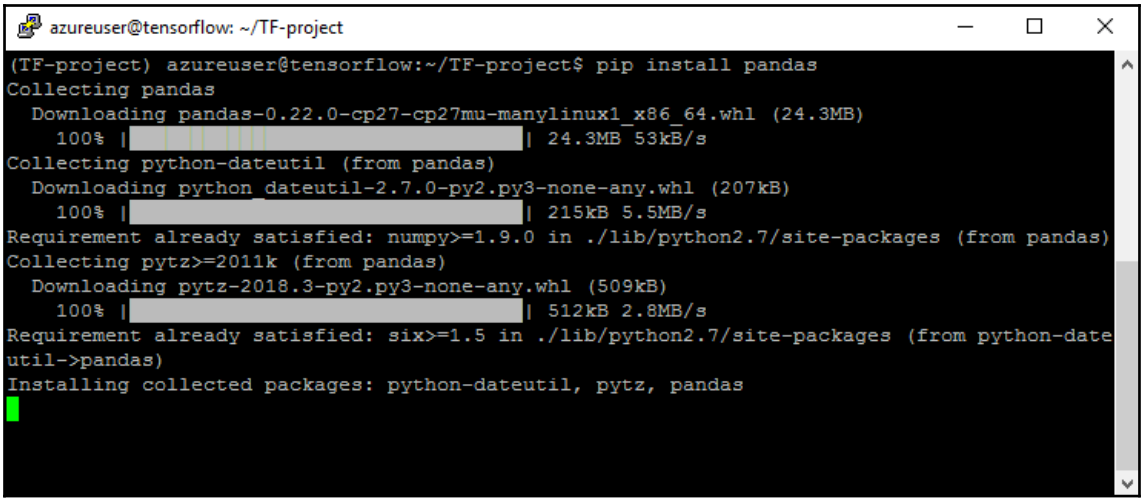

azureuser@tensorflow: ~
Setting up fakeroot (1.20.2-1ubuntu1) ...
update-alternatives: using /usr/bin/fakeroot-sysv to provide /usr/bin/fakeroot (fakeroot) in auto
mode
Setting up libalgorithm-diff-perl (1.19.03-1) ...
Setting up libalgorithm-diff-xs-perl (0.04-4build1) ...
Setting up libalgorithm-merge-perl (0.08-3) ...
Setting up libexpat1-dev:amd64 (2.1.0-7ubuntu0.16.04.3) ...
Setting up libfile-fcntllock-perl (0.22-3) ...
Setting up libpython2.7-dev:amd64 (2.7.12-1ubuntu0~16.04.3) ...
Setting up libpython-dev:amd64 (2.7.12-1~16.04) ...
Setting up libpython-all-dev:amd64 (2.7.12-1~16.04) ...
Setting up manpages-dev (4.04-2) ...
Setting up python-all (2.7.12-1~16.04) ...
Setting up python2.7-dev (2.7.12-1ubuntu0~16.04.3) ...
Setting up python-dev (2.7.12-1~16.04) ...
Setting up python-all-dev (2.7.12-1~16.04) ...
Setting up python-pip-whl (8.1.1-2ubuntu0.4) ...
Setting up python-pip (8.1.1-2ubuntu0.4) ...
Setting up python-pkg-resources (20.7.0-1) ...
Setting up python-setuptools (20.7.0-1) ...
Setting up python-virtualenv (15.0.1+ds-3ubuntu1) ...
Setting up python-wheel (0.29.0-1) ...
Setting up python3-virtualenv (15.0.1+ds-3ubuntu1) ...
Setting up virtualenv (15.0.1+ds-3ubuntu1) ...
Processing triggers for libc-bin (2.23-0ubuntu10) ...
azureuser@tensorflow:~$


azureuser@tensorflow: ~
azureuser@tensorflow:~$ mkdir TF-project
azureuser@tensorflow:~$ virtualenv --system-site-packages TF-project
Running virtualenv with interpreter /usr/bin/python2
New python executable in /home/azureuser/TF-project/bin/python2
Also creating executable in /home/azureuser/TF-project/bin/python
Installing setuptools, pkg_resources, pip, wheel...done.
azureuser@tensorflow:~$

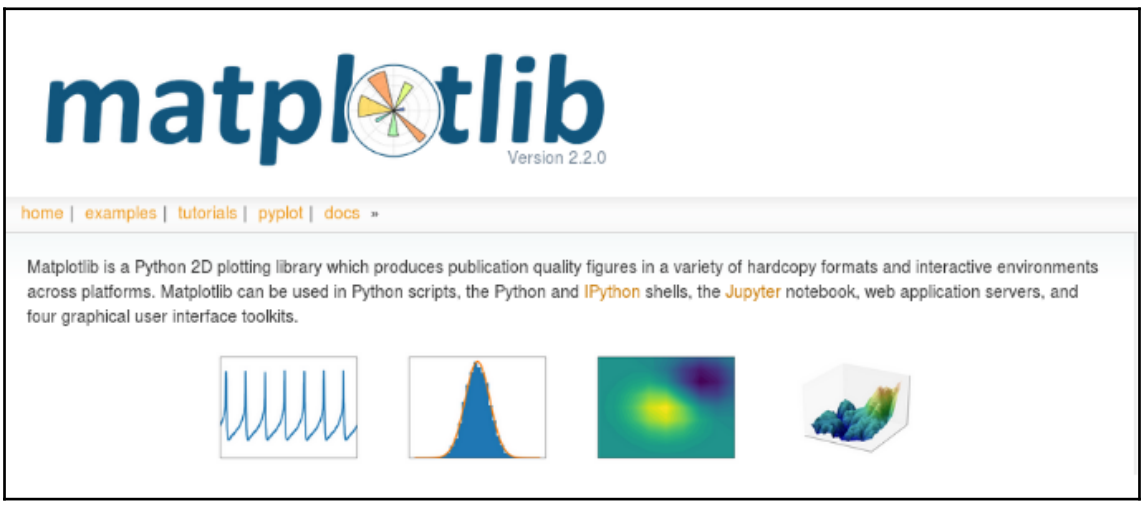

azureuser@tensorflow: ~/TF-project
azureuser@tensorflow:~/TF-project$ source /home/azureuser/TF-project/bin/activate
(TF-project) azureuser@tensorflow:~/TF-project$ pip install --upgrade tensorflow
Collecting tensorflow
Downloading tensorflow-1.6.0-cp27-cp27mu-manylinux1_x86_64.whl (45.9MB)
100% | | 45.9MB 29kB/s
Collecting astor>=0.6.0 (from tensorflow)
Downloading astor-0.6.2-py2.py3-none-any.whl
Collecting protobuf>=3.4.0 (from tensorflow)
Downloading protobuf-3.5.2.post1-cp27-cp27mu-manylinux1_x86_64.whl (6.4MB)
100% | | 6.4MB 195kB/s
Collecting gast>=0.2.0 (from tensorflow)
Downloading gast-0.2.0.tar.gz
Collecting tensorboard<1.7.0,>=1.6.0 (from tensorflow)
Downloading tensorboard-1.6.0-py2-none-any.whl (3.0MB)
100% | | 3.1MB 445kB/s
Collecting six>=1.10.0 (from tensorflow)
Downloading six-1.11.0-py2.py3-none-any.whl


azureuser@tensorflow: ~/TF-project
(TF-project) azureuser@tensorflow:~/TF-project$ python
Python 2.7.12 (default, Dec 4 2017, 14:50:18)
[GCC 5.4.0 20160609] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> import tensorflow as tf
>>> Message = tf.constant("Hello World!")
>>> sess = tf.Session()
2018-03-19 21:10:10.062507: I tensorflow/core/platform/cpu_feature_guard.cc:140] Your CPU
t this TensorFlow binary was not compiled to use: AVX2 FMA
>>> print(sess.run(Message))
Hello World!
>>>

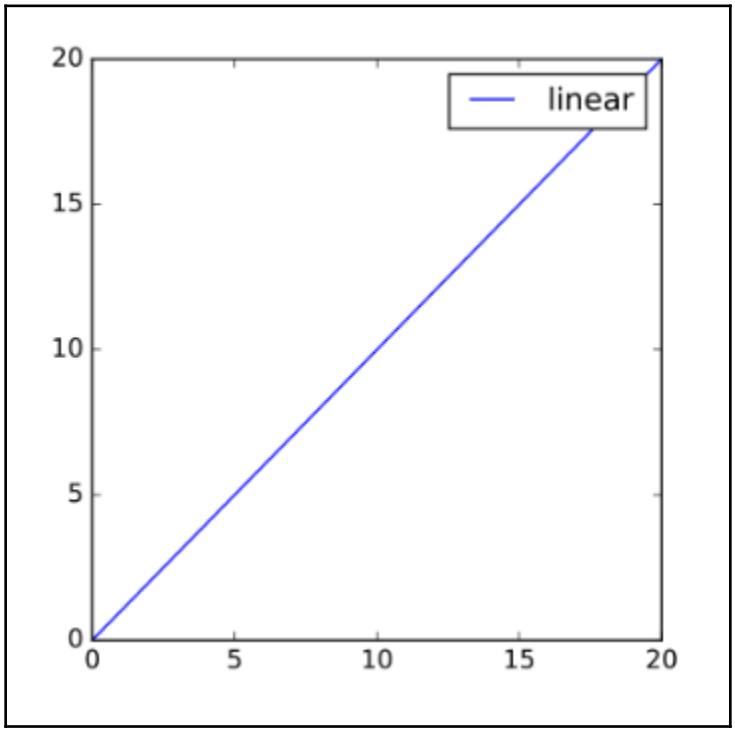

azureuser@tensorflow: ~/TF-project
(TF-project) azureuser@tensorflow:~/TF-project$ pip install keras
Collecting keras
Downloading Keras-2.1.5-py2.py3-none-any.whl (334kB)
100% | 337kB 2.3MB/s
Requirement already satisfied: numpy>=1.9.1 in ./lib/python2.7/site-packages (from keras)
Collecting scipy>=0.14 (from keras)
Downloading scipy-1.0.0-cp27-cp27mu-manylinux1_x86_64.whl (46.7MB)
99% | 46.7MB 60.9MB/s eta 0:00:01

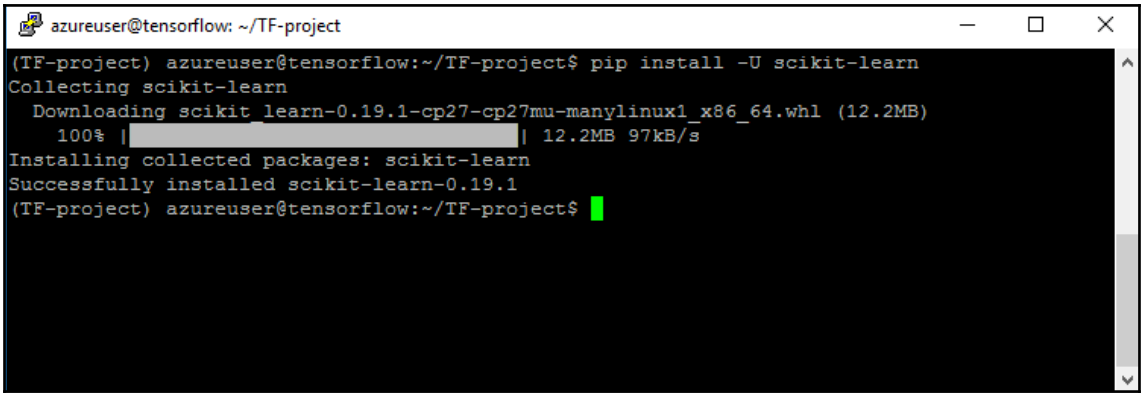

azureuser@tensorflow: ~/TF-project
(TF-project) azureuser@tensorflow:~/TF-project$ pip install pandas
Collecting pandas
Downloading pandas-0.22.0-cp27-cp27mu-manylinux1_x86_64.whl (24.3MB)
100% | 24.3MB 53kB/s
Collecting python-dateutil (from pandas)
Downloading python_dateutil-2.7.0-py2.py3-none-any.whl (207kB)
100% | 215kB 5.5MB/s
Requirement already satisfied: numpy>=1.9.0 in ./lib/python2.7/site-packages (from pandas)
Collecting pytz>=2011k (from pandas)
Downloading pytz-2018.3-py2.py3-none-any.whl (509kB)
100% | 512kB 2.8MB/s
Requirement already satisfied: six>=1.5 in ./lib/python2.7/site-packages (from python-date
util->pandas)
Installing collected packages: python-dateutil, pytz, pandas

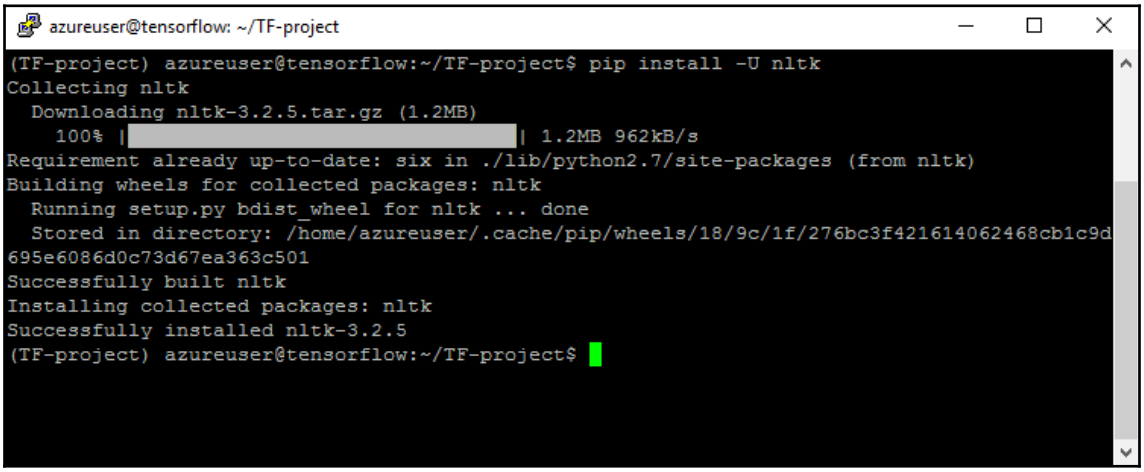
```
azureuser@tensorflow: ~/TF-project
(TF-project) azureuser@tensorflow:~/TF-project$ python
Python 2.7.12 (default, Dec  4 2017, 14:50:18)
[GCC 5.4.0 20160609] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> import pandas as pd
>>> import  numpy as np
>>> data = np.array(['p','a','c','k','t'])
>>> SR = pd.Series(data)
>>> print SR
0    p
1    a
2    c
3    k
4    t
dtype: object
>>>
```

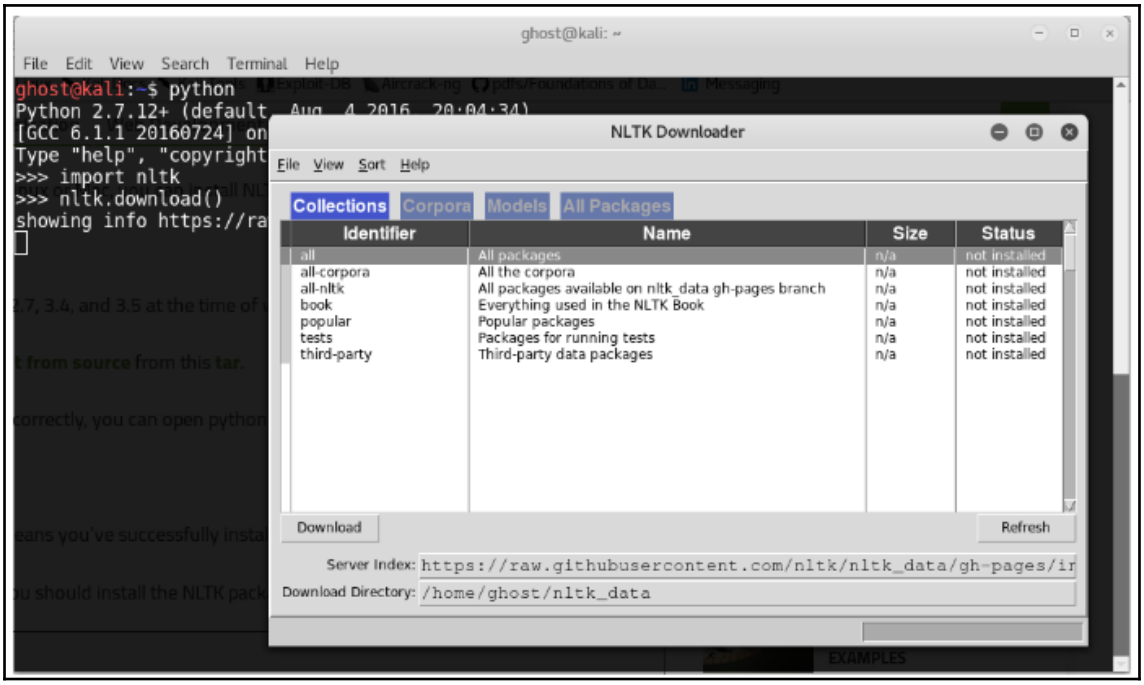


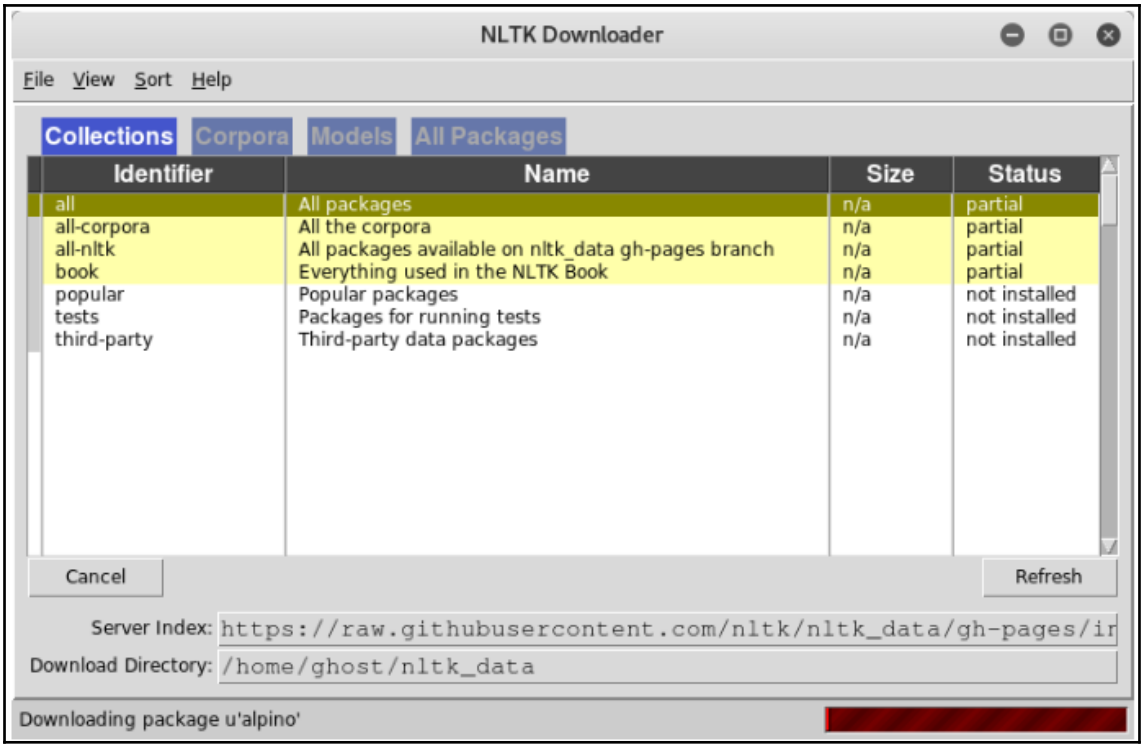
```
azureuser@tensorflow: ~/TF-project
(TF-project) azureuser@tensorflow:~/TF-project$  sudo apt-get install python3-matplotlib
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  blt fontconfig-config fonts-lyx javascript-common libblas-common libblas3
  libfontconfig1 libgfortran3 libjbig0 libjpeg-turbo8 libjpeg8 libjs-jquery
  libjs-jquery-ui liblapack3 liblcms2-2 libtcl8.6 libtiff5 libtk8.6 libwebp5 libwebpmux1
  libxft2 libxrender1 libxss1 python-matplotlib-data python3-cycler python3-dateutil
  python3-numpy python3-pil python3-pyparsing python3-tk python3-tz tk8.6-blt2.5
  ttf-bitstream-vera x11-common
Suggested packages:
  blt-demo apache2 | lighttpd | httpd libjs-jquery-ui-docs liblcms2-utils tcl8.6 tk8.6
  dvipng ffmpeg gir1.2-gtk-3.0 ghostscript inkscape ipython3 librsvg2-common
  python-matplotlib-doc python3-cairocffi python3-gi-cairo python3-gobject python3-nose
  python3-pyqt4 python3-scipy python3-sip python3-tornado texlive-extra-utils
  texlive-latex-extra ttf-staypuft gfortran python-numpy-doc python3-dev
  python3-numpy-dbg python-pil-doc python3-pil-dbg tix python3-tk-dbg
The following NEW packages will be installed:
  blt fontconfig-config fonts-lyx javascript-common libblas-common libblas3
```

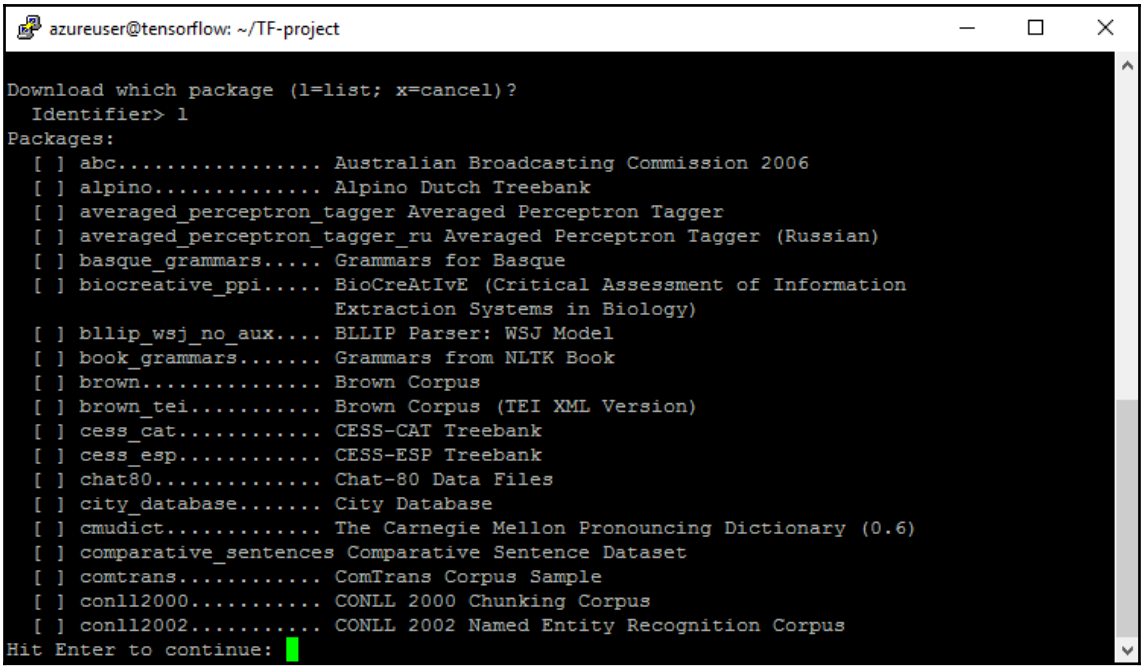


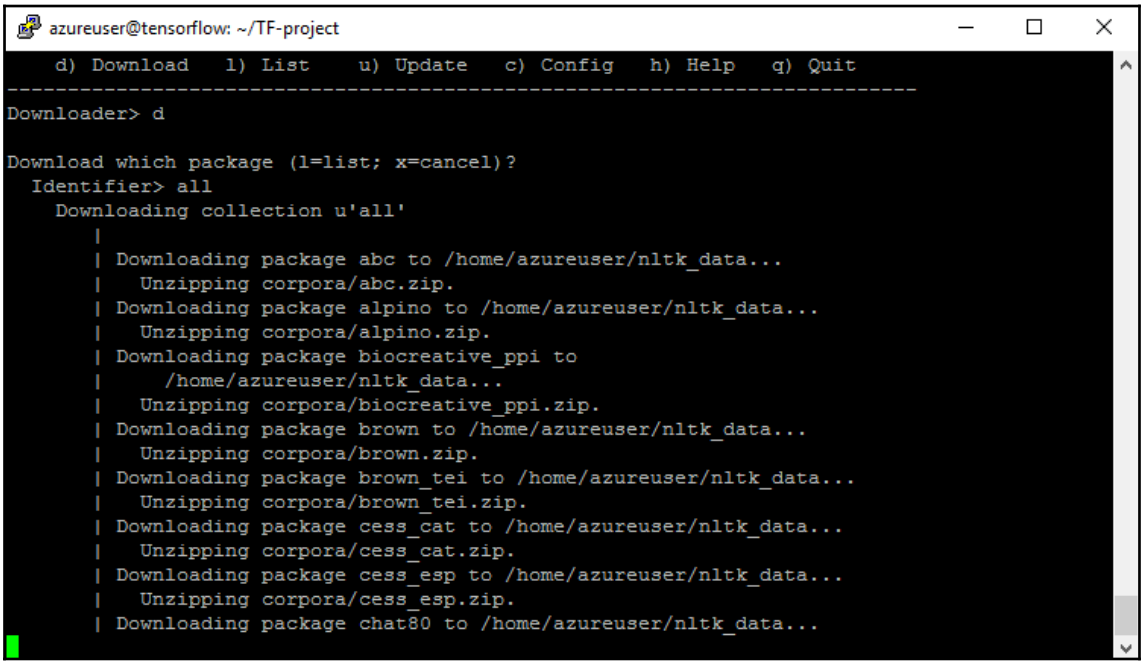

azureuser@tensorflow: ~/TF-project
(TF-project) azureuser@tensorflow:~/TF-project$ pip install -U scikit-learn
Collecting scikit-learn
Downloading scikit_learn-0.19.1-cp27-cp27mu-manylinux1_x86_64.whl (12.2MB)
100% | | 12.2MB 97kB/s
Installing collected packages: scikit-learn
Successfully installed scikit-learn-0.19.1
(TF-project) azureuser@tensorflow:~/TF-project$

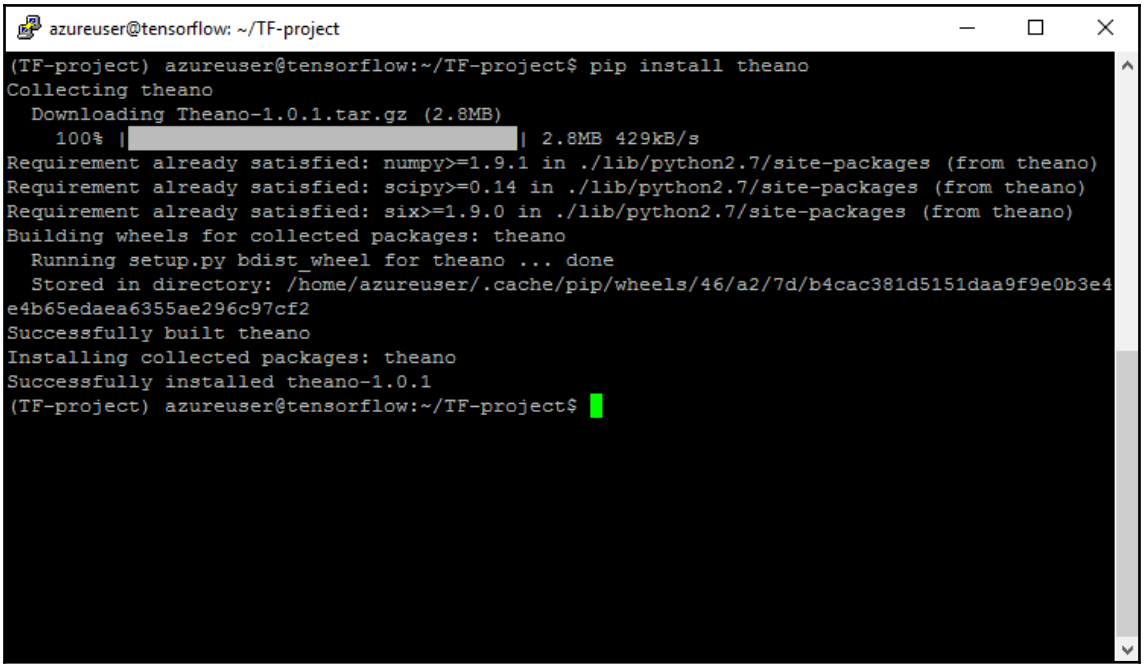

azureuser@tensorflow: ~/TF-project
(TF-project) azureuser@tensorflow:~/TF-project$ pip install -U nltk
Collecting nltk
Downloading nltk-3.2.5.tar.gz (1.2MB)
100% | | 1.2MB 962kB/s
Requirement already up-to-date: six in ./lib/python2.7/site-packages (from nltk)
Building wheels for collected packages: nltk
Running setup.py bdist_wheel for nltk ... done
Stored in directory: /home/azureuser/.cache/pip/wheels/18/9c/1f/276bc3f421614062468cb1c9d
695e6086d0c73d67ea363c501
Successfully built nltk
Installing collected packages: nltk
Successfully installed nltk-3.2.5
(TF-project) azureuser@tensorflow:~/TF-project$

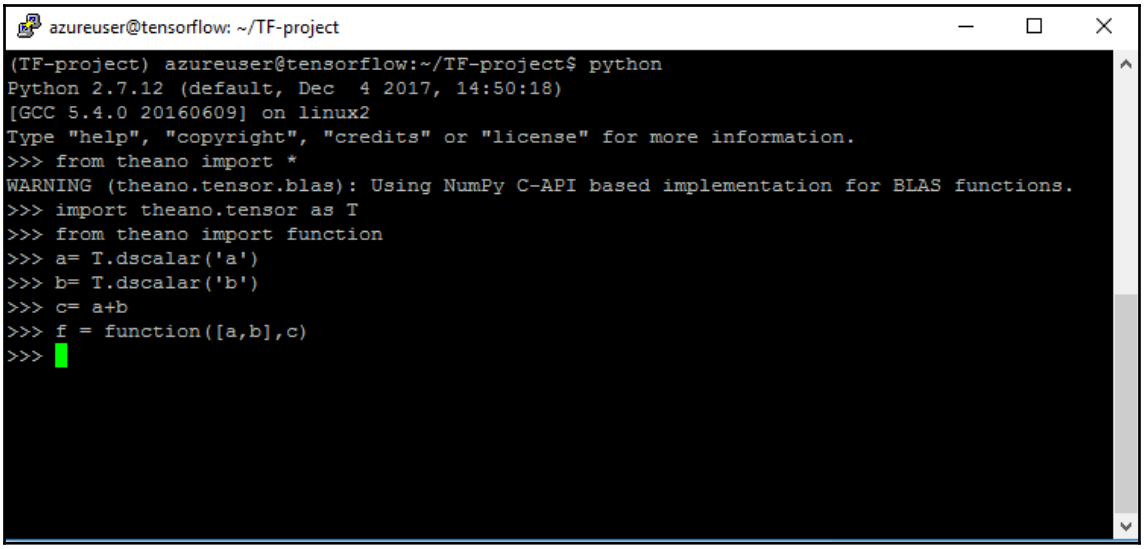

ghost@kali: ~
File Edit View Search Terminal Help
ghost@kali:~$ python
Python 2.7.12+ (default
[GCC 6.1.1 20160724] on
Type "help", "copyright
>>> import nltk
>>> nltk.download()
showing info https://ra
NLTK Downloader
File View Sort Help
Collections Corpora Models All Packages
Identifier Name Size Status
all All packages n/a not installed
all-corpora All the corpora n/a not installed
all-nltk All packages available on nltk_data gh-pages branch n/a not installed
book Everything used in the NLTK Book n/a not installed
popular Popular packages n/a not installed
tests Packages for running tests n/a not installed
third-party Third-party data packages n/a not installed
Download
Refresh
Server Index: https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/ir
Download Directory: /home/ghost/nltk_data

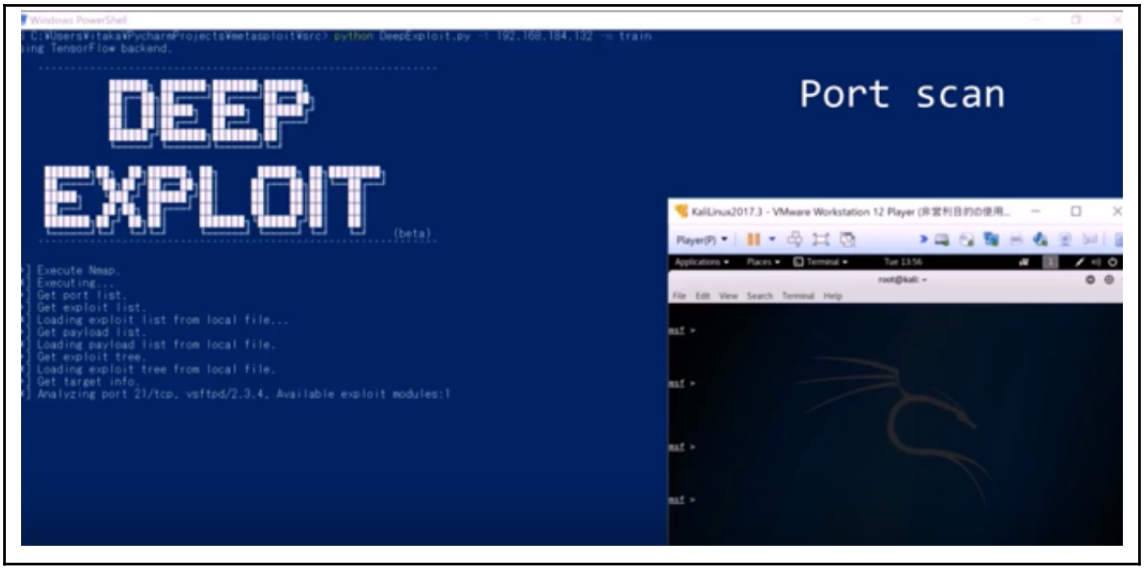

NLTK Downloader
File View Sort Help
Collections Corpora Models All Packages
Identifier Name Size Status
all All packages n/a partial
all-corpora All the corpora n/a partial
all-nltk All packages available on nltk_data gh-pages branch n/a partial
book Everything used in the NLTK Book n/a partial
popular Popular packages n/a not installed
tests Packages for running tests n/a not installed
third-party Third-party data packages n/a not installed
Cancel
Refresh
Server Index: https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/ir
Download Directory: /home/ghost/nltk_data
Downloading package u'alpino'

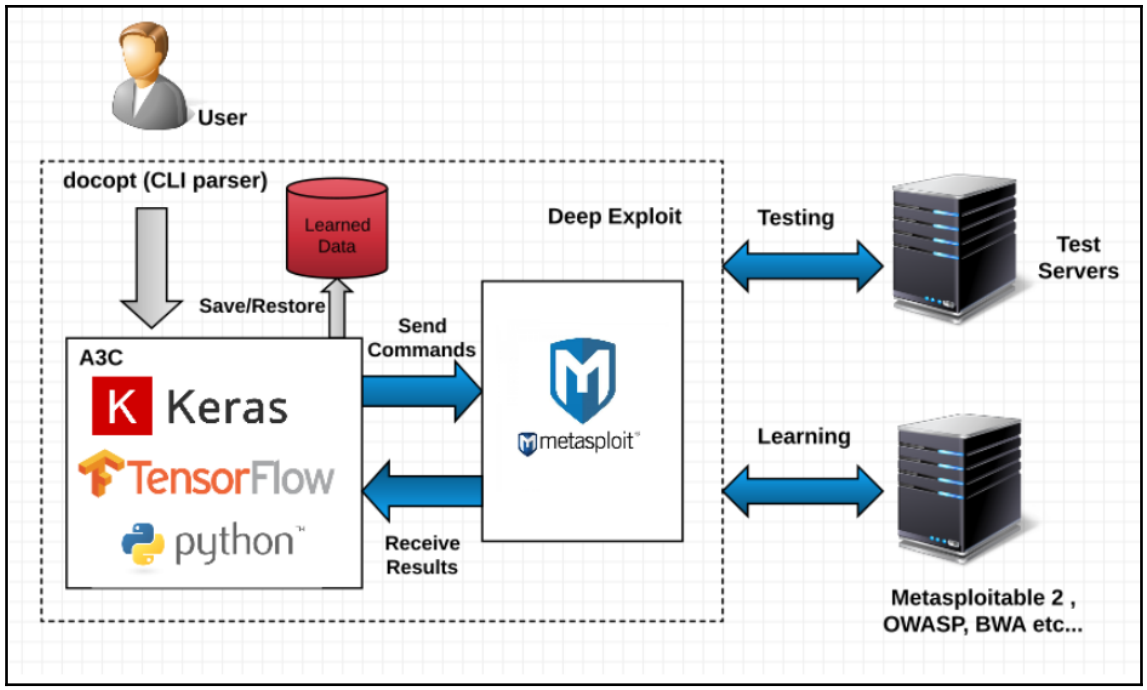

azureuser@tensorflow: ~/TF-project
Download which package (l=list; x=cancel)?
Identifier> l
Packages:
[ ] abc..... Australian Broadcasting Commission 2006
[ ] alpino..... Alpino Dutch Treebank
[ ] averaged_perceptron_tagger Averaged Perceptron Tagger
[ ] averaged_perceptron_tagger_ru Averaged Perceptron Tagger (Russian)
[ ] basque_grammars..... Grammars for Basque
[ ] biocreative_ppi..... BioCreAtIvE (Critical Assessment of Information Extraction Systems in Biology)
[ ] bllip_wsj_no_aux.... BLLIP Parser: WSJ Model
[ ] book_grammars..... Grammars from NLTK Book
[ ] brown..... Brown Corpus
[ ] brown_tei..... Brown Corpus (TEI XML Version)
[ ] cess_cat..... CESS-CAT Treebank
[ ] cess_esp..... CESS-ESP Treebank
[ ] chat80..... Chat-80 Data Files
[ ] city_database..... City Database
[ ] cmudict..... The Carnegie Mellon Pronouncing Dictionary (0.6)
[ ] comparative_sentences Comparative Sentence Dataset
[ ] comtrans..... ComTrans Corpus Sample
[ ] conll2000..... CONLL 2000 Chunking Corpus
[ ] conll2002..... CONLL 2002 Named Entity Recognition Corpus
Hit Enter to continue:


azureuser@tensorflow: ~/TF-project
d) Download l) List u) Update c) Config h) Help q) Quit
Downloader> d
Download which package (l=list; x=cancel)?
Identifier> all
Downloading collection u'all'
Downloading package abc to /home/azureuser/nltk_data...
Unzipping corpora/abc.zip.
Downloading package alpino to /home/azureuser/nltk_data...
Unzipping corpora/alpino.zip.
Downloading package biocreative_ppi to /home/azureuser/nltk_data...
Unzipping corpora/biocreative_ppi.zip.
Downloading package brown to /home/azureuser/nltk_data...
Unzipping corpora/brown.zip.
Downloading package brown_tei to /home/azureuser/nltk_data...
Unzipping corpora/brown_tei.zip.
Downloading package cess_cat to /home/azureuser/nltk_data...
Unzipping corpora/cess_cat.zip.
Downloading package cess_esp to /home/azureuser/nltk_data...
Unzipping corpora/cess_esp.zip.
Downloading package chat80 to /home/azureuser/nltk_data...

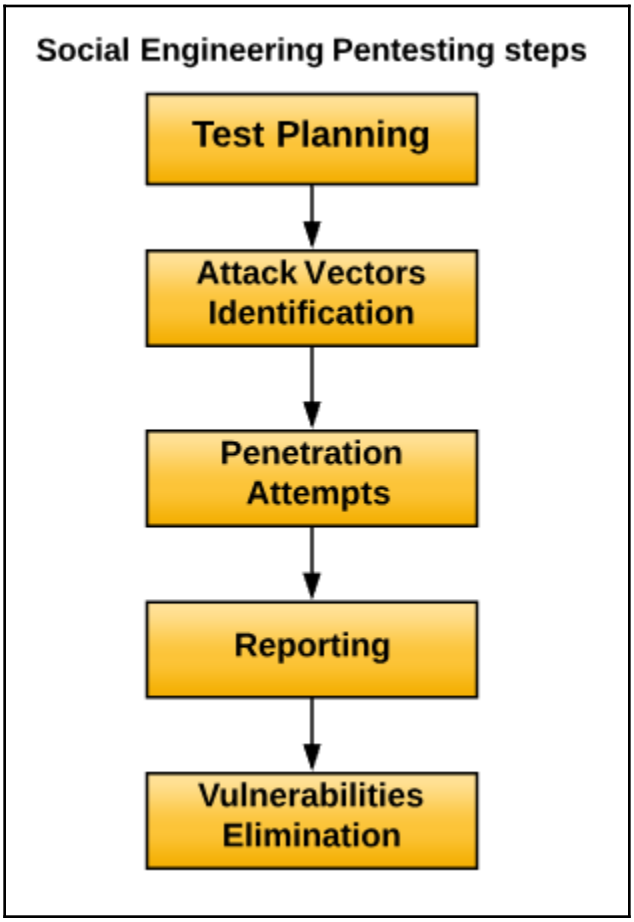

azureuser@tensorflow: ~/TF-project
(TF-project) azureuser@tensorflow:~/TF-project$ pip install theano
Collecting theano
Downloading Theano-1.0.1.tar.gz (2.8MB)
100% | | 2.8MB 429kB/s
Requirement already satisfied: numpy>=1.9.1 in ./lib/python2.7/site-packages (from theano)
Requirement already satisfied: scipy>=0.14 in ./lib/python2.7/site-packages (from theano)
Requirement already satisfied: six>=1.9.0 in ./lib/python2.7/site-packages (from theano)
Building wheels for collected packages: theano
Running setup.py bdist_wheel for theano ... done
Stored in directory: /home/azureuser/.cache/pip/wheels/46/a2/7d/b4cac381d5151daa9f9e0b3e4
e4b65edaea6355ae296c97cf2
Successfully built theano
Installing collected packages: theano
Successfully installed theano-1.0.1
(TF-project) azureuser@tensorflow:~/TF-project$


azureuser@tensorflow: ~/TF-project
(TF-project) azureuser@tensorflow:~/TF-project$ python
Python 2.7.12 (default, Dec 4 2017, 14:50:18)
[GCC 5.4.0 20160609] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> from theano import *
WARNING (theano.tensor.blas): Using NumPy C-API based implementation for BLAS functions.
>>> import theano.tensor as T
>>> from theano import function
>>> a= T.dscalar('a')
>>> b= T.dscalar('b')
>>> c= a+b
>>> f = function([a,b],c)
>>>


DEEP
EXPLOIT
(beta)
Port scan


User
docopt (CLI parser)
Learned Data
Save/Restore
A3C
Keras
TensorFlow
python
Send Commands
Receive Results
Deep Exploit
metasploit
Testing
Test Servers
Learning
Metasploitable 2 , OWASP, BWA etc...

Chapter 2: Phishing Domain Detection



Machine Learning Repository

Center for Machine Learning and Intelligent Systems

Phishing Websites Data Set

Download: Data Folder, Data Set Description

Abstract: This dataset collected mainly from: PhishTank archive, MillerSmiles archive, Googleâ€™s searching operators.

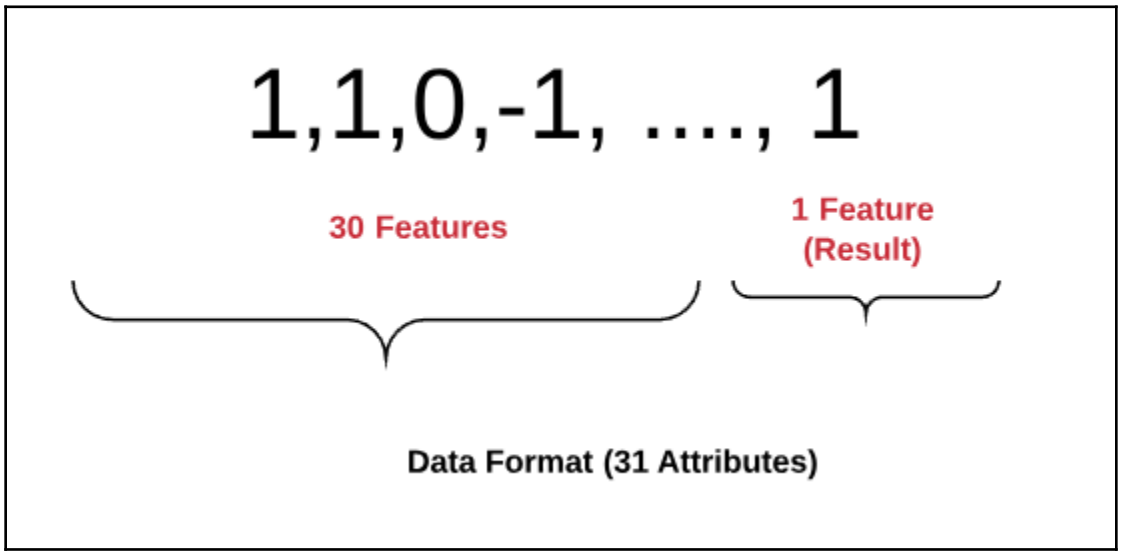
Data Set Characteristics:	N/A	**Number of Instances:**	2456	**Area:**	Computer Security
Attribute Characteristics:	Integer	**Number of Attributes:**	30	**Date Donated**	2015-03-26
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	72541

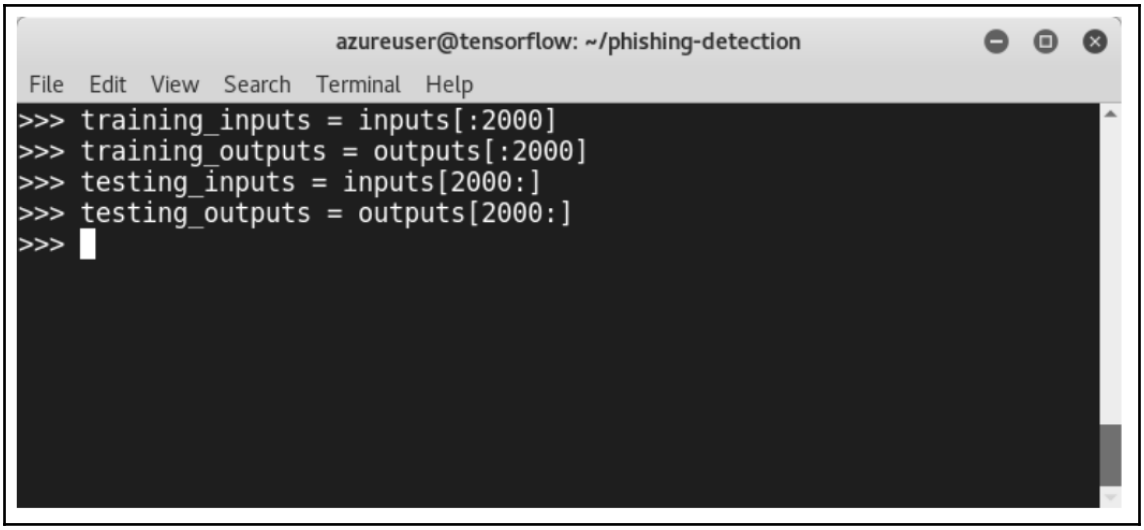
```
@relation phishing

@attribute having_IP_Address  { -1,1 }
@attribute URL_Length   { 1,0,-1 }
@attribute Shortining_Service { 1,-1 }
@attribute having_At_Symbol   { 1,-1 }
@attribute double_slash_redirecting { -1,1 }
@attribute Prefix_Suffix  { -1,1 }
@attribute having_Sub_Domain  { -1,0,1 }
@attribute SSLfinal_State  { -1,1,0 }
@attribute Domain_registeration_length { -1,1 }
@attribute Favicon { 1,-1 }
@attribute port { 1,-1 }
@attribute HTTPS_token { -1,1 }
@attribute Request_URL  { 1,-1 }
@attribute URL_of_Anchor { -1,0,1 }
@attribute Links_in_tags { 1,-1,0 }
@attribute SFH  { -1,1,0 }
@attribute Submitting_to_email { -1,1 }
@attribute Abnormal_URL { -1,1 }
@attribute Redirect  { 0,1 }
@attribute on_mouseover  { 1,-1 }
@attribute RightClick  { 1,-1 }
@attribute popUpWidnow  { 1,-1 }
@attribute Iframe { 1,-1 }
@attribute age_of_domain  { -1,1 }
@attribute DNSRecord   { -1,1 }
@attribute web_traffic  { -1,0,1 }
@attribute Page_Rank { -1,1 }
@attribute Google_Index { 1,-1 }
@attribute Links_pointing_to_page { 1,0,-1 }
@attribute Statistical_report { -1,1 }
@attribute Result  { -1,1 }
```

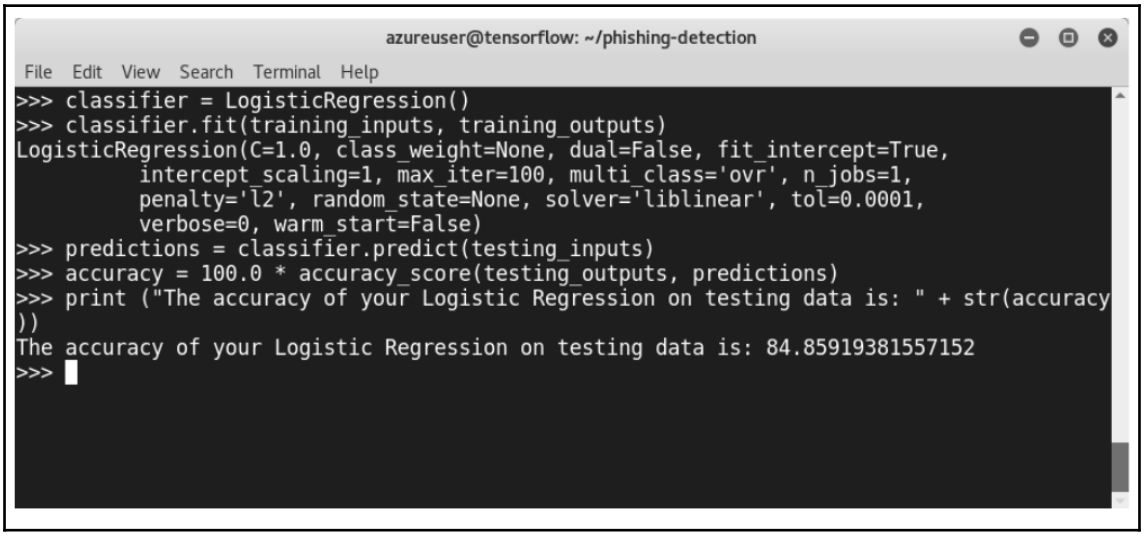
```
@data
-1,1,1,1,-1,-1,-1,-1,-1,1,1,-1,1,-1,1,-1,-1,-1,0,1,1,1,1,-1,-1,-1,-1,1,1,-1,-1
1,1,1,1,1,-1,0,1,-1,1,1,-1,1,0,-1,-1,1,1,0,1,1,1,1,-1,-1,0,-1,1,1,1,-1
1,0,1,1,1,-1,-1,-1,-1,1,1,-1,1,0,-1,-1,-1,-1,0,1,1,1,1,1,-1,1,-1,1,0,-1,-1
1,0,1,1,1,-1,-1,-1,-1,1,1,1,-1,-1,0,0,-1,1,1,0,1,1,1,1,-1,-1,1,-1,1,-1,1,-1
1,0,-1,1,1,-1,1,1,-1,1,1,1,1,1,0,0,-1,1,1,0,-1,1,-1,1,-1,-1,0,-1,1,1,1,1
-1,0,-1,1,-1,-1,1,1,-1,1,1,-1,1,0,0,-1,-1,-1,0,1,1,1,1,1,1,1,-1,1,-1,-1,1
1,0,-1,1,1,-1,-1,-1,1,1,1,1,1,-1,-1,0,-1,-1,-1,0,1,1,1,1,1,-1,-1,-1,1,0,-1,-1
1,0,1,1,1,-1,-1,-1,-1,1,1,1,1,-1,-1,0,-1,-1,1,1,0,1,1,1,1,1,-1,-1,0,-1,1,0,1,-1
1,0,-1,1,1,-1,1,1,-1,1,1,1,-1,1,0,1,-1,1,1,1,0,1,1,1,1,1,1,-1,1,1,1,1,0,1,1
1,1,-1,1,1,-1,-1,-1,1,-1,1,1,1,1,1,1,0,1,-1,1,1,1,0,1,1,1,1,1,1,-1,0,-1,1,0,1,-1
1,1,1,1,1,1,-1,0,1,1,1,1,1,1,1,-1,0,0,-1,-1,-1,0,1,1,1,1,1,-1,1,1,1,1,1,1,-1,-1,1
1,1,-1,1,1,-1,-1,-1,-1,1,1,1,1,1,1,1,-1,-1,-1,-1,-1,0,1,1,1,1,1,-1,-1,-1,1,1,0,-1,-1
-1,1,-1,1,-1,-1,0,0,1,1,1,1,-1,-1,-1,1,1,-1,1,1,1,0,-1,1,-1,1,1,1,-1,-1,-1,1,1,0,1,-1
1,1,-1,1,1,-1,0,-1,1,1,1,1,1,-1,-1,-1,-1,1,1,1,0,1,1,1,1,1,-1,-1,0,-1,1,1,1,1,-1
1,1,-1,1,1,1,-1,1,-1,1,1,-1,1,1,0,1,1,1,1,1,0,1,1,1,1,1,1,-1,1,-1,1,-1,1,1
1,-1,-1,-1,1,-1,0,0,1,1,1,1,1,-1,-1,0,-1,1,1,0,1,1,1,1,1,1,-1,-1,-1,1,1,0,1,-1
1,-1,-1,1,1,-1,1,1,-1,1,1,-1,1,1,0,-1,-1,-1,-1,0,1,1,1,1,1,1,-1,0,-1,1,1,1,-1,-1
1,-1,1,1,1,1,-1,-1,0,1,1,-1,1,1,1,0,-1,-1,-1,-1,0,1,1,1,1,1,-1,1,1,1,-1,1,1,-1
1,1,1,1,1,1,-1,-1,1,1,1,1,1,1,-1,-1,0,-1,-1,-1,-1,0,1,1,1,1,1,1,-1,-1,1,1,-1,-1,1
1,1,1,1,1,1,-1,-1,1,-1,1,1,1,1,1,1,0,0,-1,-1,-1,0,-1,-1,-1,-1,1,-1,0,-1,1,1,0,-1,1
1,0,-1,1,1,-1,0,1,-1,1,1,1,1,1,1,0,0,-1,-1,-1,0,-1,1,-1,1,-1,1,1,1,-1,1,-1,-1,1
1,0,1,1,1,-1,0,1,1,1,1,1,-1,-1,-1,0,-1,-1,-1,-1,0,1,1,1,1,1,-1,1,-1,-1,1,1,0,-1,1
1,1,1,1,1,1,-1,-1,-1,-1,1,1,-1,1,1,0,0,-1,1,1,0,1,1,1,1,1,1,1,0,-1,1,-1,1,1
1,1,1,1,1,1,-1,1,0,-1,1,1,1,1,1,1,1,0,0,-1,1,1,0,1,1,1,1,1,1,1,-1,1,-1,1,1,1
1,-1,-1,-1,1,-1,1,1,1,-1,1,1,1,-1,-1,-1,0,0,-1,1,1,0,1,1,1,1,1,1,1,-1,-1,1,1,0,1,1
1,-1,1,1,1,1,-1,0,1,-1,1,1,1,1,1,1,1,0,-1,1,1,0,1,1,1,1,-1,1,1,1,-1,1,1,0,1,1
1,-1,1,1,1,1,-1,0,-1,1,1,1,1,-1,-1,-1,-1,-1,-1,-1,-1,0,1,1,1,1,1,1,1,1,0,-1,1,-1,-1,-1
1,-1,-1,1,1,1,1,-1,1,1,1,1,1,1,1,-1,1,1,0,-1,-1,-1,0,1,1,1,1,1,1,-1,0,-1,1,1,0,-1,1
1,-1,-1,1,-1,1,-1,1,-1,1,1,1,1,1,1,1,1,1,0,-1,1,1,1,1,1,1,1,1,-1,-1,1,-1,1,-1,1,1
1,-1,1,1,1,1,-1,-1,1,-1,1,1,1,1,1,1,1,1,0,-1,1,1,1,0,1,1,1,1,-1,1,1,1,1,1,1,0,1,1
1,-1,1,1,1,1,-1,-1,1,-1,1,1,1,-1,1,1,0,1,-1,1,1,1,0,1,1,1,1,-1,-1,-1,-1,1,1,0,1,1
1,-1,1,1,1,1,-1,-1,1,-1,-1,1,-1,1,1,0,-1,-1,-1,-1,0,-1,1,-1,-1,1,-1,1,1,1,1,0,-1,1
1,-1,1,1,1,1,1,1,1,-1,1,1,1,1,1,1,1,1,-1,-1,0,1,1,1,1,1,1,1,-1,-1,1,-1,-1,1
1,0,1,1,1,-1,-1,1,-1,1,1,1,1,1,1,0,1,-1,-1,-1,0,1,1,1,1,1,1,1,1,-1,1,0,-1,1
```

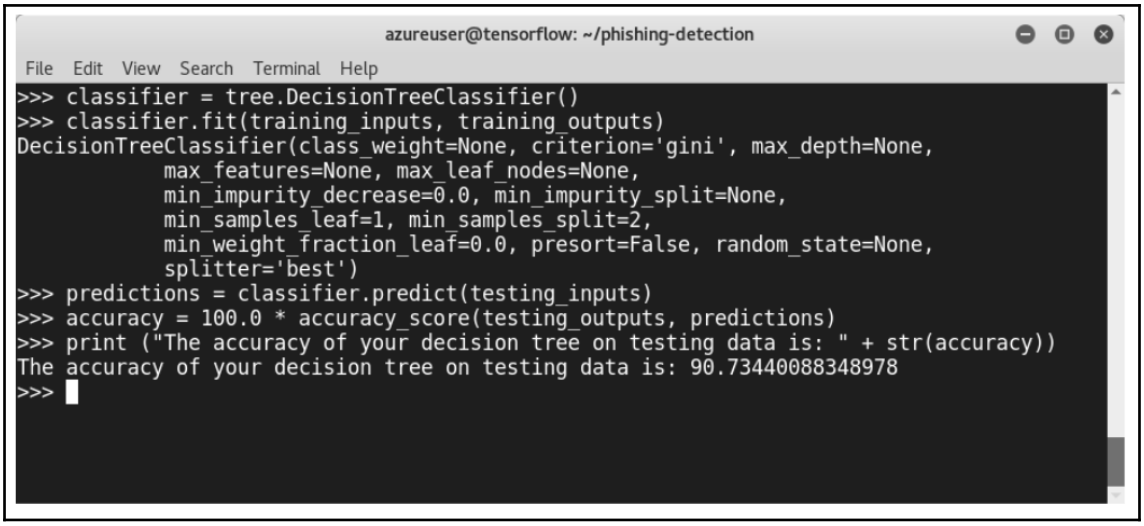
dataset.csv Open with Google Sheets

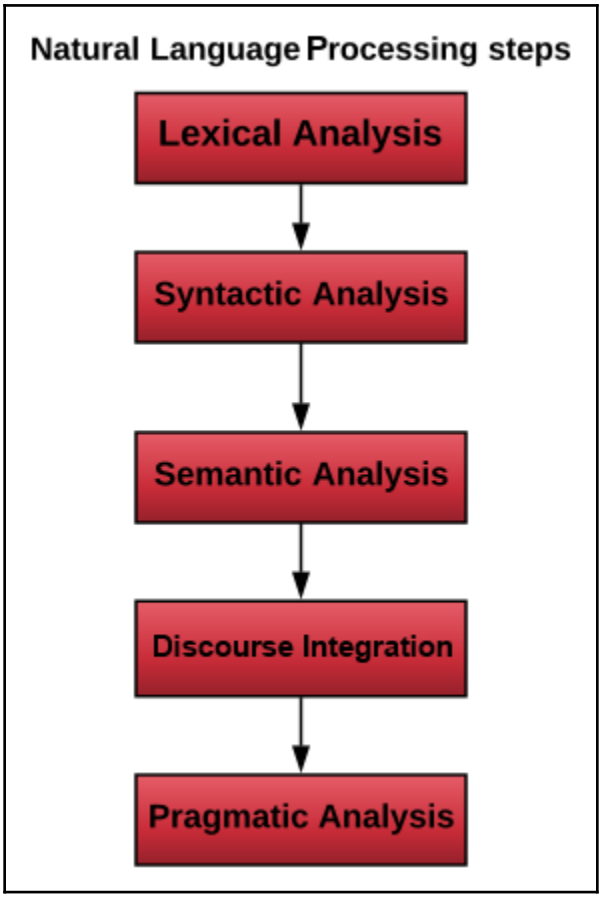
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	-1	1	1	1	-1	-1	-1	-1	-1	1	1	-1	-1
2	1	1	1	1	1	-1	0	1	-1	1	1	-1	1
3	1	0	1	1	1	-1	-1	-1	-1	1	1	-1	1
4	1	0	1	1	1	-1	-1	-1	1	1	1	-1	-1
5	1	0	-1	1	1	-1	1	1	-1	1	1	1	1
6	-1	0	-1	1	-1	-1	1	1	-1	1	1	-1	1
7	1	0	-1	1	1	-1	-1	-1	1	1	1	1	-1
8	1	0	1	1	1	-1	-1	-1	1	1	1	-1	-1
9	1	0	-1	1	1	-1	1	1	-1	1	1	-1	1
10	1	1	-1	1	1	-1	-1	1	-1	1	1	1	1
11	1	1	1	1	1	-1	0	1	1	1	1	1	-1
12	1	1	-1	1	1	-1	1	-1	-1	1	1	1	1
13	-1	1	-1	1	-1	-1	0	0	1	1	1	-1	-1
14	1	1	-1	1	1	-1	0	-1	1	1	1	1	-1
15	1	1	-1	1	1	1	-1	1	-1	1	1	-1	1
16	1	-1	-1	-1	1	-1	0	0	1	1	1	1	-1
17	1	-1	-1	1	1	-1	1	1	-1	1	1	-1	1
18	1	-1	1	1	1	-1	-1	0	1	1	-1	1	1

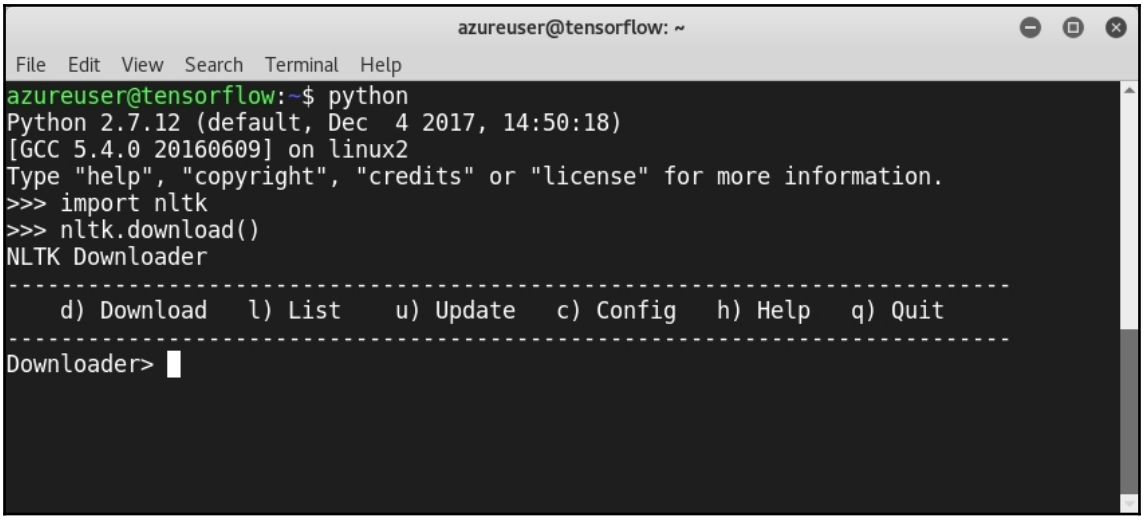


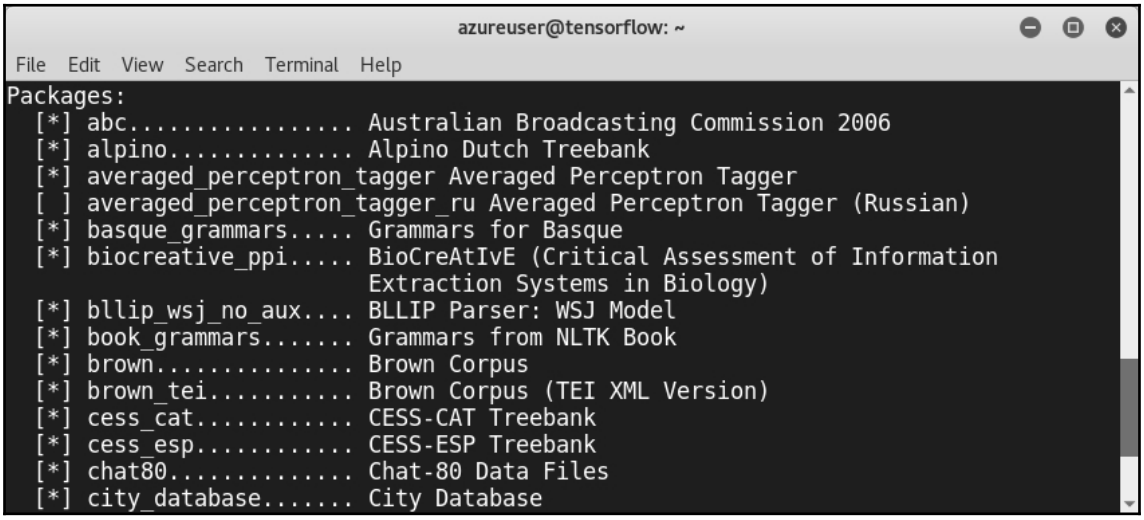

azureuser@tensorflow: ~/phishing-detection
File Edit View Search Terminal Help
>>> training_inputs = inputs[:2000]
>>> training_outputs = outputs[:2000]
>>> testing_inputs = inputs[2000:]
>>> testing_outputs = outputs[2000:]
>>>

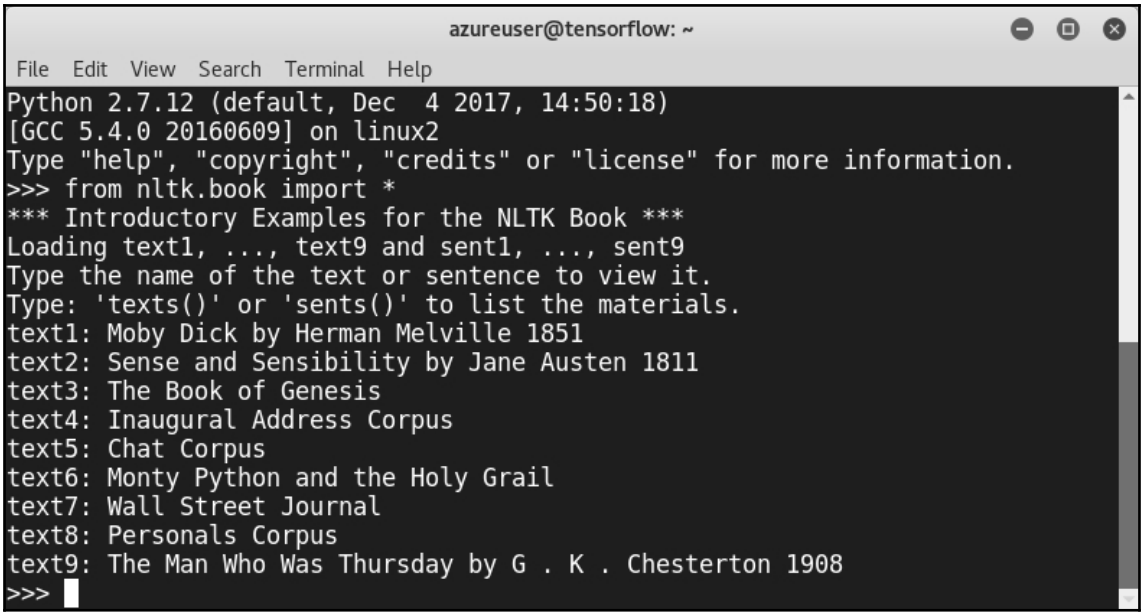

azureuser@tensorflow: ~/phishing-detection
File Edit View Search Terminal Help
>>> classifier = LogisticRegression()
>>> classifier.fit(training_inputs, training_outputs)
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,
penalty='l2', random_state=None, solver='liblinear', tol=0.0001,
verbose=0, warm_start=False)
>>> predictions = classifier.predict(testing_inputs)
>>> accuracy = 100.0 * accuracy_score(testing_outputs, predictions)
>>> print ("The accuracy of your Logistic Regression on testing data is: " + str(accuracy
))
The accuracy of your Logistic Regression on testing data is: 84.85919381557152
>>>

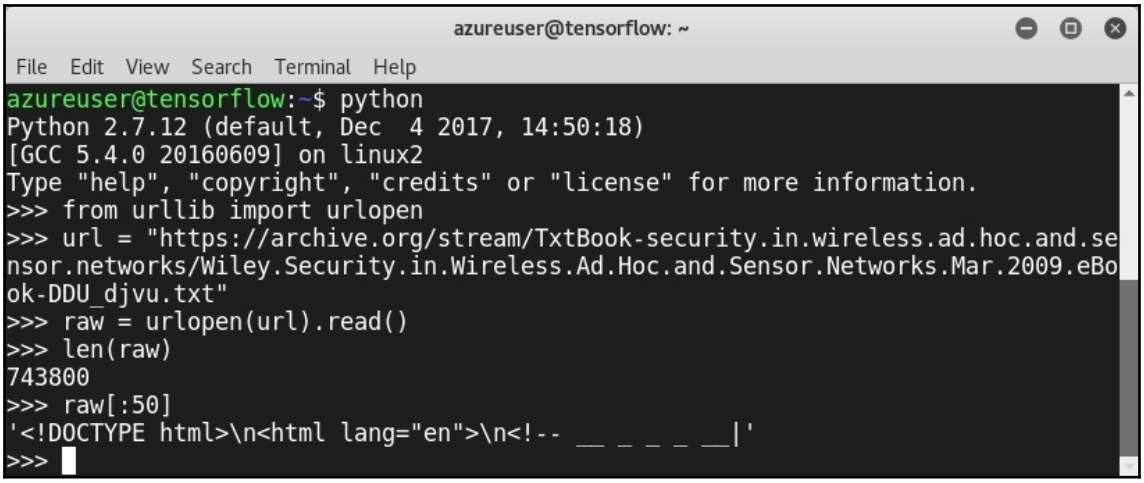

azureuser@tensorflow: ~/phishing-detection
File Edit View Search Terminal Help
>>> classifier = tree.DecisionTreeClassifier()
>>> classifier.fit(training_inputs, training_outputs)
DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=None,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, presort=False, random_state=None,
splitter='best')
>>> predictions = classifier.predict(testing_inputs)
>>> accuracy = 100.0 * accuracy_score(testing_outputs, predictions)
>>> print ("The accuracy of your decision tree on testing data is: " + str(accuracy))
The accuracy of your decision tree on testing data is: 90.73440088348978
>>>

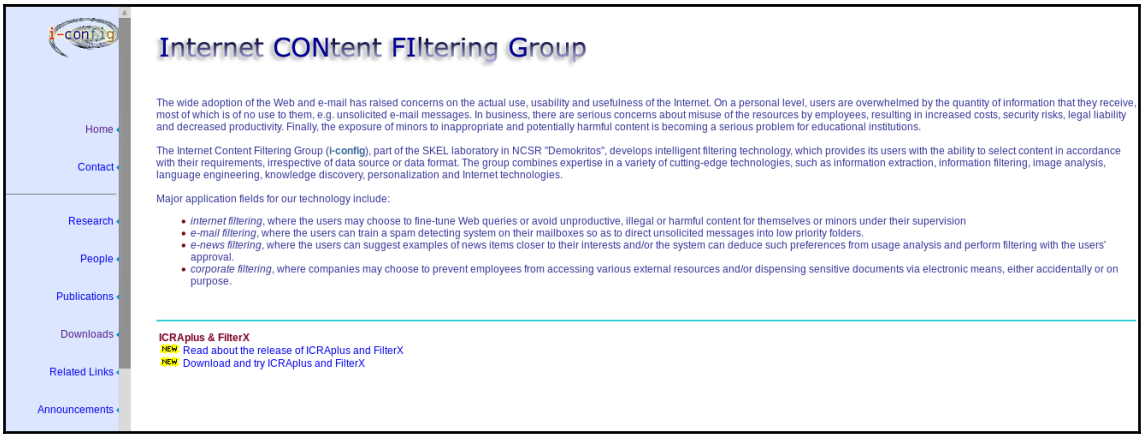

Natural Language Processing steps
Lexical Analysis
Syntactic Analysis
Semantic Analysis
Discourse Integration
Pragmatic Analysis

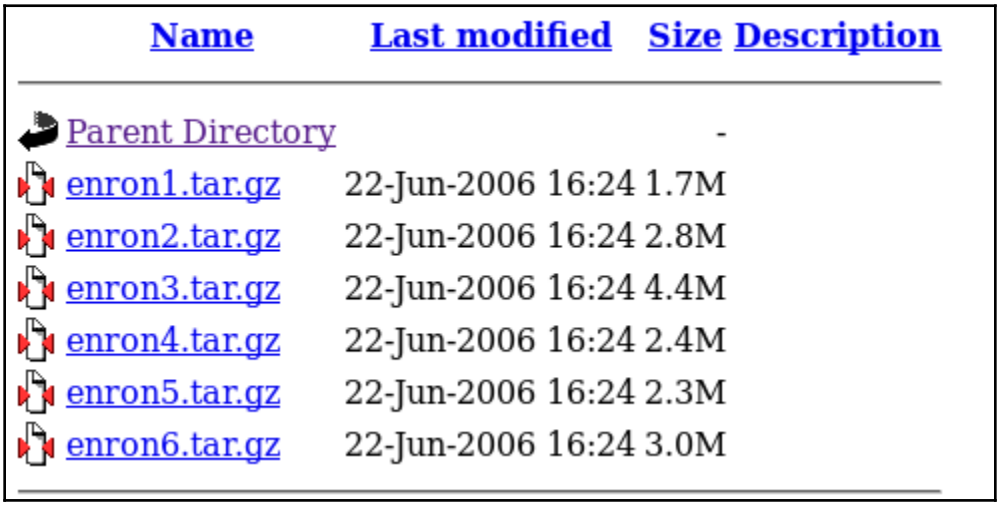

azureuser@tensorflow: ~
File Edit View Search Terminal Help
azureuser@tensorflow:~$ python
Python 2.7.12 (default, Dec 4 2017, 14:50:18)
[GCC 5.4.0 20160609] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> import nltk
>>> nltk.download()
NLTK Downloader
d) Download l) List u) Update c) Config h) Help q) Quit
Downloader>

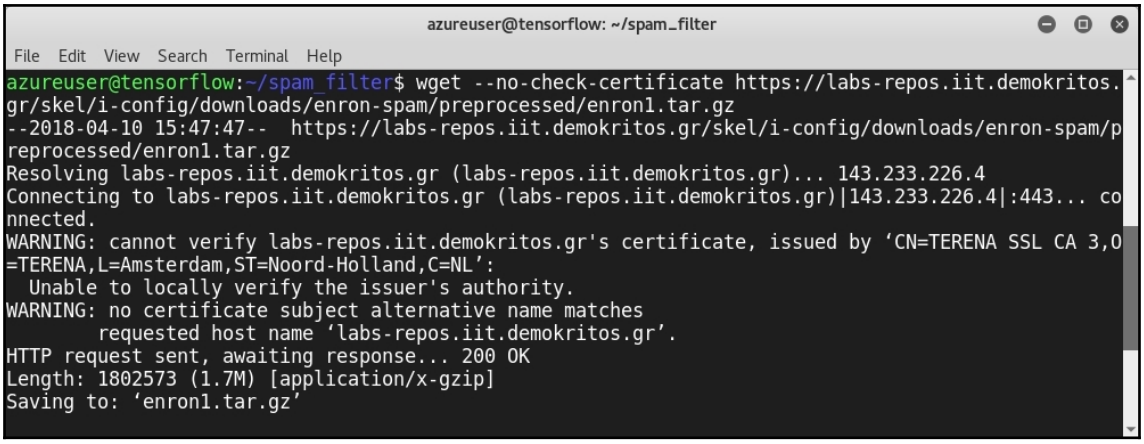

azureuser@tensorflow: ~
File Edit View Search Terminal Help
Packages:
[*] abc..... Australian Broadcasting Commission 2006
[*] alpino..... Alpino Dutch Treebank
[*] averaged_perceptron_tagger Averaged Perceptron Tagger
[ ] averaged_perceptron_tagger_ru Averaged Perceptron Tagger (Russian)
[*] basque_grammars..... Grammars for Basque
[*] biocreative_ppi..... BioCreAtIvE (Critical Assessment of Information
Extraction Systems in Biology)
[*] bllip_wsj_no_aux.... BLLIP Parser: WSJ Model
[*] book_grammars..... Grammars from NLTK Book
[*] brown..... Brown Corpus
[*] brown_tei..... Brown Corpus (TEI XML Version)
[*] cess_cat..... CESS-CAT Treebank
[*] cess_esp..... CESS-ESP Treebank
[*] chat80..... Chat-80 Data Files
[*] city_database..... City Database

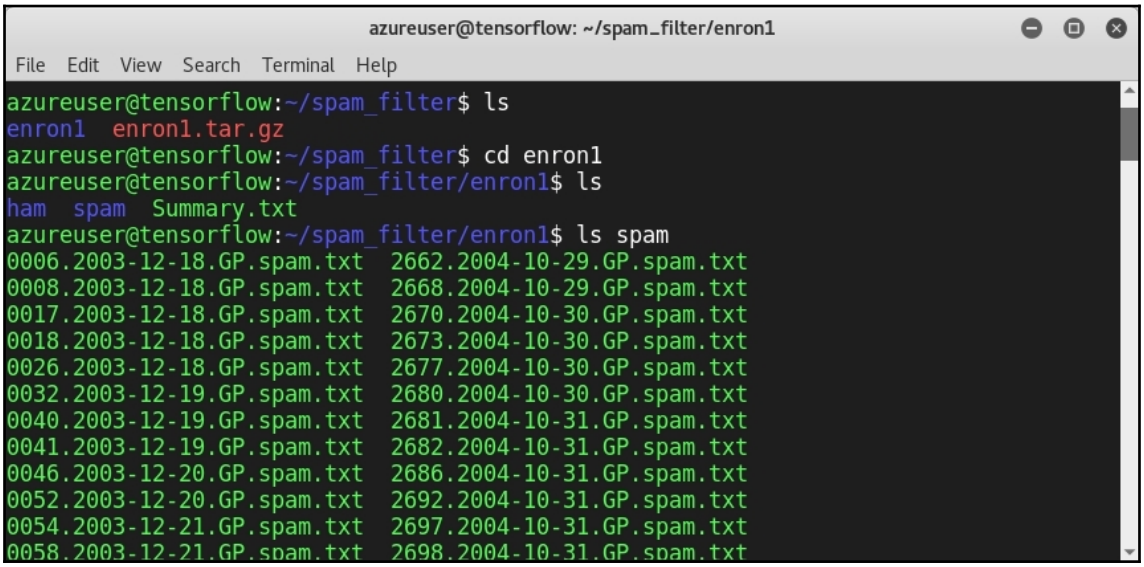

azureuser@tensorflow: ~
File Edit View Search Terminal Help
Python 2.7.12 (default, Dec 4 2017, 14:50:18)
[GCC 5.4.0 20160609] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> from nltk.book import *
*** Introductory Examples for the NLTK Book ***
Loading text1, ..., text9 and sent1, ..., sent9
Type the name of the text or sentence to view it.
Type: 'texts()' or 'sents()' to list the materials.
text1: Moby Dick by Herman Melville 1851
text2: Sense and Sensibility by Jane Austen 1811
text3: The Book of Genesis
text4: Inaugural Address Corpus
text5: Chat Corpus
text6: Monty Python and the Holy Grail
text7: Wall Street Journal
text8: Personals Corpus
text9: The Man Who Was Thursday by G . K . Chesterton 1908
>>>


azureuser@tensorflow: ~
File Edit View Search Terminal Help
azureuser@tensorflow:~$ python
Python 2.7.12 (default, Dec 4 2017, 14:50:18)
[GCC 5.4.0 20160609] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> from urllib import urlopen
>>> url = "https://archive.org/stream/TxtBook-security.in.wireless.ad.hoc.and.se
nsor.networks/Wiley.Security.in.Wireless.Ad.Hoc.and.Sensor.Networks.Mar.2009.eBo
ok-DDU_djvu.txt"
>>> raw = urlopen(url).read()
>>> len(raw)
743800
>>> raw[:50]
'<!DOCTYPE html>\n<html lang="en">\n<!-- __ _ _ _ __|'
>>>


i-config
Internet CONtent FIltering Group
Home
Contact
Research
People
Publications
Downloads
Related Links
Announcements
The wide adoption of the Web and e-mail has raised concerns on the actual use, usability and usefulness of the Internet. On a personal level, users are overwhelmed by the quantity of information that they receive, most of which is of no use to them, e.g. unsolicited e-mail messages. In business, there are serious concerns about misuse of the resources by employees, resulting in increased costs, security risks, legal liability and decreased productivity. Finally, the exposure of minors to inappropriate and potentially harmful content is becoming a serious problem for educational institutions.
The Internet Content Filtering Group (i-config), part of the SKEL laboratory in NCSR "Demokritos", develops intelligent filtering technology, which provides its users with the ability to select content in accordance with their requirements, irrespective of data source or data format. The group combines expertise in a variety of cutting-edge technologies, such as information extraction, information filtering, image analysis, language engineering, knowledge discovery, personalization and Internet technologies.
Major application fields for our technology include:
internet filtering, where the users may choose to fine-tune Web queries or avoid unproductive, illegal or harmful content for themselves or minors under their supervision
e-mail filtering, where the users can train a spam detecting system on their mailboxes so as to direct unsolicited messages into low priority folders.
e-news filtering, where the users can suggest examples of news items closer to their interests and/or the system can deduce such preferences from usage analysis and perform filtering with the users' approval.
corporate filtering, where companies may choose to prevent employees from accessing various external resources and/or dispensing sensitive documents via electronic means, either accidentally or on purpose.
ICRAplus & FilterX
NEW Read about the release of ICRAplus and FilterX
NEW Download and try ICRAplus and FilterX


Name
Last modified
Size
Description
Parent Directory -
enron1.tar.gz 22-Jun-2006 16:24 1.7M
enron2.tar.gz 22-Jun-2006 16:24 2.8M
enron3.tar.gz 22-Jun-2006 16:24 4.4M
enron4.tar.gz 22-Jun-2006 16:24 2.4M
enron5.tar.gz 22-Jun-2006 16:24 2.3M
enron6.tar.gz 22-Jun-2006 16:24 3.0M


azureuser@tensorflow: ~/spam_filter
File Edit View Search Terminal Help
azureuser@tensorflow:~/spam_filter$ wget --no-check-certificate https://labs-repos.iit.demokritos.gr/skel/i-config/downloads/enron-spam/preprocessed/enron1.tar.gz
--2018-04-10 15:47:47-- https://labs-repos.iit.demokritos.gr/skel/i-config/downloads/enron-spam/preprocessed/enron1.tar.gz
Resolving labs-repos.iit.demokritos.gr (labs-repos.iit.demokritos.gr)... 143.233.226.4
Connecting to labs-repos.iit.demokritos.gr (labs-repos.iit.demokritos.gr)|143.233.226.4|:443... connected.
WARNING: cannot verify labs-repos.iit.demokritos.gr's certificate, issued by 'CN=TERENA SSL CA 3,O=TERENA,L=Amsterdam,ST=Noord-Holland,C=NL':
Unable to locally verify the issuer's authority.
WARNING: no certificate subject alternative name matches
requested host name 'labs-repos.iit.demokritos.gr'.
HTTP request sent, awaiting response... 200 OK
Length: 1802573 (1.7M) [application/x-gzip]
Saving to: 'enron1.tar.gz'


azureuser@tensorflow: ~/spam_filter/enron1
File Edit View Search Terminal Help
azureuser@tensorflow:~/spam_filter$ ls
enron1 enron1.tar.gz
azureuser@tensorflow:~/spam_filter$ cd enron1
azureuser@tensorflow:~/spam_filter/enron1$ ls
ham spam Summary.txt
azureuser@tensorflow:~/spam_filter/enron1$ ls spam
0006.2003-12-18.GP.spam.txt 2662.2004-10-29.GP.spam.txt
0008.2003-12-18.GP.spam.txt 2668.2004-10-29.GP.spam.txt
0017.2003-12-18.GP.spam.txt 2670.2004-10-30.GP.spam.txt
0018.2003-12-18.GP.spam.txt 2673.2004-10-30.GP.spam.txt
0026.2003-12-18.GP.spam.txt 2677.2004-10-30.GP.spam.txt
0032.2003-12-19.GP.spam.txt 2680.2004-10-30.GP.spam.txt
0040.2003-12-19.GP.spam.txt 2681.2004-10-31.GP.spam.txt
0041.2003-12-19.GP.spam.txt 2682.2004-10-31.GP.spam.txt
0046.2003-12-20.GP.spam.txt 2686.2004-10-31.GP.spam.txt
0052.2003-12-20.GP.spam.txt 2692.2004-10-31.GP.spam.txt
0054.2003-12-21.GP.spam.txt 2697.2004-10-31.GP.spam.txt
0058.2003-12-21.GP.spam.txt 2698.2004-10-31.GP.spam.txt

```
2577.2000-10-18.farmer.ham.txt  5163.2005-09-06.GP.spam.txt
2578.2000-10-18.farmer.ham.txt  5164.2005-09-06.GP.spam.txt
2579.2000-10-18.farmer.ham.txt  5165.2002-01-09.farmer.ham.txt
2580.2004-10-22.GP.spam.txt     5166.2002-01-09.farmer.ham.txt
2581.2004-10-23.GP.spam.txt     5167.2005-09-06.GP.spam.txt
2582.2000-10-18.farmer.ham.txt  5168.2002-01-10.farmer.ham.txt
2583.2004-10-23.GP.spam.txt     5169.2002-01-11.farmer.ham.txt
2584.2000-10-18.farmer.ham.txt  5170.2005-09-06.GP.spam.txt
2585.2004-10-24.GP.spam.txt     5171.2005-09-06.GP.spam.txt
2586.2000-10-18.farmer.ham.txt  5172.2002-01-11.farmer.ham.txt
```

azureuser@tensorflow: ~/spam_filter/enron1

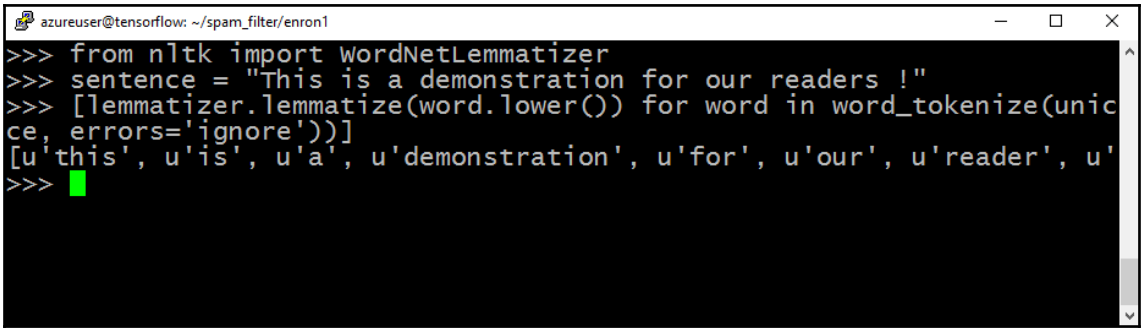
File Edit View Search Terminal Help

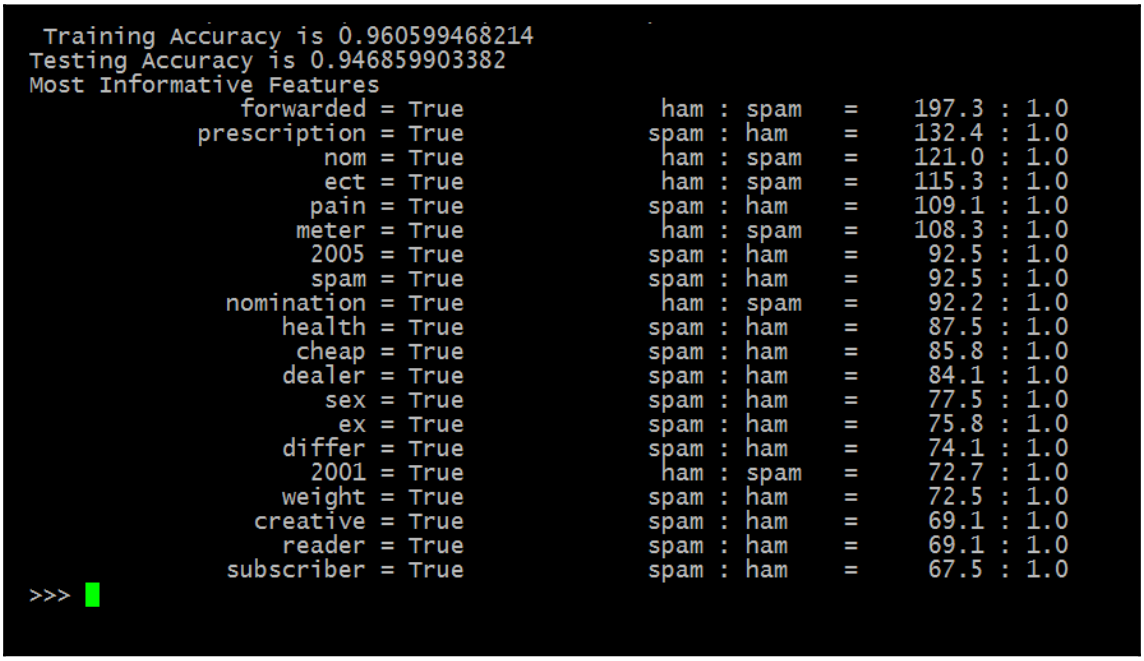
```
import os
import random

#iniate a list called emails_list
emails_list = []
Directory = '/home/azureuser/spam_filter/enron1/emails/'
Dir_list  = os.listdir(Directory)
for file in Dir_list:
        f = open(Directory + file, 'r')
        emails_list.append(f.read())
f.close()



~
"shuffle.py" 14L, 266C                                1,1           All
```

azureuser@tensorflow: ~/spam_filter/enron1

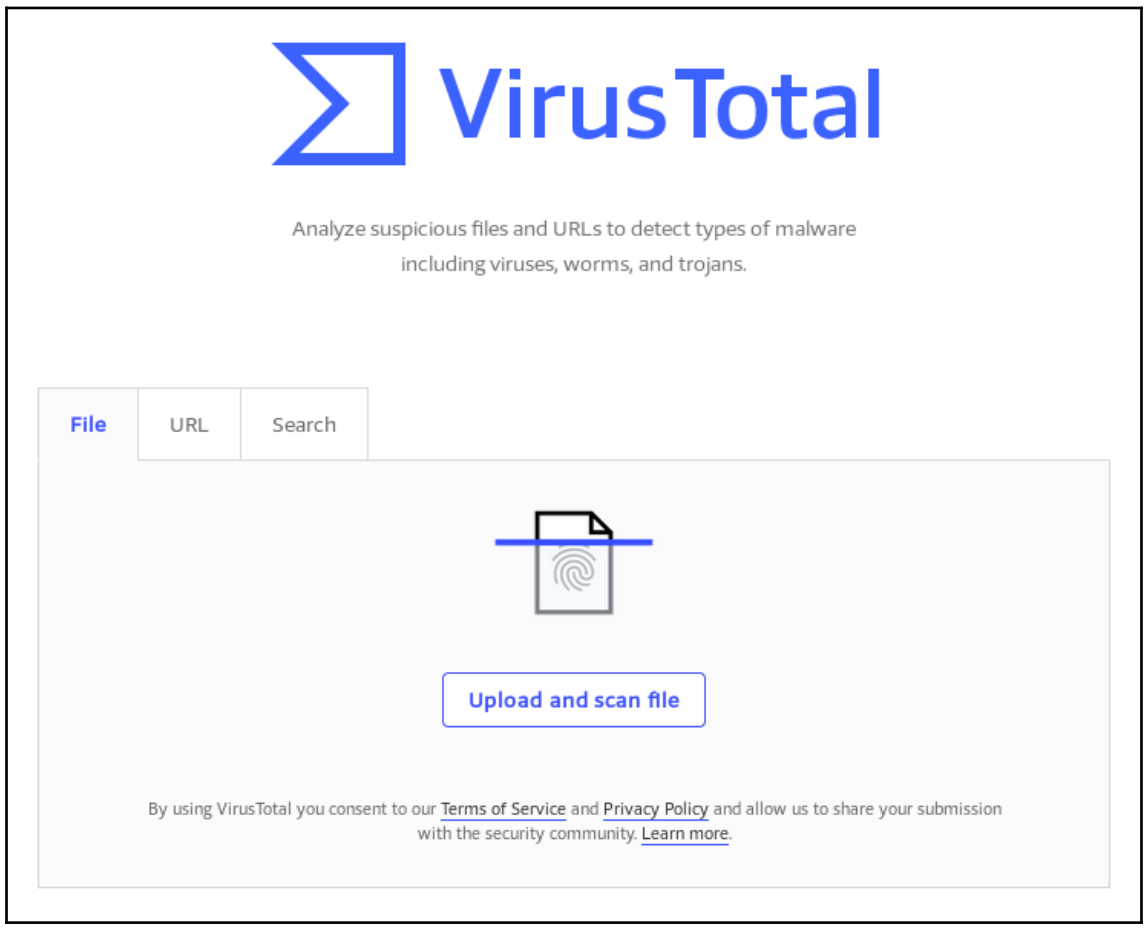
File Edit View Search Terminal Help

```
azureuser@tensorflow:~/spam_filter/enron1$ python
Python 2.7.12 (default, Dec  4 2017, 14:50:18)
[GCC 5.4.0 20160609] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> from nltk import word_tokenize
>>> word_tokenize("Hello Readers! This is tokenizing demonstration")
['Hello', 'Readers', '!', 'This', 'is', 'tokenizing', 'demonstration']
>>>
```


azureuser@tensorflow: ~/spam_filter/enron1
>>> from nltk import WordNetLemmatizer
>>> sentence = "This is a demonstration for our readers !"
>>> [lemmatizer.lemmatize(word.lower()) for word in word_tokenize(unic
ce, errors='ignore'))]
[u'this', u'is', u'a', u'demonstration', u'for', u'our', u'reader', u'
>>>


Training Accuracy is 0.960599468214
Testing Accuracy is 0.946859903382
Most Informative Features
forwarded = True ham : spam = 197.3 : 1.0
prescription = True spam : ham = 132.4 : 1.0
nom = True ham : spam = 121.0 : 1.0
ect = True ham : spam = 115.3 : 1.0
pain = True spam : ham = 109.1 : 1.0
meter = True ham : spam = 108.3 : 1.0
2005 = True spam : ham = 92.5 : 1.0
spam = True spam : ham = 92.5 : 1.0
nomination = True ham : spam = 92.2 : 1.0
health = True spam : ham = 87.5 : 1.0
cheap = True spam : ham = 85.8 : 1.0
dealer = True spam : ham = 84.1 : 1.0
sex = True spam : ham = 77.5 : 1.0
ex = True spam : ham = 75.8 : 1.0
differ = True spam : ham = 74.1 : 1.0
2001 = True ham : spam = 72.7 : 1.0
weight = True spam : ham = 72.5 : 1.0
creative = True spam : ham = 69.1 : 1.0
reader = True spam : ham = 69.1 : 1.0
subscriber = True spam : ham = 67.5 : 1.0
>>>

Chapter 3: Malware Detection with API Calls and PE Headers



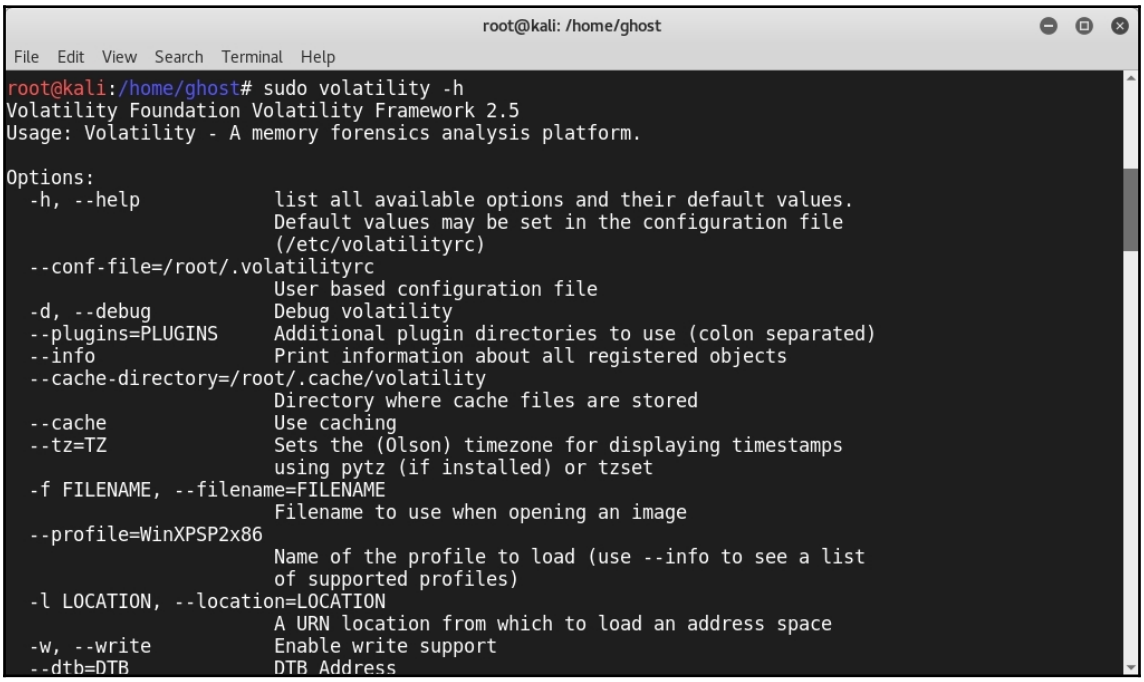
Python

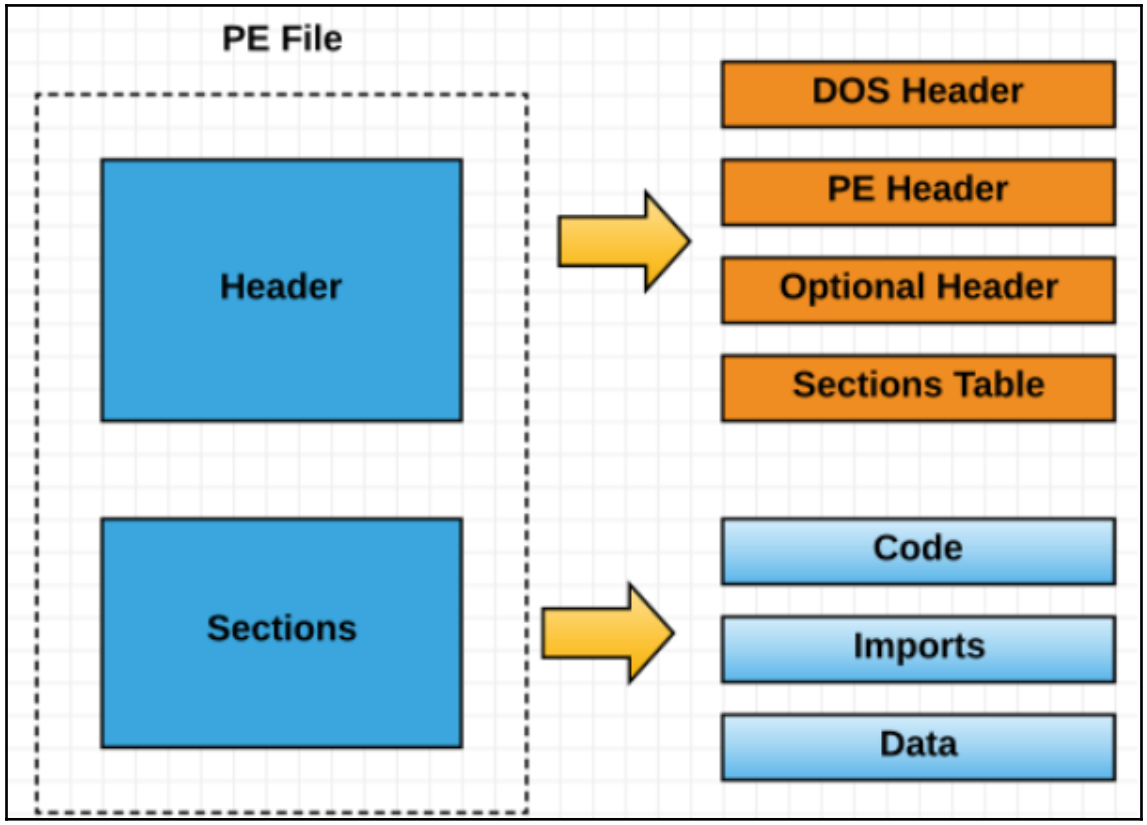
VirusTotal public API version 2.0 implementation in Python 2.x by Chris Clark and Adam Meyers.

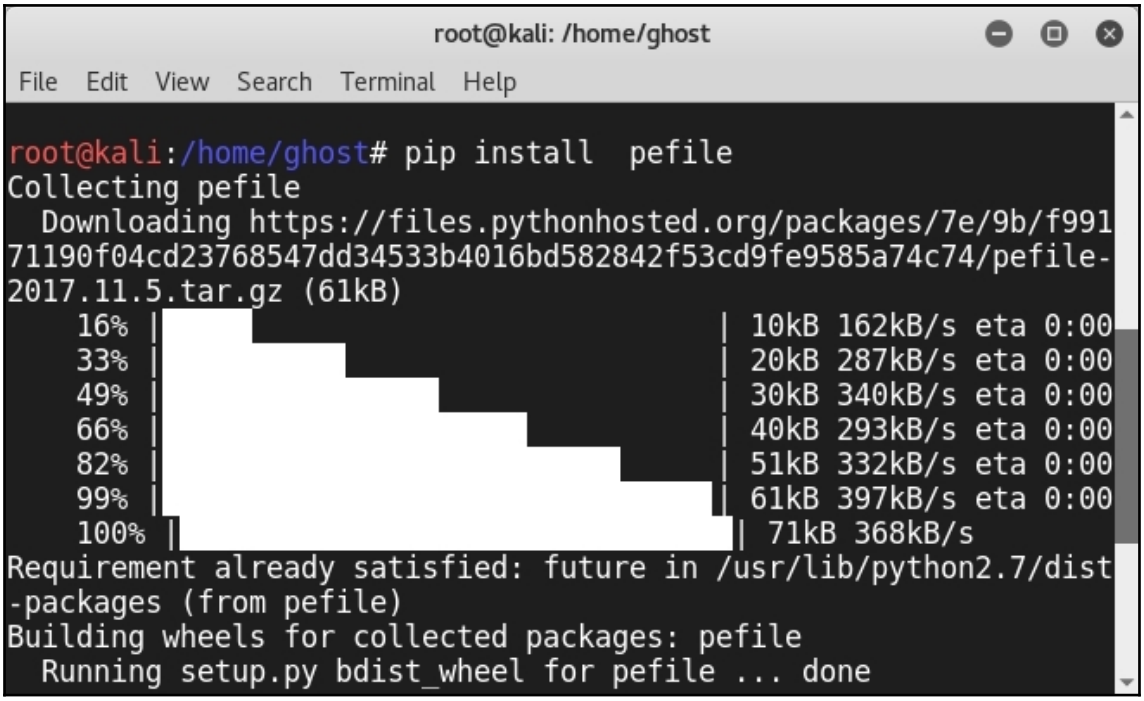
VirusTotal public API version 2.0 implementation in Python 2.x by Gawen Arab.

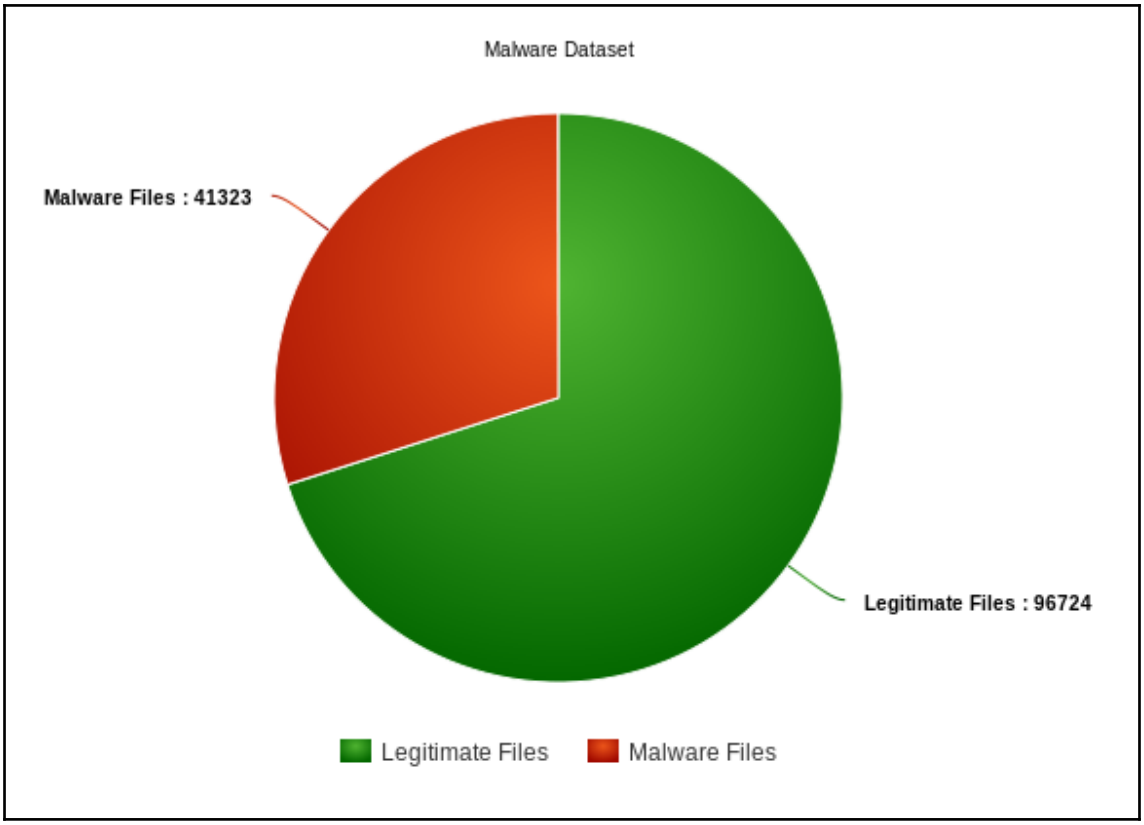
VirusTotal public API version 2.0 implementation in Python 2.x by @techno_vikiing.

Single and bulk lookups with VirusTotal public API version 2.0 by Claudio Guarnieri.

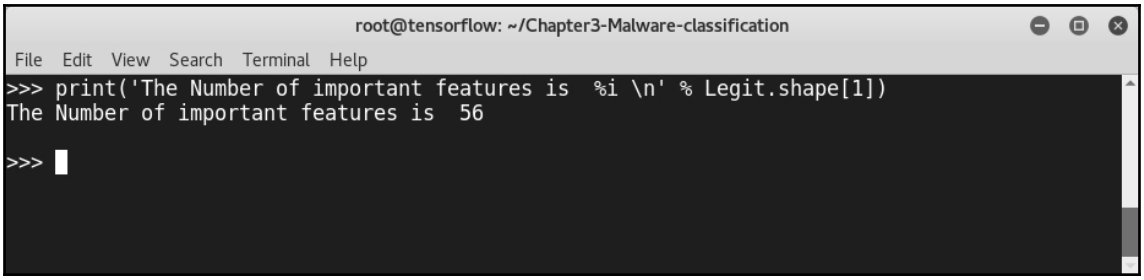


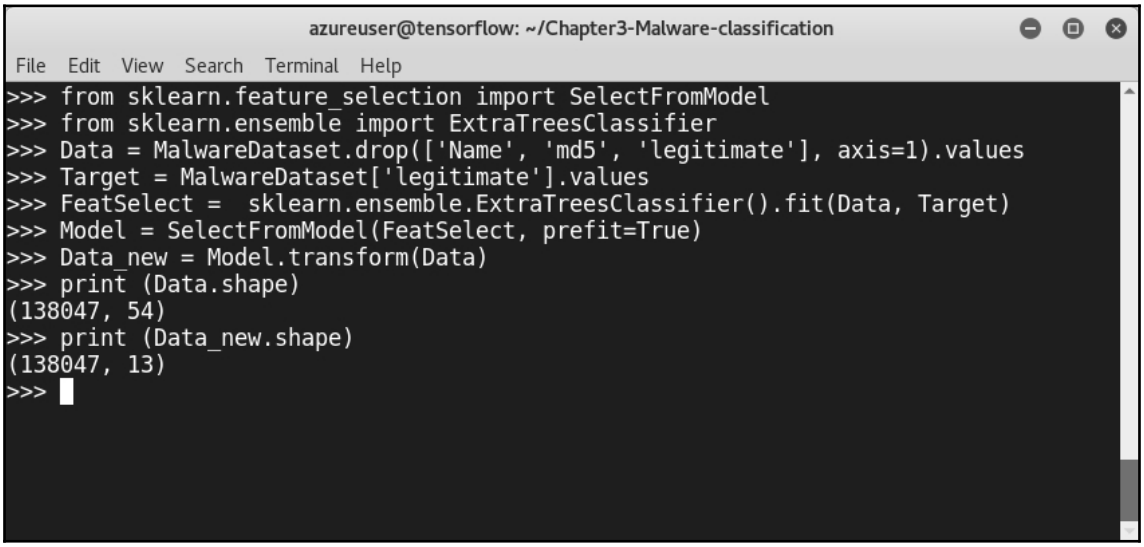

PE File
Header
Sections
DOS Header
PE Header
Optional Header
Sections Table
Code
Imports
Data

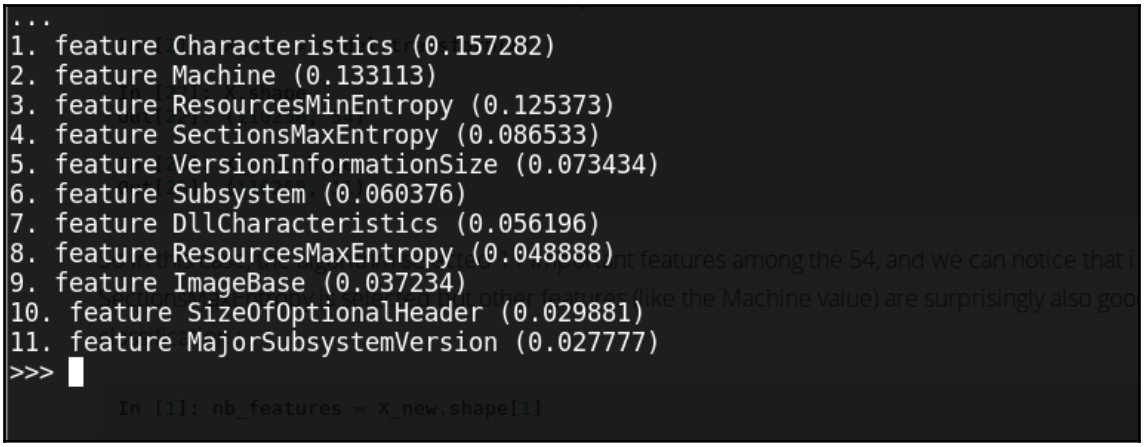

root@kali: /home/ghost
File Edit View Search Terminal Help
root@kali:/home/ghost# pip install pefile
Collecting pefile
Downloading https://files.pythonhosted.org/packages/7e/9b/f99171190f04cd23768547dd34533b4016bd582842f53cd9fe9585a74c74/pefile-2017.11.5.tar.gz (61kB)
16% | | 10kB 162kB/s eta 0:00
33% | | 20kB 287kB/s eta 0:00
49% | | 30kB 340kB/s eta 0:00
66% | | 40kB 293kB/s eta 0:00
82% | | 51kB 332kB/s eta 0:00
99% | | 61kB 397kB/s eta 0:00
100% | | 71kB 368kB/s
Requirement already satisfied: future in /usr/lib/python2.7/dist-packages (from pefile)
Building wheels for collected packages: pefile
Running setup.py bdist_wheel for pefile ... done

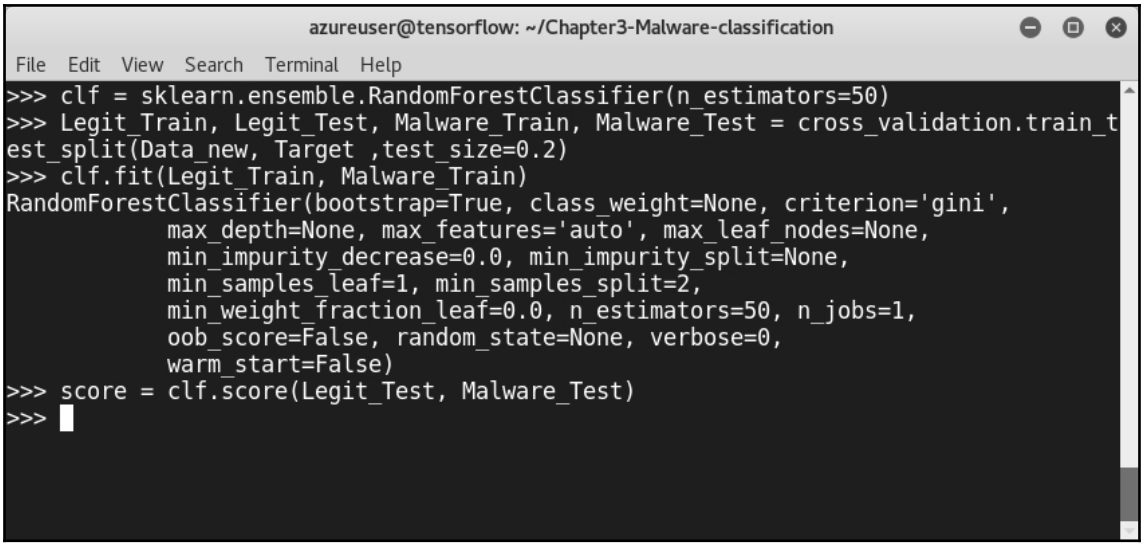


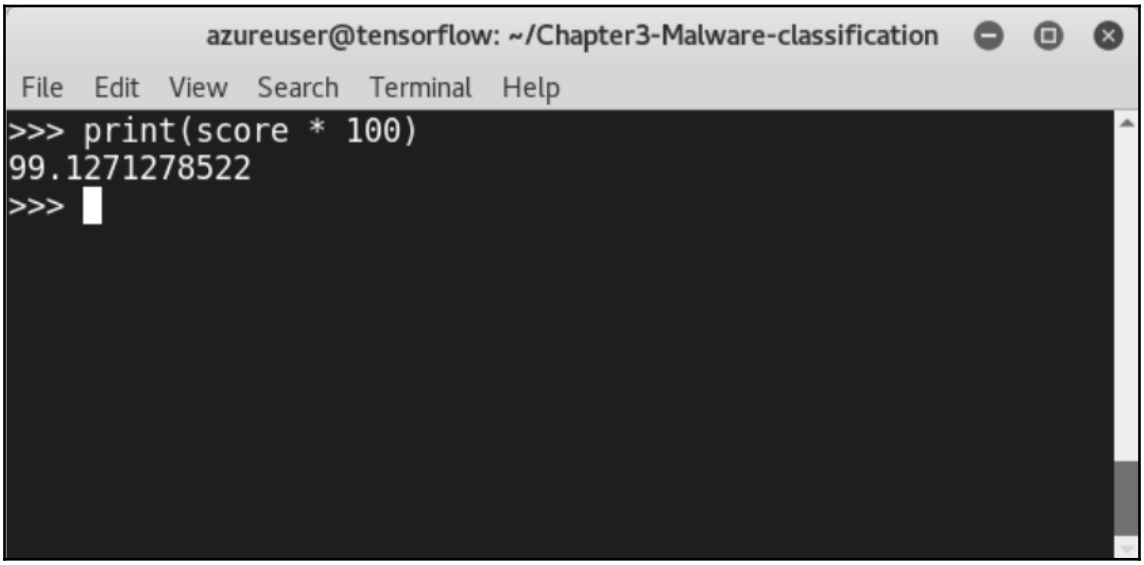

```
azureuser@tensorflow: ~/Chapter3-Malware-classification
File  Edit  View  Search  Terminal  Help
azureuser@tensorflow:~/Chapter3-Malware-classification$ python
Python 2.7.12 (default, Dec  4 2017, 14:50:18)
[GCC 5.4.0 20160609] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> import pandas as pd
>>> MalwareDataset = pd.read_csv('MalwareData.csv', sep='|')
>>> Legit = MalwareDataset[0:41323].drop(['legitimate'], axis=1)
>>> Malware = MalwareDataset[41323::].drop(['legitimate'], axis=1)
>>>
```

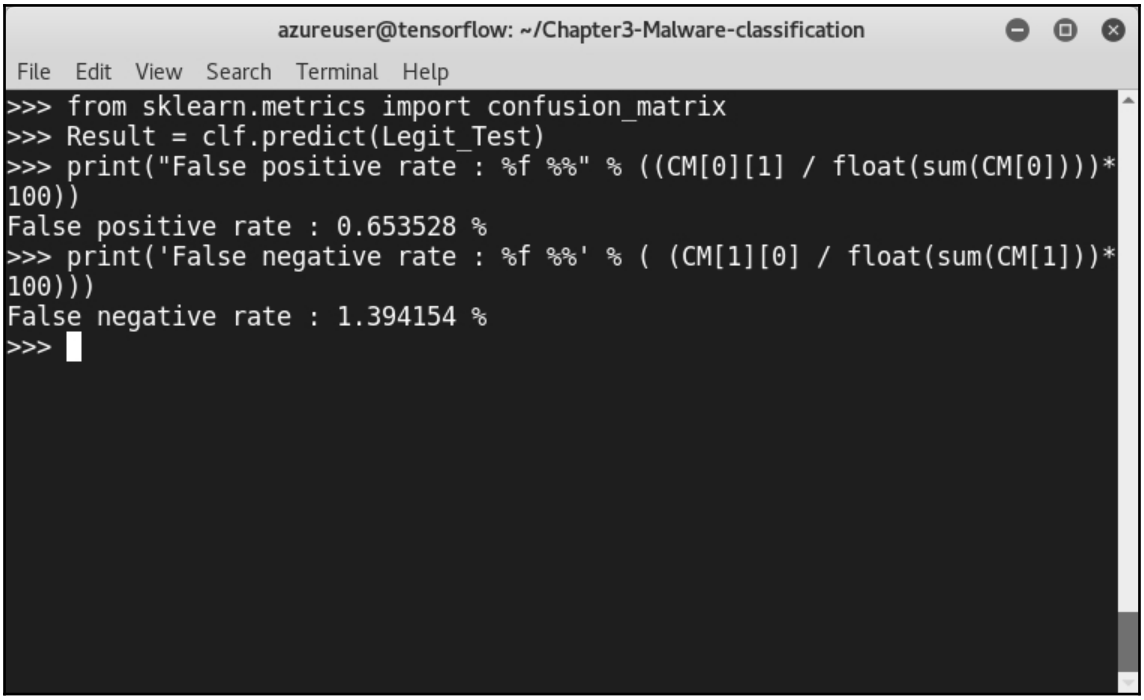

root@tensorflow: ~/Chapter3-Malware-classification
File Edit View Search Terminal Help
>>> print('The Number of important features is %i \n' % Legit.shape[1])
The Number of important features is 56
>>>

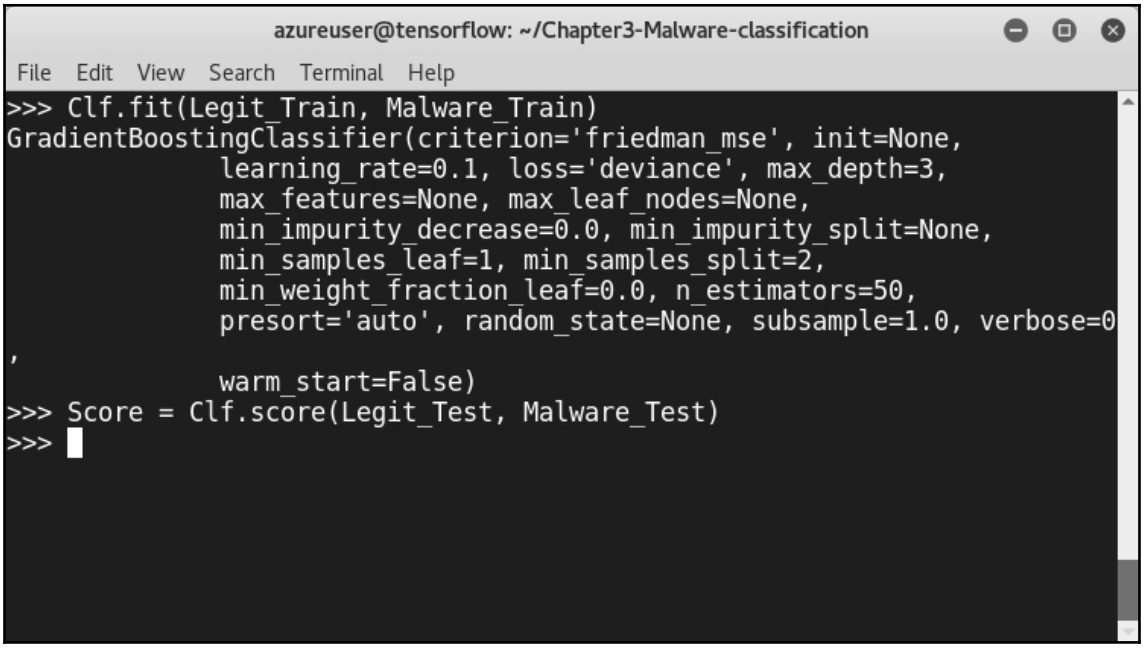

azureuser@tensorflow: ~/Chapter3-Malware-classification
File Edit View Search Terminal Help
>>> from sklearn.feature_selection import SelectFromModel
>>> from sklearn.ensemble import ExtraTreesClassifier
>>> Data = MalwareDataset.drop(['Name', 'md5', 'legitimate'], axis=1).values
>>> Target = MalwareDataset['legitimate'].values
>>> FeatSelect = sklearn.ensemble.ExtraTreesClassifier().fit(Data, Target)
>>> Model = SelectFromModel(FeatSelect, prefit=True)
>>> Data_new = Model.transform(Data)
>>> print (Data.shape)
(138047, 54)
>>> print (Data_new.shape)
(138047, 13)
>>>

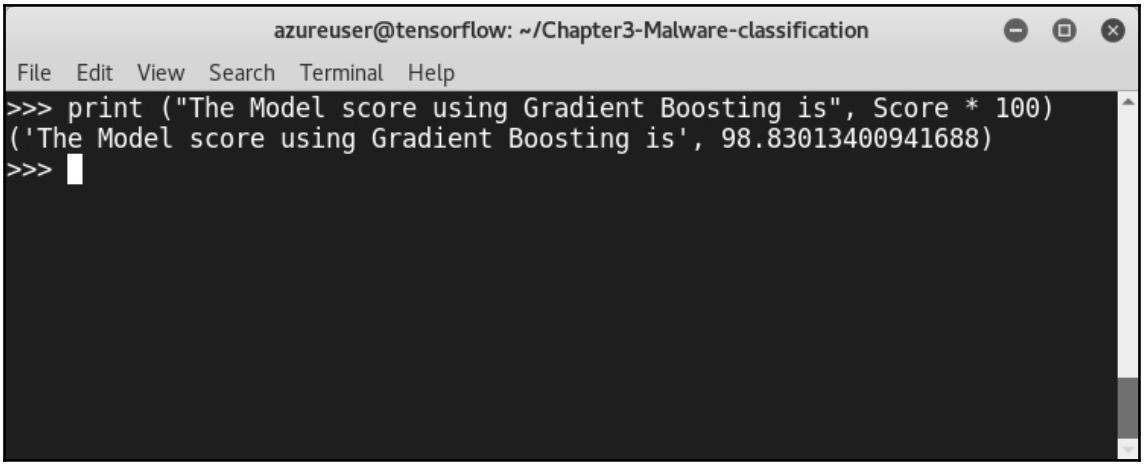

...
1. feature Characteristics (0.157282)
2. feature Machine (0.133113)
3. feature ResourcesMinEntropy (0.125373)
4. feature SectionsMaxEntropy (0.086533)
5. feature VersionInformationSize (0.073434)
6. feature Subsystem (0.060376)
7. feature DllCharacteristics (0.056196)
8. feature ResourcesMaxEntropy (0.048888)
9. feature ImageBase (0.037234)
10. feature SizeOfOptionalHeader (0.029881)
11. feature MajorSubsystemVersion (0.027777)
>>>

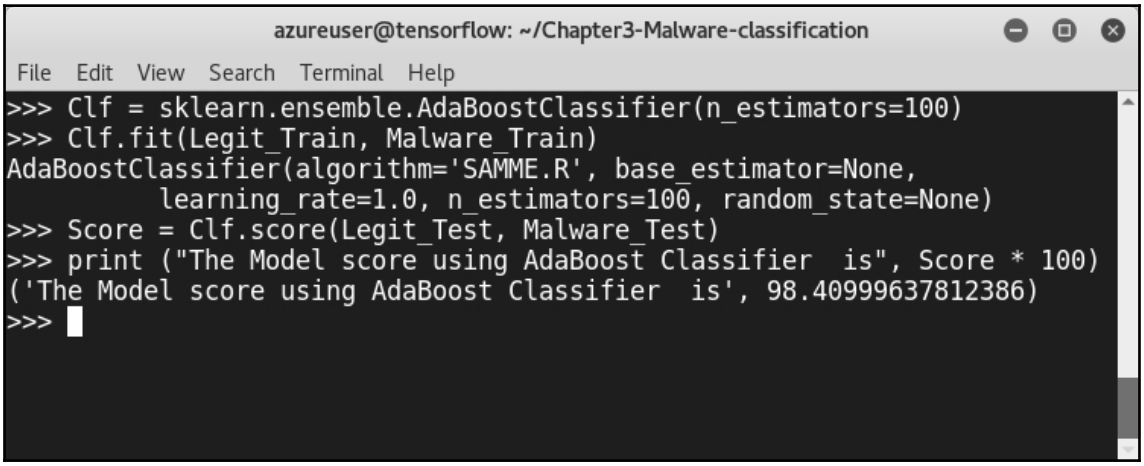

azureuser@tensorflow: ~/Chapter3-Malware-classification
File Edit View Search Terminal Help
>>> clf = sklearn.ensemble.RandomForestClassifier(n_estimators=50)
>>> Legit_Train, Legit_Test, Malware_Train, Malware_Test = cross_validation.train_t
est_split(Data_new, Target ,test_size=0.2)
>>> clf.fit(Legit_Train, Malware_Train)
RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',
max_depth=None, max_features='auto', max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, n_estimators=50, n_jobs=1,
oob_score=False, random_state=None, verbose=0,
warm_start=False)
>>> score = clf.score(Legit_Test, Malware_Test)
>>>


azureuser@tensorflow: ~/Chapter3-Malware-classification
File Edit View Search Terminal Help
>>> print(score * 100)
99.1271278522
>>>


azureuser@tensorflow: ~/Chapter3-Malware-classification
File Edit View Search Terminal Help
>>> from sklearn.metrics import confusion_matrix
>>> Result = clf.predict(Legit_Test)
>>> print("False positive rate : %f %%" % ((CM[0][1] / float(sum(CM[0])))*
100))
False positive rate : 0.653528 %
>>> print('False negative rate : %f %%' % ((CM[1][0] / float(sum(CM[1]))*
100)))
False negative rate : 1.394154 %
>>>

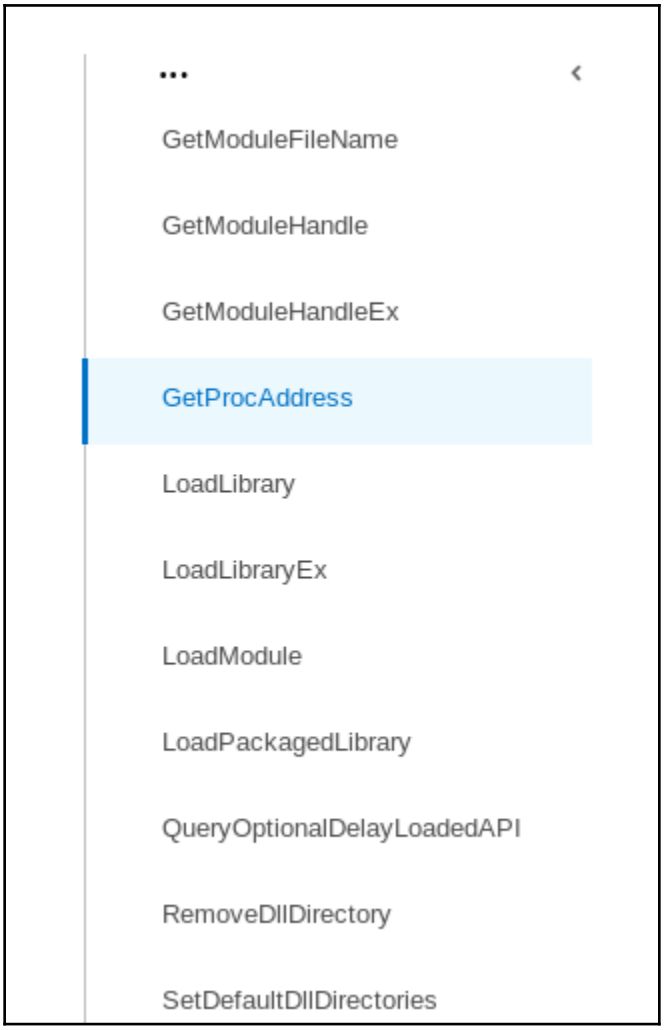

azureuser@tensorflow: ~/Chapter3-Malware-classification
File Edit View Search Terminal Help
>>> Clf.fit(Legit_Train, Malware_Train)
GradientBoostingClassifier(criterion='friedman_mse', init=None,
learning_rate=0.1, loss='deviance', max_depth=3,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, n_estimators=50,
presort='auto', random_state=None, subsample=1.0, verbose=0
,
warm_start=False)
>>> Score = Clf.score(Legit_Test, Malware_Test)
>>>


azureuser@tensorflow: ~/Chapter3-Malware-classification
File Edit View Search Terminal Help
>>> print ("The Model score using Gradient Boosting is", Score * 100)
('The Model score using Gradient Boosting is', 98.83013400941688)
>>>


azureuser@tensorflow: ~/Chapter3-Malware-classification
File Edit View Search Terminal Help
>>> Clf = sklearn.ensemble.AdaBoostClassifier(n_estimators=100)
>>> Clf.fit(Legit_Train, Malware_Train)
AdaBoostClassifier(algorithm='SAMME.R', base_estimator=None,
learning_rate=1.0, n_estimators=100, random_state=None)
>>> Score = Clf.score(Legit_Test, Malware_Test)
>>> print ("The Model score using AdaBoost Classifier is", Score * 100)
('The Model score using AdaBoost Classifier is', 98.40999637812386)
>>>

Imports suspicious APIs

details	RegCloseKey
	OpenProcessToken
	RegOpenKeyExA
	GetFileAttributesA
	VirtualProtect
	GetVersionExA
	GetModuleFileNameA
	GetFileSize
	LockResource
	CreateDirectoryA
	DeleteFileA
	GetCommandLineA
	GetProcAddress
source	Static Parser
relevance	1/10


GetModuleFileName
GetModuleHandle
GetModuleHandleEx
GetProcAddress
LoadLibrary
LoadLibraryEx
LoadModule
LoadPackagedLibrary
QueryOptionalDelayLoadedAPI
RemoveDllDirectory
SetDefaultDllDirectories

messenger

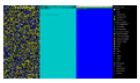
Filename	messenger
Size	1.5MiB (1619766 bytes)
Type	peexe
Description	PE32 executable (GUI) Intel 80386, for MS Windows
Architecture	WINDOWS
SHA256	4c510779ab6a58a3bdbbe8d5f3ec568fcf33df81b0f1a5bdacabf78a9c62f492
Compiler/Packer	Borland Delphi 4.0

Resources

Language	NEUTRAL
Icon	

Visualization

Input File (PortEx)



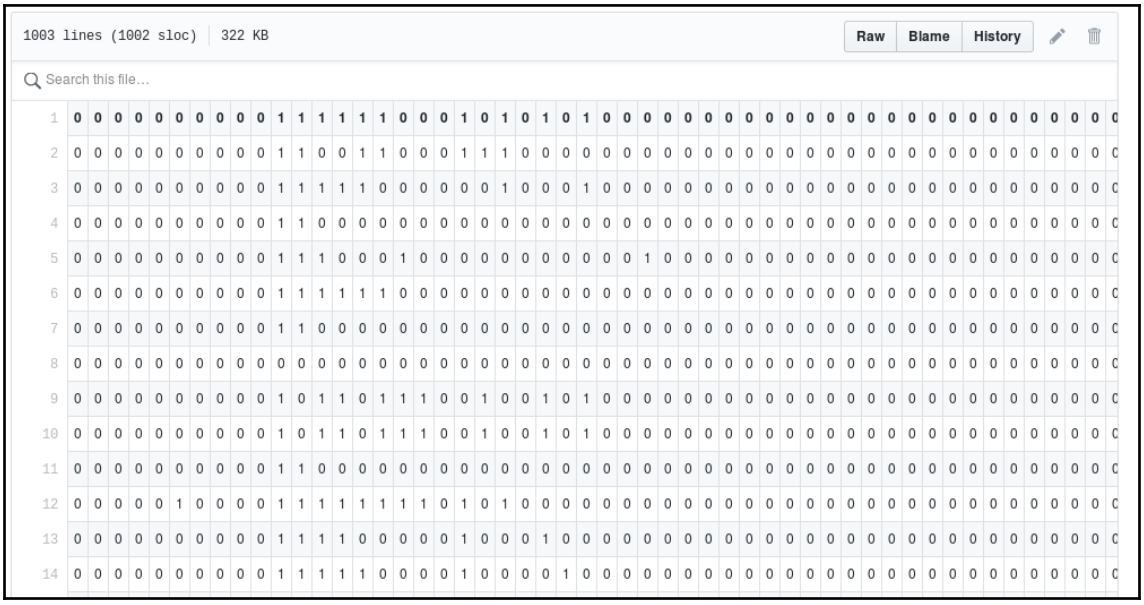
Version Info

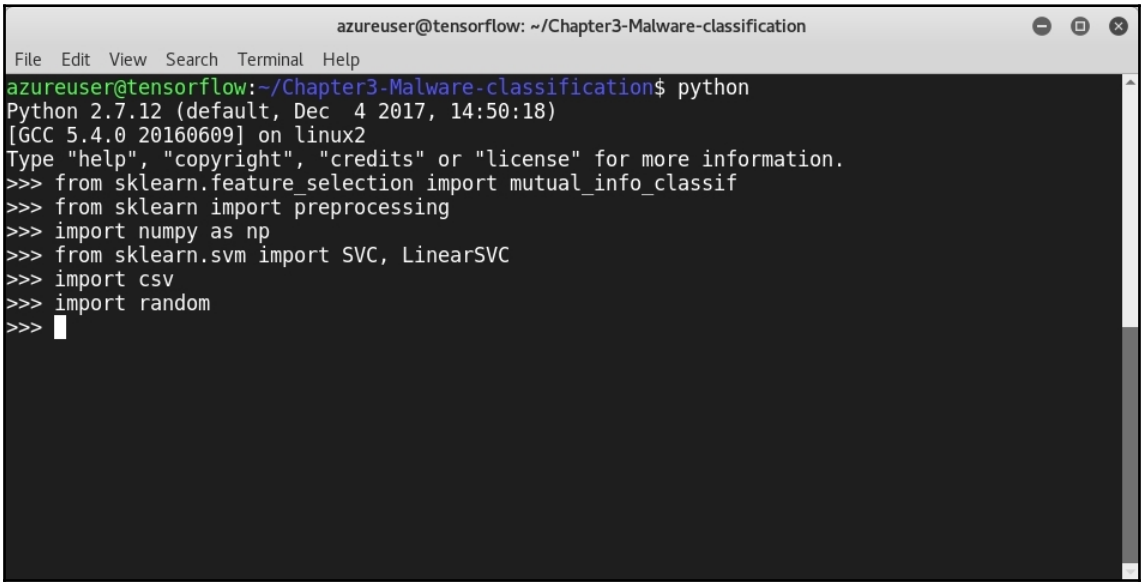
LegalCopyright	-
FileVersion	1.5.2.1
CompanyName	Malisopeha
Comments	This installation was built with Inno Setup.
ProductName	Hinec

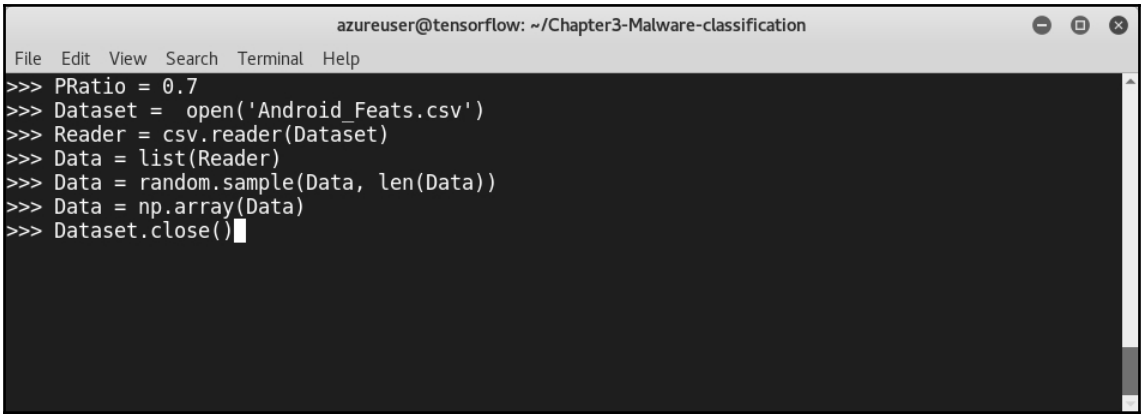
Classification (TrID)

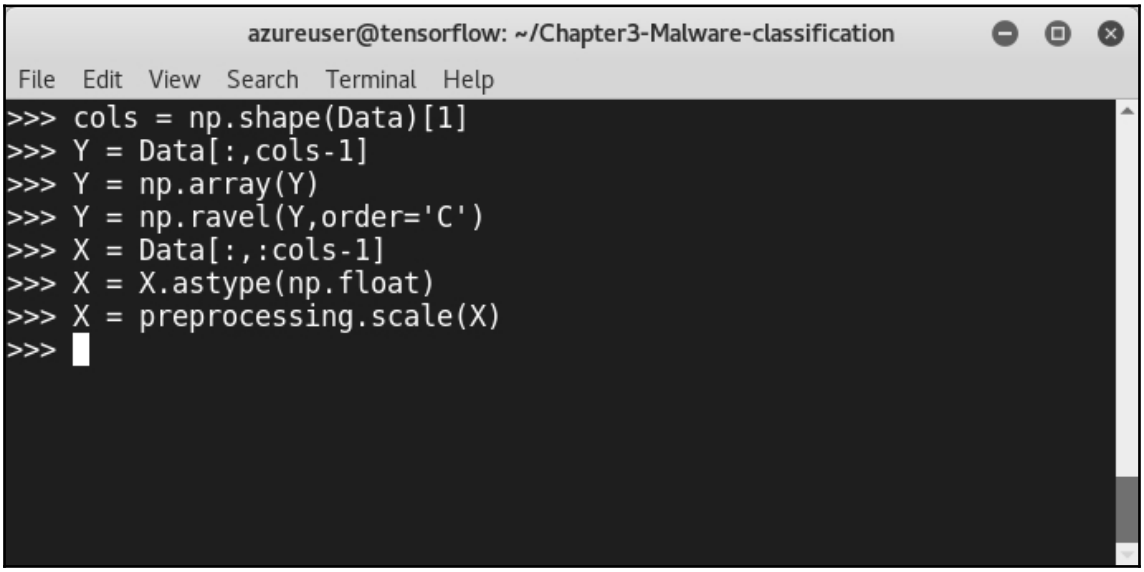
- 86.4% (.EXE) Inno Setup installer
- 5.1% (.DLL) Win32 Dynamic Link Library (generic)
- 3.5% (.EXE) Win32 Executable (generic)
- 1.6% (.EXE) Win16/32 Executable Delphi generic
- 1.5% (.EXE) Generic Win/DOS Executable

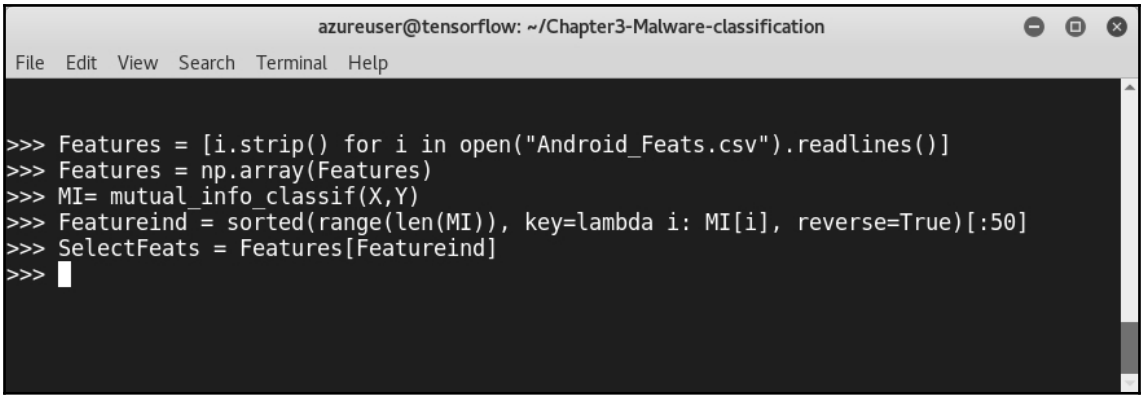
Behaviour	Malware Category	API Function Calls
Behaviour 1	Search Files to Infect	FindClose, FindFirstFile, FindFirstFileEx, FindFirstFileName, TransactedW, FindFirstFileNameW, FindFirstFileTransacted, FindFirstStream, TransactedW, FindFirstStreamW, FindNextFile, FindNextFileNameW, FindNextStreamW, SearchPath.
Behaviour 2	Copy/Delete Files	CloseHandle, CopyFile, CopyFileEx, CopyFileTransacted, CreateFile, CreateFileTransacted, CreateHardLink, CreateHardLink, Transacted, CreateSymbolicLink, CreateSymbolic, LinkTransacted, DeleteFile, DeleteFileTransacted.
Behaviour 3	Get File Information	GetBinaryType, GetCompressed, FileSize, GetCompressedFile, SizeTransacted, GetFileAttributes, GetFileAttributesEx, GetFileAttributes, Transacted, GetFileBandwidth, Reservation, GetFileInformation, ByHandle, GetFileInformation, ByHandleEx, GetFileSize, GetFileSizeEx, GetFileType, GetFinalPathName, ByHandle, GetFullPathName, GetFullPathName, Transacted, GetLongPathName, GetLongPathName, Transacted, GetShortPathName, GetTempFileName, GetTempPath.
Behaviour 4	Move Files	MoveFile, MoveFileEx, MoveFileTransacted, MoveFileWithProgress.
Behaviour 5	Read/Write Files	OpenFile, OpenFileById, ReOpenFile, ReplaceFile, WriteFile, CreateFile, CloseHandle.
Behaviour 6	Change File Attributes	SetFileApisToANSI, SetFileApisToOEM, SetFileAttributes, SetFileAttributesTransacted, SetFileBandwidthReservation, SetFileInformationByHandle, SetFileShortName, SetFileValidData

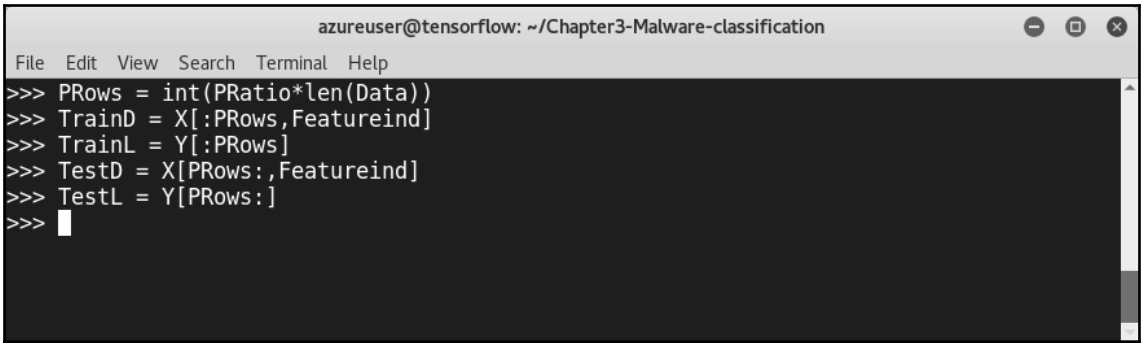

1003 lines (1002 sloc) 322 KB
Raw
Blame
History
Search this file...

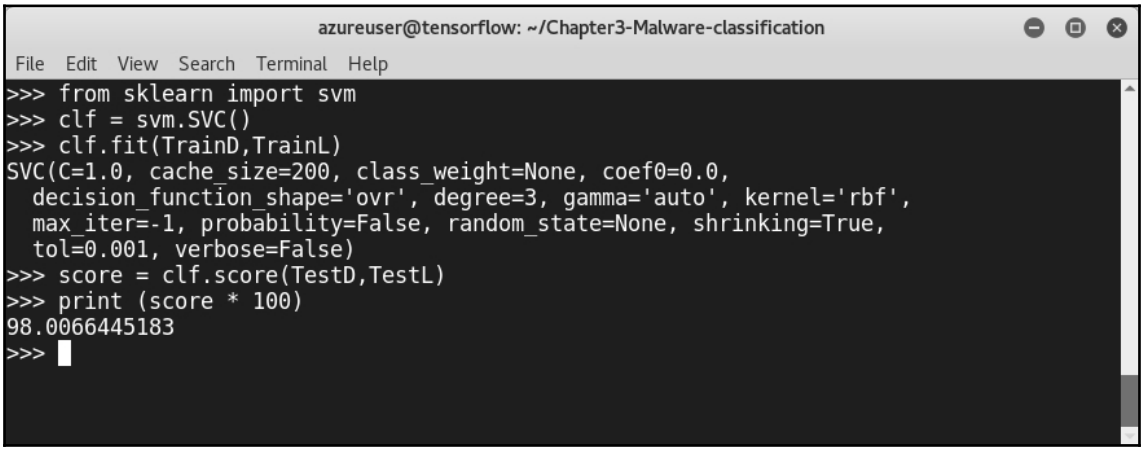

azureuser@tensorflow: ~/Chapter3-Malware-classification
File Edit View Search Terminal Help
azureuser@tensorflow:~/Chapter3-Malware-classification$ python
Python 2.7.12 (default, Dec 4 2017, 14:50:18)
[GCC 5.4.0 20160609] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> from sklearn.feature_selection import mutual_info_classif
>>> from sklearn import preprocessing
>>> import numpy as np
>>> from sklearn.svm import SVC, LinearSVC
>>> import csv
>>> import random
>>>


azureuser@tensorflow: ~/Chapter3-Malware-classification
File Edit View Search Terminal Help
>>> PRatio = 0.7
>>> Dataset = open('Android_Feats.csv')
>>> Reader = csv.reader(Dataset)
>>> Data = list(Reader)
>>> Data = random.sample(Data, len(Data))
>>> Data = np.array(Data)
>>> Dataset.close()

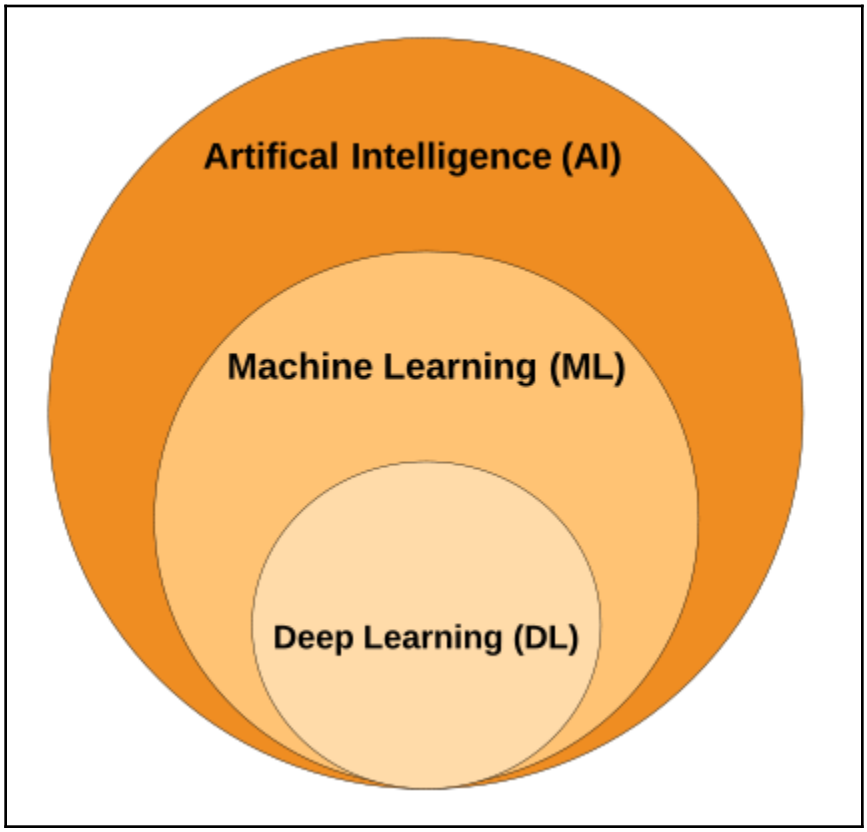

azureuser@tensorflow: ~/Chapter3-Malware-classification
File Edit View Search Terminal Help
>>> cols = np.shape(Data)[1]
>>> Y = Data[:,cols-1]
>>> Y = np.array(Y)
>>> Y = np.ravel(Y,order='C')
>>> X = Data[:,:cols-1]
>>> X = X.astype(np.float)
>>> X = preprocessing.scale(X)
>>>

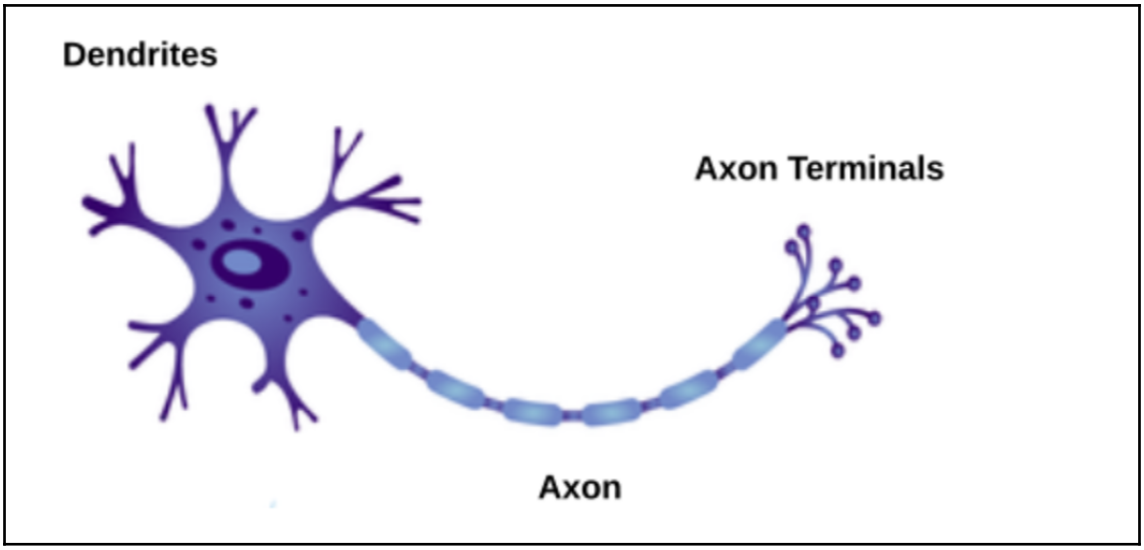

azureuser@tensorflow: ~/Chapter3-Malware-classification
File Edit View Search Terminal Help
>>> Features = [i.strip() for i in open("Android_Feats.csv").readlines()]
>>> Features = np.array(Features)
>>> MI= mutual_info_classif(X,Y)
>>> Featureind = sorted(range(len(MI)), key=lambda i: MI[i], reverse=True)[:50]
>>> SelectFeats = Features[Featureind]
>>>

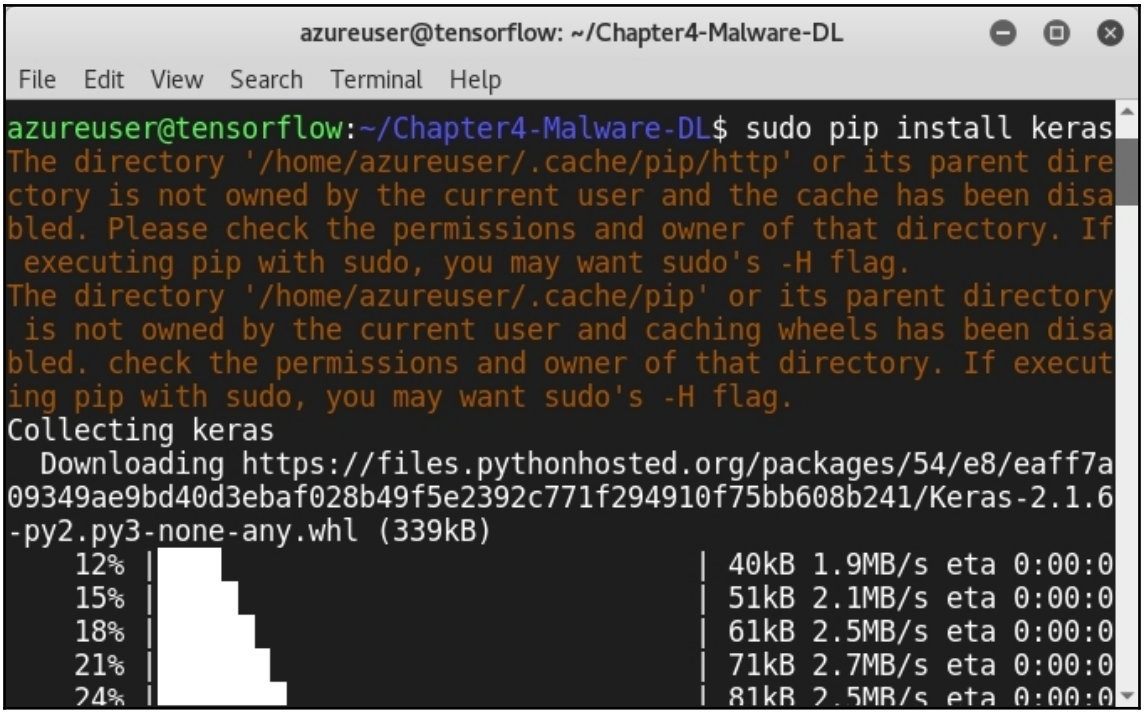

azureuser@tensorflow: ~/Chapter3-Malware-classification
File Edit View Search Terminal Help
>>> PRows = int(PRatio*len(Data))
>>> TrainD = X[:PRows,Featureind]
>>> TrainL = Y[:PRows]
>>> TestD = X[PRows:,Featureind]
>>> TestL = Y[PRows:]
>>>

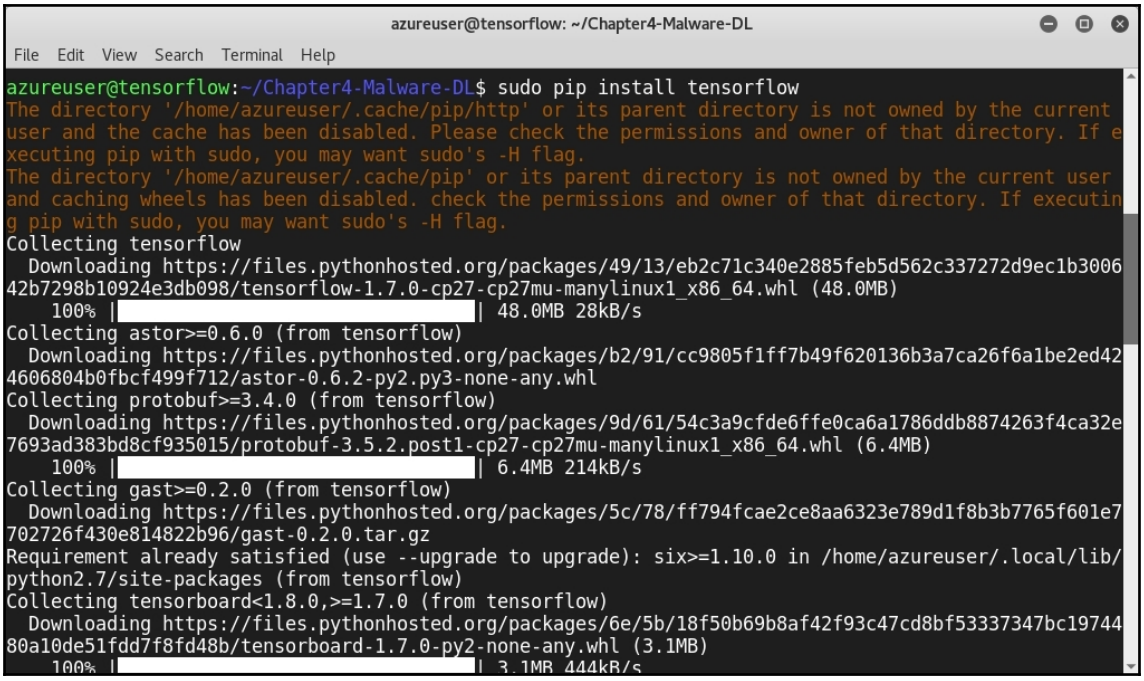

azureuser@tensorflow: ~/Chapter3-Malware-classification
File Edit View Search Terminal Help
>>> from sklearn import svm
>>> clf = svm.SVC()
>>> clf.fit(TrainD,TrainL)
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
decision_function_shape='ovr', degree=3, gamma='auto', kernel='rbf',
max_iter=-1, probability=False, random_state=None, shrinking=True,
tol=0.001, verbose=False)
>>> score = clf.score(TestD,TestL)
>>> print (score * 100)
98.0066445183
>>>

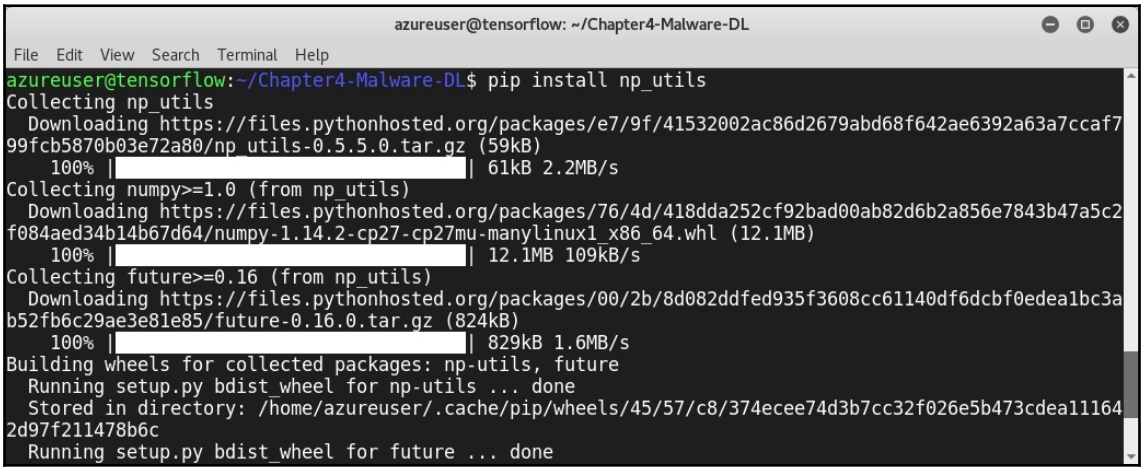
Chapter 4: Malware Detection with Deep Learning

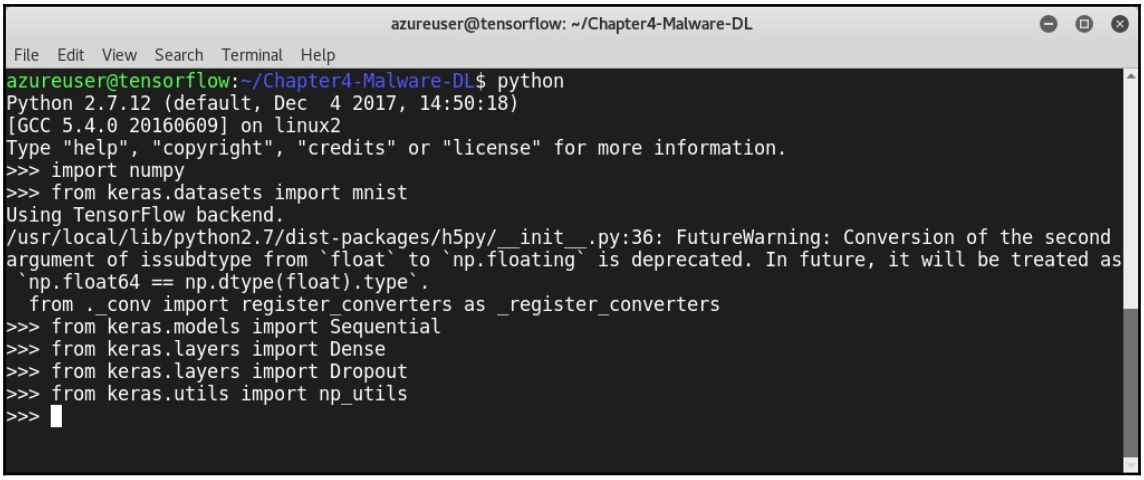


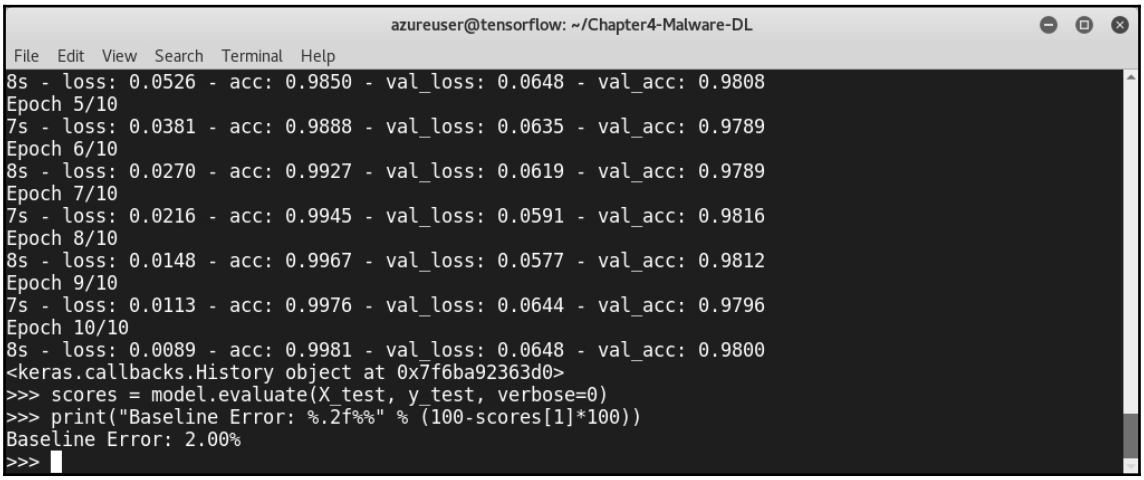

Dendrites
Axon Terminals
Axon


azureuser@tensorflow: ~/Chapter4-Malware-DL
File Edit View Search Terminal Help
azureuser@tensorflow:~/Chapter4-Malware-DL$ sudo pip install keras
The directory '/home/azureuser/.cache/pip/http' or its parent dire
ctory is not owned by the current user and the cache has been disa
bled. Please check the permissions and owner of that directory. If
executing pip with sudo, you may want sudo's -H flag.
The directory '/home/azureuser/.cache/pip' or its parent directory
is not owned by the current user and caching wheels has been disa
bled. check the permissions and owner of that directory. If execut
ing pip with sudo, you may want sudo's -H flag.
Collecting keras
Downloading https://files.pythonhosted.org/packages/54/e8/eaff7a
09349ae9bd40d3ebaf028b49f5e2392c771f294910f75bb608b241/Keras-2.1.6
-py2.py3-none-any.whl (339kB)
12% | | 40kB 1.9MB/s eta 0:00:0
15% | | 51kB 2.1MB/s eta 0:00:0
18% | | 61kB 2.5MB/s eta 0:00:0
21% | | 71kB 2.7MB/s eta 0:00:0
24% | | 81kB 2.5MB/s eta 0:00:0


azureuser@tensorflow: ~/Chapter4-Malware-DL
File Edit View Search Terminal Help
azureuser@tensorflow:~/Chapter4-Malware-DL$ sudo pip install tensorflow
The directory '/home/azureuser/.cache/pip/http' or its parent directory is not owned by the current user and the cache has been disabled. Please check the permissions and owner of that directory. If executing pip with sudo, you may want sudo's -H flag.
The directory '/home/azureuser/.cache/pip' or its parent directory is not owned by the current user and caching wheels has been disabled. check the permissions and owner of that directory. If executing pip with sudo, you may want sudo's -H flag.
Collecting tensorflow
Downloading https://files.pythonhosted.org/packages/49/13/eb2c71c340e2885feb5d562c337272d9ec1b300642b7298b10924e3db098/tensorflow-1.7.0-cp27-cp27mu-manylinux1_x86_64.whl (48.0MB)
100% | | 48.0MB 28kB/s
Collecting astor>=0.6.0 (from tensorflow)
Downloading https://files.pythonhosted.org/packages/b2/91/cc9805f1ff7b49f620136b3a7ca26f6a1be2ed424606804b0fbcf499f712/astor-0.6.2-py2.py3-none-any.whl
Collecting protobuf>=3.4.0 (from tensorflow)
Downloading https://files.pythonhosted.org/packages/9d/61/54c3a9cfde6ffe0ca6a1786ddb8874263f4ca32e7693ad383bd8cf935015/protobuf-3.5.2.post1-cp27-cp27mu-manylinux1_x86_64.whl (6.4MB)
100% | | 6.4MB 214kB/s
Collecting gast>=0.2.0 (from tensorflow)
Downloading https://files.pythonhosted.org/packages/5c/78/ff794fcae2ce8aa6323e789d1f8b3b7765f601e7702726f430e814822b96/gast-0.2.0.tar.gz
Requirement already satisfied (use --upgrade to upgrade): six>=1.10.0 in /home/azureuser/.local/lib/python2.7/site-packages (from tensorflow)
Collecting tensorboard<1.8.0,>=1.7.0 (from tensorflow)
Downloading https://files.pythonhosted.org/packages/6e/5b/18f50b69b8af42f93c47cd8bf53337347bc1974480a10de51fdd7f8fd48b/tensorboard-1.7.0-py2-none-any.whl (3.1MB)
100% | | 3.1MB 444kB/s

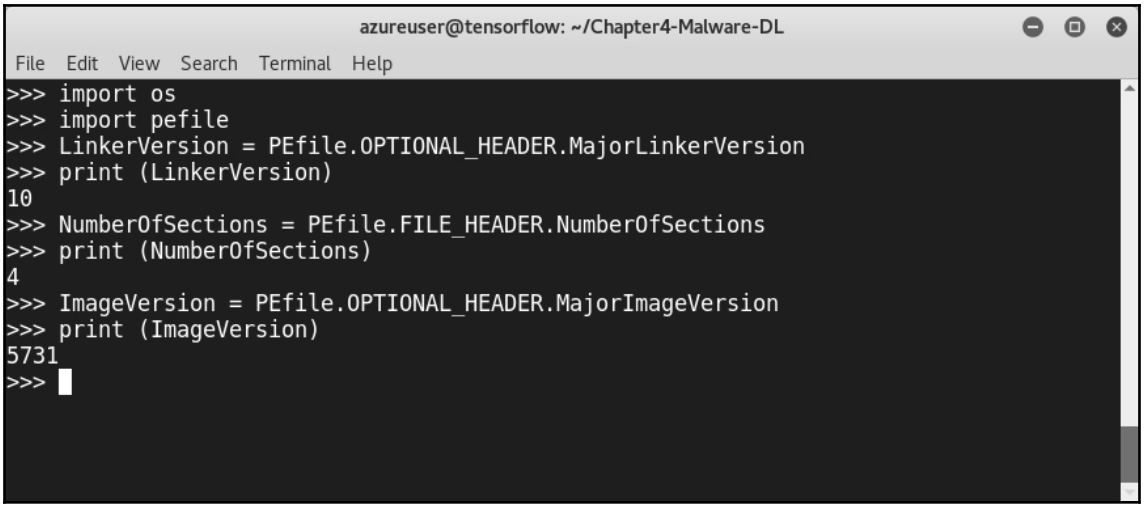

azureuser@tensorflow: ~/Chapter4-Malware-DL
File Edit View Search Terminal Help
azureuser@tensorflow:~/Chapter4-Malware-DL$ pip install np_utils
Collecting np_utils
Downloading https://files.pythonhosted.org/packages/e7/9f/41532002ac86d2679abd68f642ae6392a63a7ccaf799fcb5870b03e72a80/np_utils-0.5.5.0.tar.gz (59kB)
100% | | 61kB 2.2MB/s
Collecting numpy>=1.0 (from np_utils)
Downloading https://files.pythonhosted.org/packages/76/4d/418dda252cf92bad00ab82d6b2a856e7843b47a5c2f084aed34b14b67d64/numpy-1.14.2-cp27-cp27mu-manylinux1_x86_64.whl (12.1MB)
100% | | 12.1MB 109kB/s
Collecting future>=0.16 (from np_utils)
Downloading https://files.pythonhosted.org/packages/00/2b/8d082ddfed935f3608cc61140df6dcbf0edea1bc3ab52fb6c29ae3e81e85/future-0.16.0.tar.gz (824kB)
100% | | 829kB 1.6MB/s
Building wheels for collected packages: np-utils, future
Running setup.py bdist_wheel for np-utils ... done
Stored in directory: /home/azureuser/.cache/pip/wheels/45/57/c8/374ecee74d3b7cc32f026e5b473cdea111642d97f211478b6c
Running setup.py bdist_wheel for future ... done

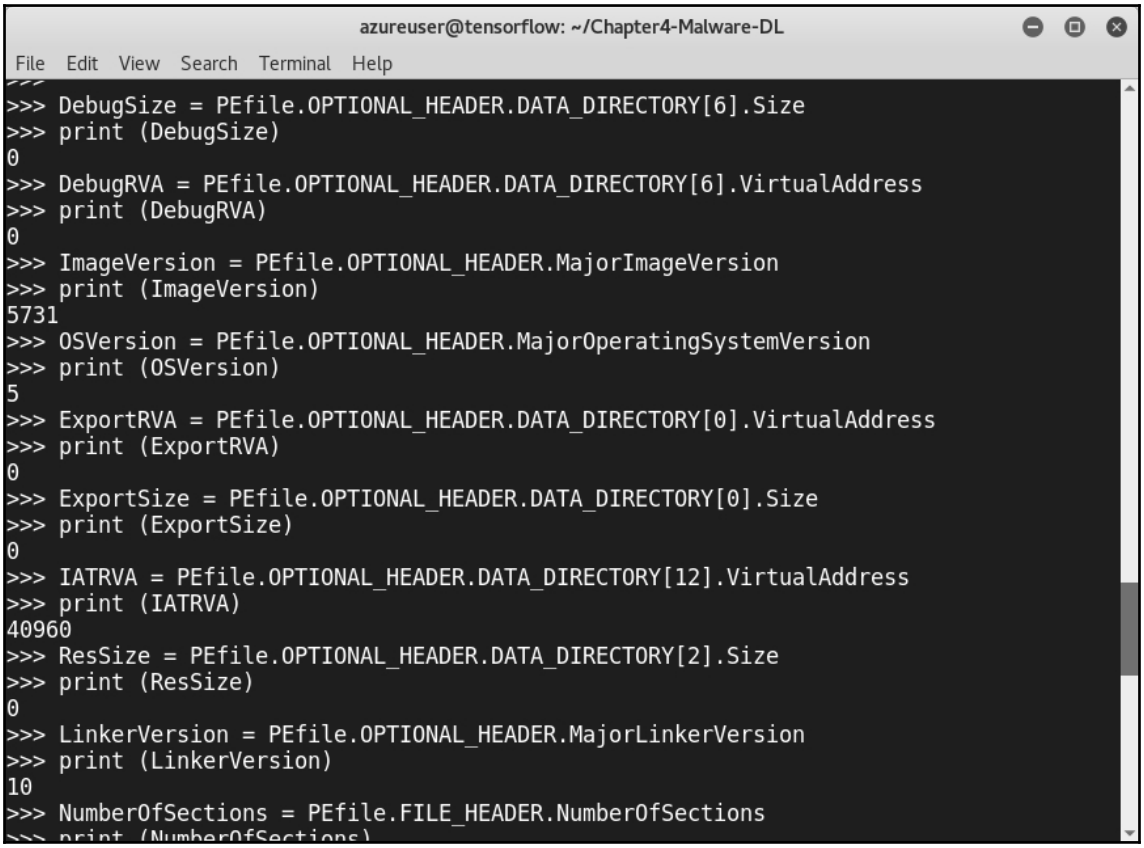

azureuser@tensorflow: ~/Chapter4-Malware-DL
File Edit View Search Terminal Help
azureuser@tensorflow:~/Chapter4-Malware-DL$ python
Python 2.7.12 (default, Dec 4 2017, 14:50:18)
[GCC 5.4.0 20160609] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> import numpy
>>> from keras.datasets import mnist
Using TensorFlow backend.
/usr/local/lib/python2.7/dist-packages/h5py/__init__.py:36: FutureWarning: Conversion of the second argument of issubdtype from `float` to `np.floating` is deprecated. In future, it will be treated as `np.float64 == np.dtype(float).type`.
from ._conv import register_converters as _register_converters
>>> from keras.models import Sequential
>>> from keras.layers import Dense
>>> from keras.layers import Dropout
>>> from keras.utils import np_utils
>>>

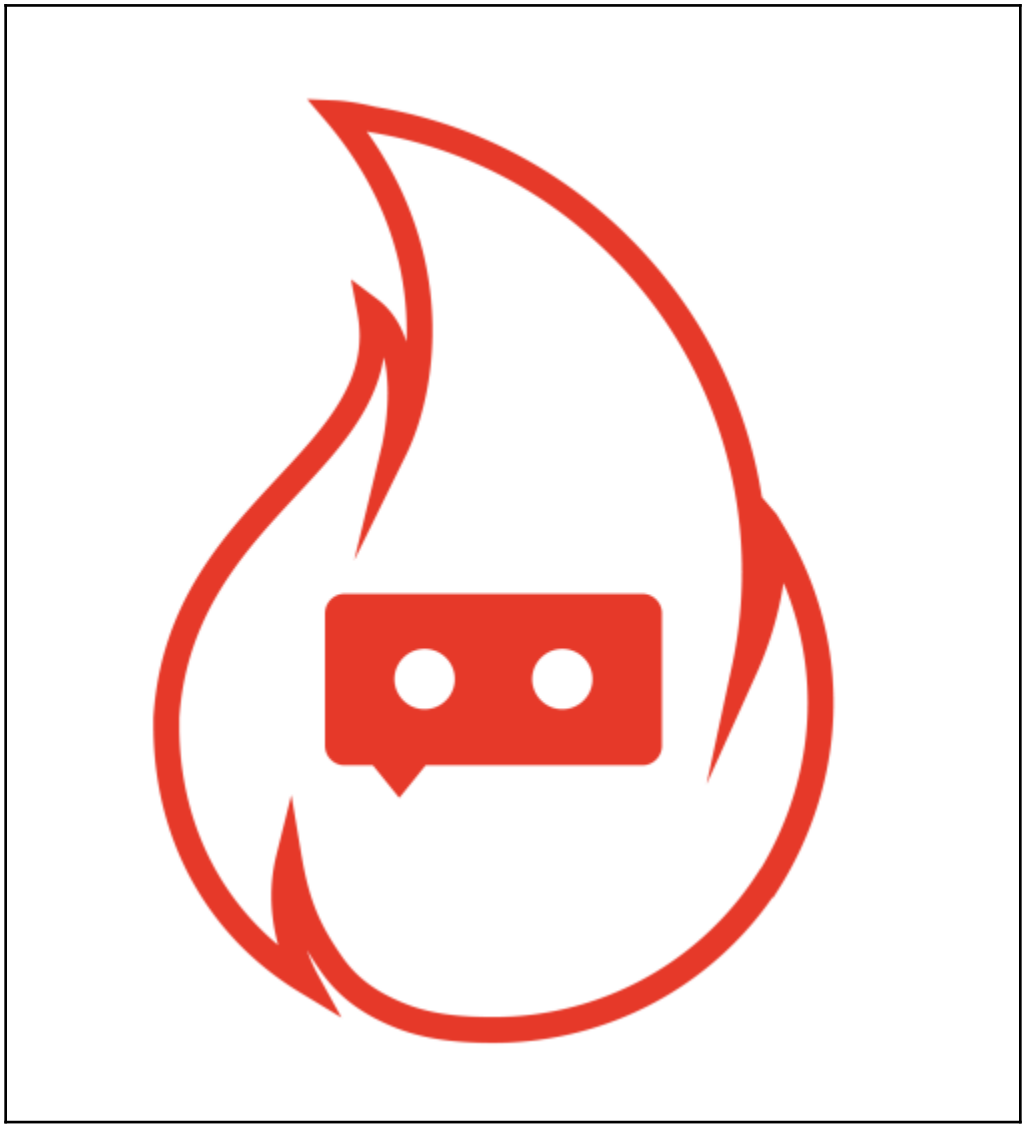

azureuser@tensorflow: ~/Chapter4-Malware-DL
File Edit View Search Terminal Help
8s - loss: 0.0526 - acc: 0.9850 - val_loss: 0.0648 - val_acc: 0.9808
Epoch 5/10
7s - loss: 0.0381 - acc: 0.9888 - val_loss: 0.0635 - val_acc: 0.9789
Epoch 6/10
8s - loss: 0.0270 - acc: 0.9927 - val_loss: 0.0619 - val_acc: 0.9789
Epoch 7/10
7s - loss: 0.0216 - acc: 0.9945 - val_loss: 0.0591 - val_acc: 0.9816
Epoch 8/10
8s - loss: 0.0148 - acc: 0.9967 - val_loss: 0.0577 - val_acc: 0.9812
Epoch 9/10
7s - loss: 0.0113 - acc: 0.9976 - val_loss: 0.0644 - val_acc: 0.9796
Epoch 10/10
8s - loss: 0.0089 - acc: 0.9981 - val_loss: 0.0648 - val_acc: 0.9800
<keras.callbacks.History object at 0x7f6ba92363d0>
>>> scores = model.evaluate(X_test, y_test, verbose=0)
>>> print("Baseline Error: %.2f%%" % (100-scores[1]*100))
Baseline Error: 2.00%
>>>

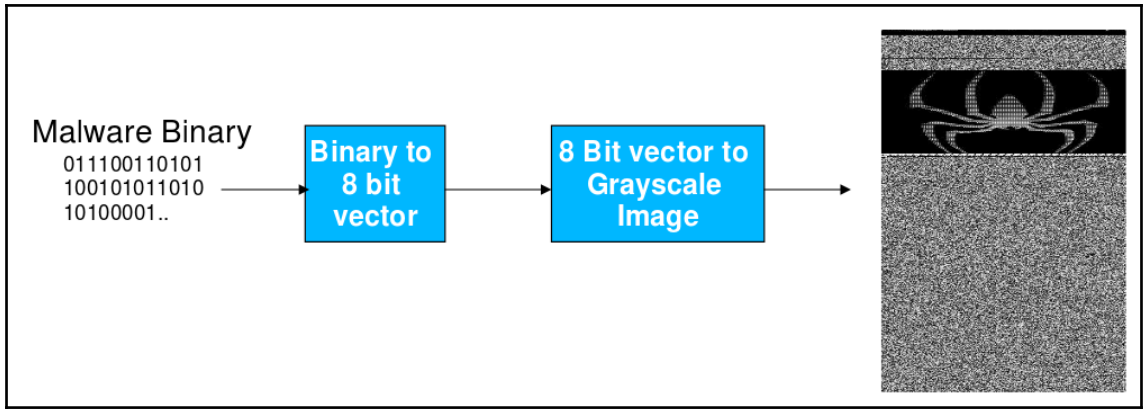
C++

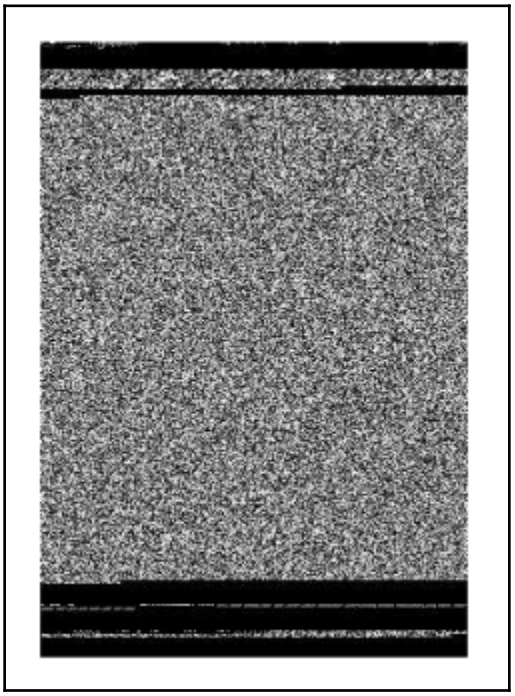
```
typedef struct _IMAGE_OPTIONAL_HEADER {
  WORD                 Magic;
  BYTE                 MajorLinkerVersion;
  BYTE                 MinorLinkerVersion;
  DWORD                SizeOfCode;
  DWORD                SizeOfInitializedData;
  DWORD                SizeOfUninitializedData;
  DWORD                AddressOfEntryPoint;
  DWORD                BaseOfCode;
  DWORD                BaseOfData;
  DWORD                ImageBase;
  DWORD                SectionAlignment;
  DWORD                FileAlignment;
  WORD                 MajorOperatingSystemVersion;
  WORD                 MinorOperatingSystemVersion;
  WORD                 MajorImageVersion;
  WORD                 MinorImageVersion;
  WORD                 MajorSubsystemVersion;
  WORD                 MinorSubsystemVersion;
  DWORD                Win32VersionValue;
  DWORD                SizeOfImage;
  DWORD                SizeOfHeaders;
  DWORD                CheckSum;
  WORD                 Subsystem;
  WORD                 DllCharacteristics;
  DWORD                SizeOfStackReserve;
  DWORD                SizeOfStackCommit;
  DWORD                SizeOfHeapReserve;
  DWORD                SizeOfHeapCommit;
  DWORD                LoaderFlags;
  DWORD                NumberOfRvaAndSizes;
  IMAGE_DATA_DIRECTORY DataDirectory[IMAGE_NUMBEROF_DIRECTORY_ENTRIES];
} IMAGE_OPTIONAL_HEADER, *PIMAGE_OPTIONAL_HEADER;
```

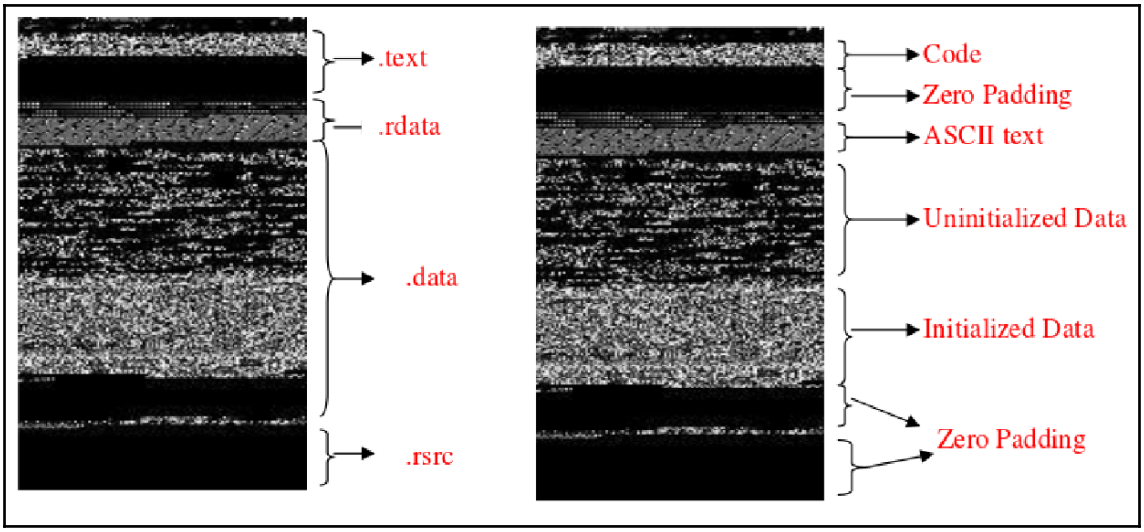

azureuser@tensorflow: ~/Chapter4-Malware-DL
File Edit View Search Terminal Help
>>> import os
>>> import pefile
>>> LinkerVersion = PEfile.OPTIONAL_HEADER.MajorLinkerVersion
>>> print (LinkerVersion)
10
>>> NumberOfSections = PEfile.FILE_HEADER.NumberOfSections
>>> print (NumberOfSections)
4
>>> ImageVersion = PEfile.OPTIONAL_HEADER.MajorImageVersion
>>> print (ImageVersion)
5731
>>>

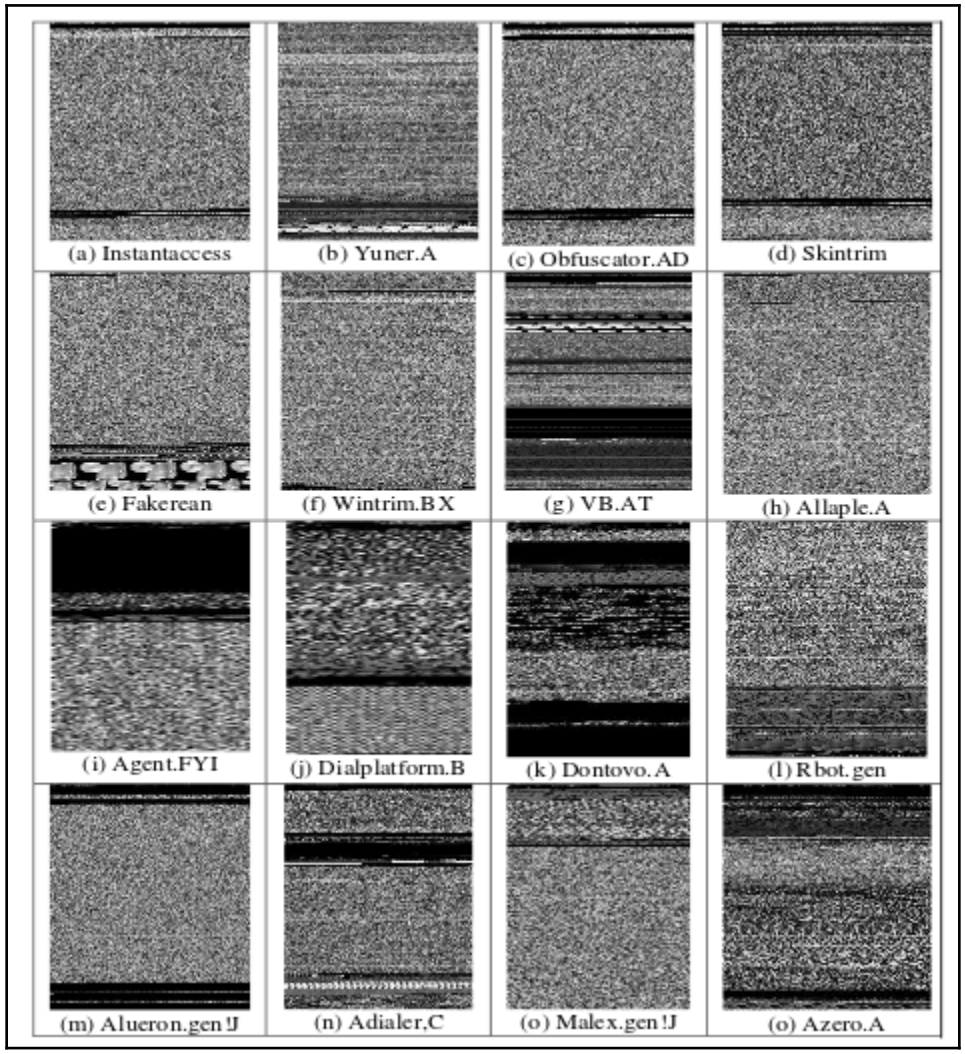

azureuser@tensorflow: ~/Chapter4-Malware-DL
File Edit View Search Terminal Help
>>> DebugSize = PEfile.OPTIONAL_HEADER.DATA_DIRECTORY[6].Size
>>> print (DebugSize)
0
>>> DebugRVA = PEfile.OPTIONAL_HEADER.DATA_DIRECTORY[6].VirtualAddress
>>> print (DebugRVA)
0
>>> ImageVersion = PEfile.OPTIONAL_HEADER.MajorImageVersion
>>> print (ImageVersion)
5731
>>> OSVersion = PEfile.OPTIONAL_HEADER.MajorOperatingSystemVersion
>>> print (OSVersion)
5
>>> ExportRVA = PEfile.OPTIONAL_HEADER.DATA_DIRECTORY[0].VirtualAddress
>>> print (ExportRVA)
0
>>> ExportSize = PEfile.OPTIONAL_HEADER.DATA_DIRECTORY[0].Size
>>> print (ExportSize)
0
>>> IATRVA = PEfile.OPTIONAL_HEADER.DATA_DIRECTORY[12].VirtualAddress
>>> print (IATRVA)
40960
>>> ResSize = PEfile.OPTIONAL_HEADER.DATA_DIRECTORY[2].Size
>>> print (ResSize)
0
>>> LinkerVersion = PEfile.OPTIONAL_HEADER.MajorLinkerVersion
>>> print (LinkerVersion)
10
>>> NumberOfSections = PEfile.FILE_HEADER.NumberOfSections

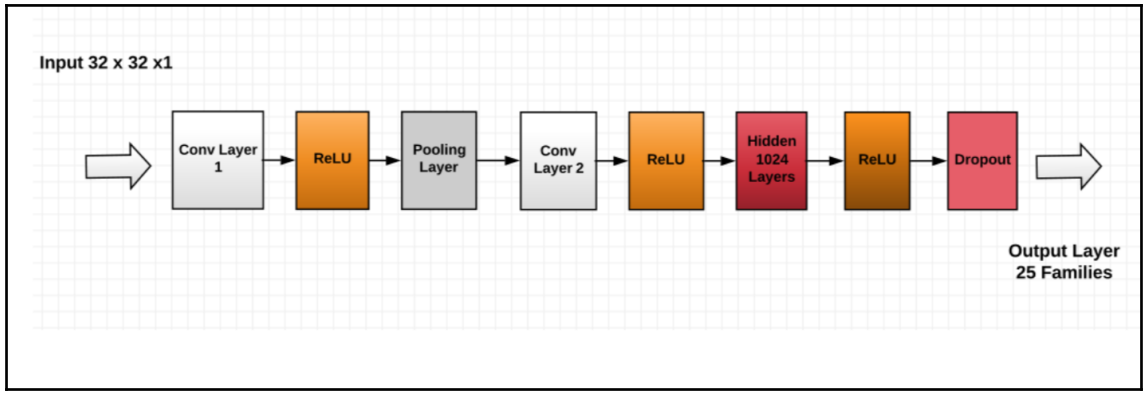



Malware Binary
011100110101
100101011010
10100001..
Binary to 8 bit vector
8 Bit vector to Grayscale Image

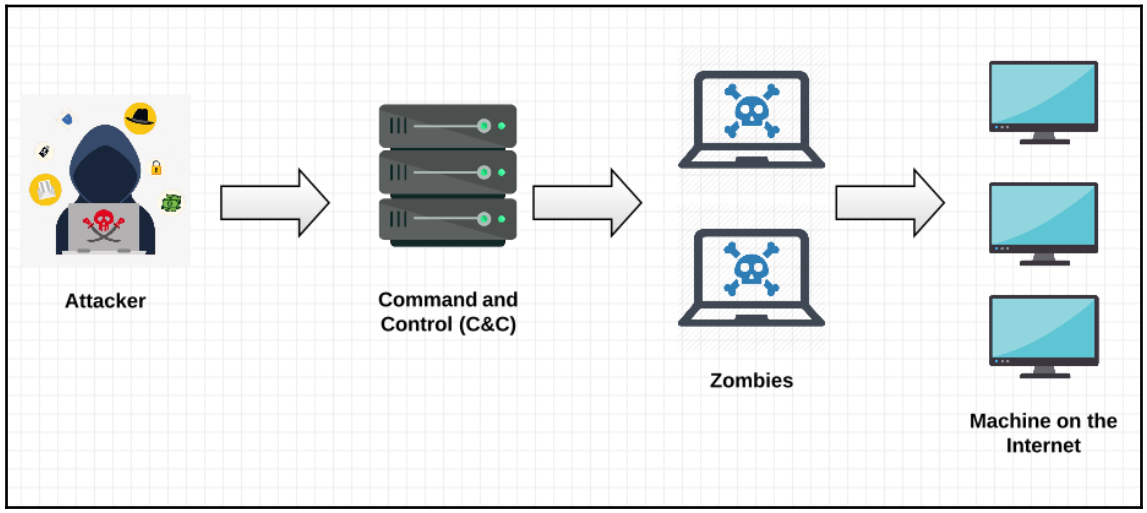


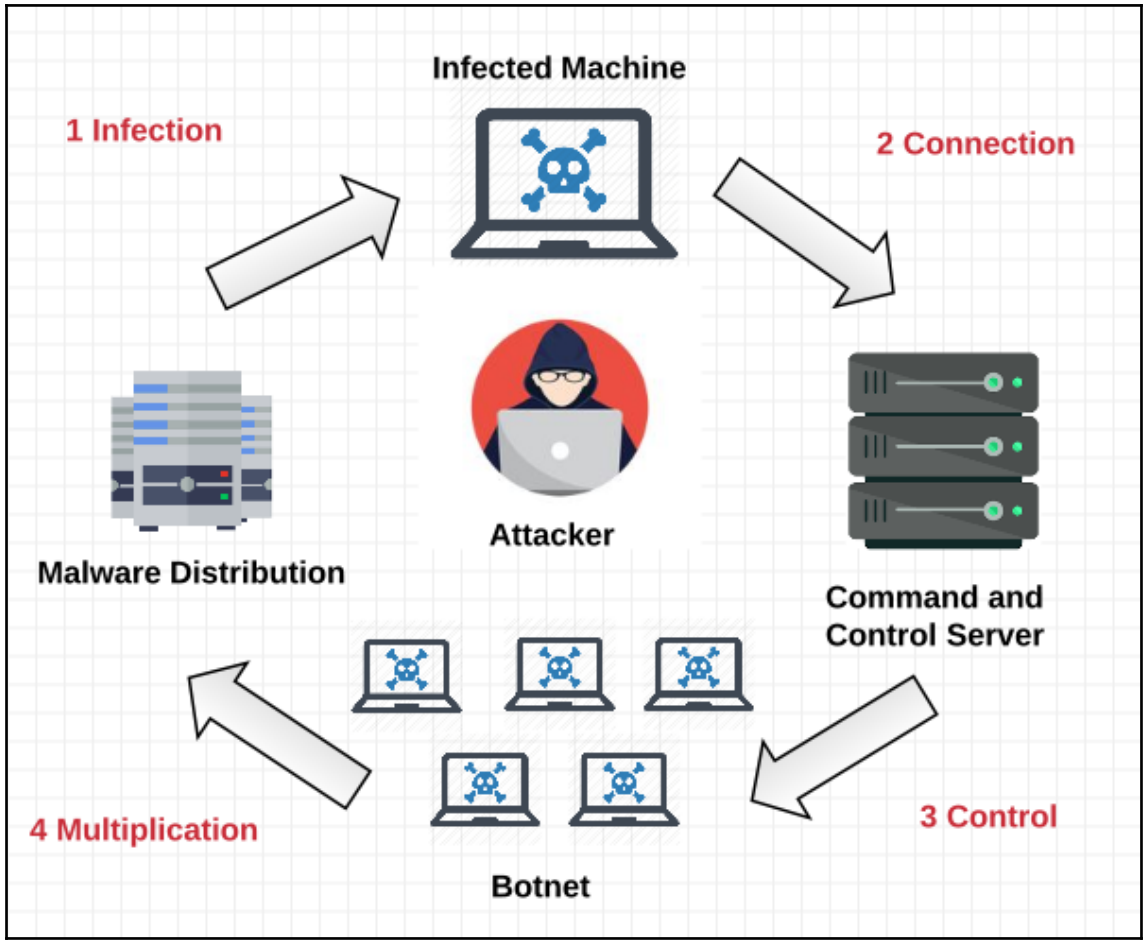

.text
.rdata
.data
.rsrc
Code
Zero Padding
ASCII text
Uninitialized Data
Initialized Data
Zero Padding

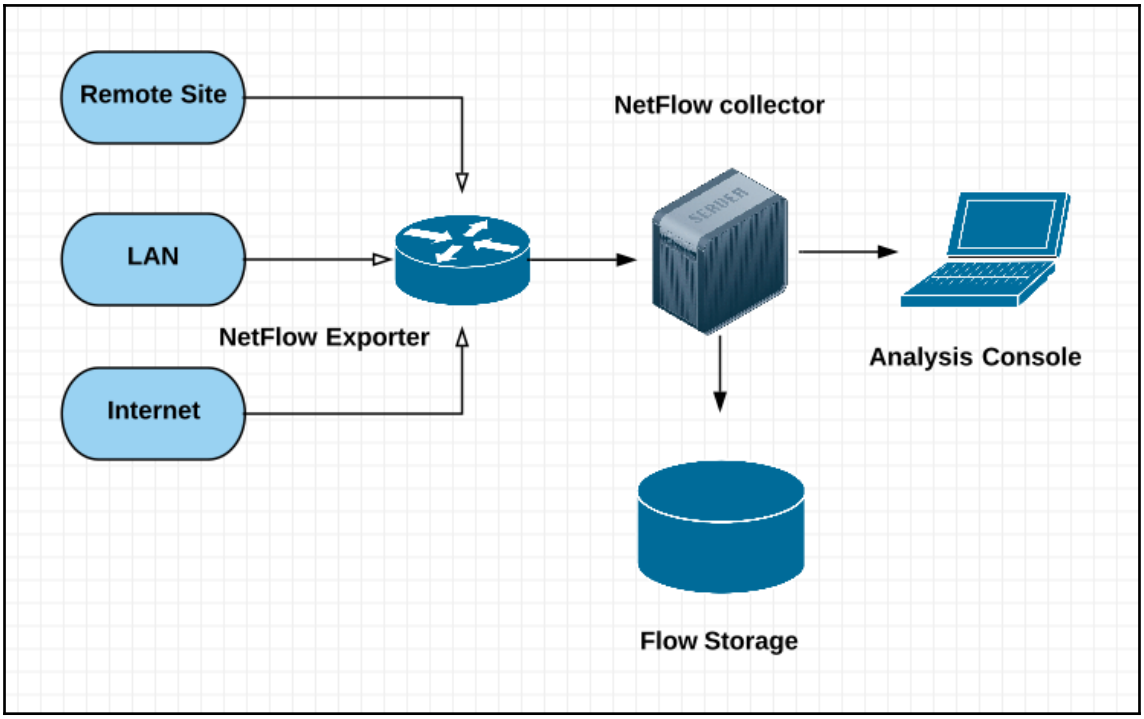

(a) Instantaccess
(b) Yuner.A
(c) Obfuscator.AD
(d) Skintrim
(e) Fakerean
(f) Wintrim.BX
(g) VB.AT
(h) Allaple.A
(i) Agent.FYI
(j) Dialplatform.B
(k) Dontovo.A
(l) Rbot.gen
(m) Alueron.gen!J
(n) Adialer,C
(o) Malex.gen!J
(o) Azero.A


Input 32 x 32 x1
Conv Layer 1
ReLU
Pooling Layer
Conv Layer 2
ReLU
Hidden 1024 Layers
ReLU
Dropout
Output Layer 25 Families

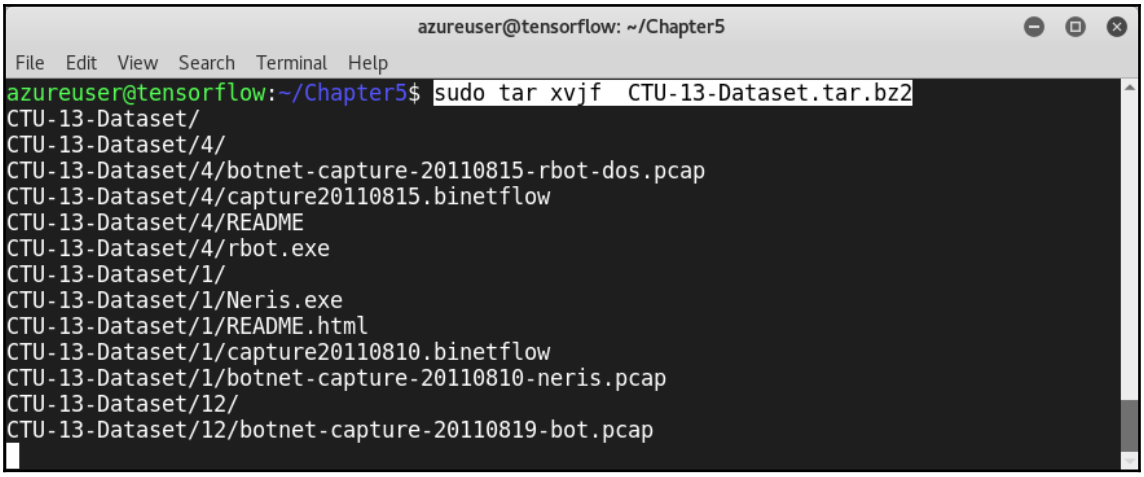
Chapter 5: Botnet Detection with Machine Learning

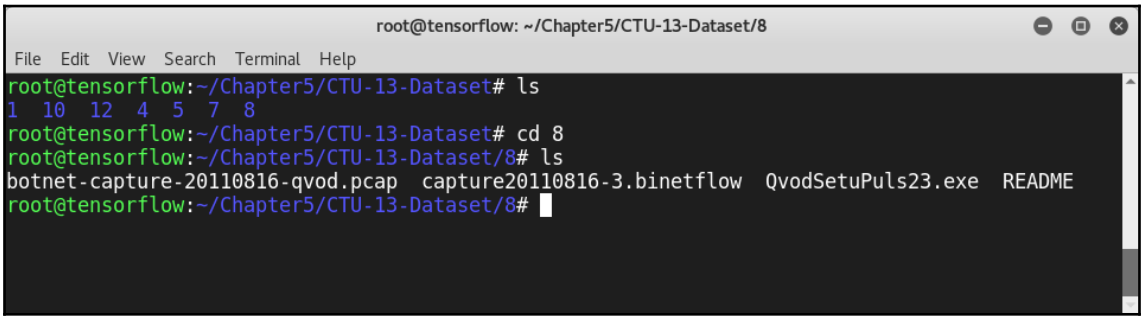



Infected Machine
1 Infection
2 Connection
Attacker
Malware Distribution
Command and
Control Server
4 Multiplication
3 Control
Botnet



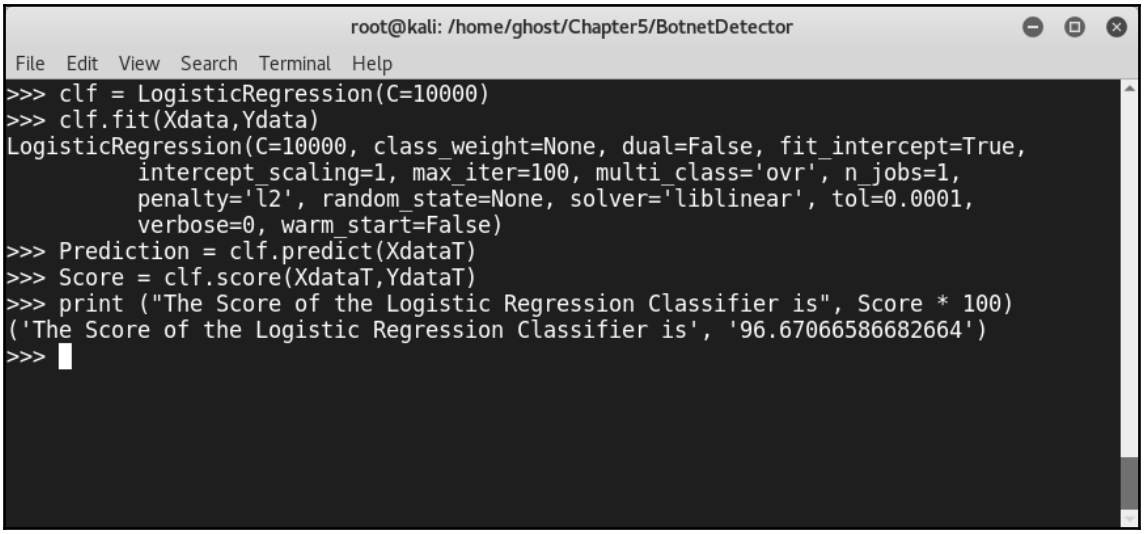

```
azureuser@tensorflow: ~/Chapter5
File  Edit  View  Search  Terminal  Help
azureuser@tensorflow:~/Chapter5$ wget --no-check-certificate https://mcfp.felk.cvut.cz
/publicDatasets/CTU-13-Dataset/CTU-13-Dataset.tar.bz2
--2018-05-18 11:37:49--  https://mcfp.felk.cvut.cz/publicDatasets/CTU-13-Dataset/CTU-1
3-Dataset.tar.bz2
Resolving mcfp.felk.cvut.cz (mcfp.felk.cvut.cz)... 147.32.83.56
Connecting to mcfp.felk.cvut.cz (mcfp.felk.cvut.cz)|147.32.83.56|:443... connected.
WARNING: cannot verify mcfp.felk.cvut.cz's certificate, issued by 'CN=TERENA SSL CA 3,
O=TERENA,L=Amsterdam,ST=Noord-Holland,C=NL':
  Unable to locally verify the issuer's authority.
HTTP request sent, awaiting response... 200 OK
Length: 1997547391 (1.9G) [application/x-bzip2]
Saving to: 'CTU-13-Dataset.tar.bz2'

CTU-13-Dataset.tar.bz 100%[=====================>]   1.86G  29.6MB/s    in 69s

2018-05-18 11:38:58 (27.7 MB/s) - 'CTU-13-Dataset.tar.bz2' saved [1997547391/199754739
1]

azureuser@tensorflow:~/Chapter5$
```

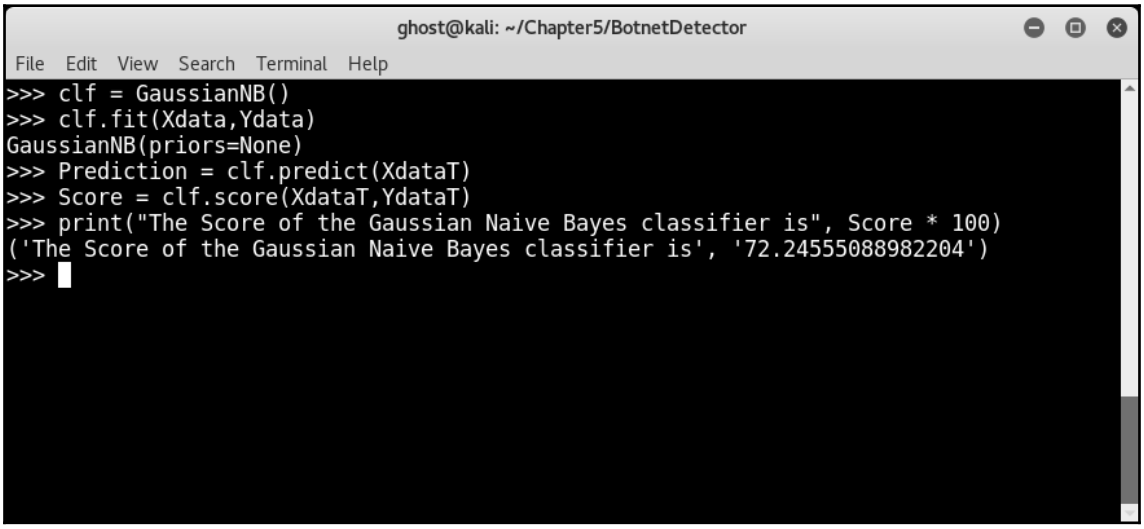

azureuser@tensorflow: ~/Chapter5
File Edit View Search Terminal Help
azureuser@tensorflow:~/Chapter5$ sudo tar xvjf CTU-13-Dataset.tar.bz2
CTU-13-Dataset/
CTU-13-Dataset/4/
CTU-13-Dataset/4/botnet-capture-20110815-rbot-dos.pcap
CTU-13-Dataset/4/capture20110815.binetflow
CTU-13-Dataset/4/README
CTU-13-Dataset/4/rbot.exe
CTU-13-Dataset/1/
CTU-13-Dataset/1/Neris.exe
CTU-13-Dataset/1/README.html
CTU-13-Dataset/1/capture20110810.binetflow
CTU-13-Dataset/1/botnet-capture-20110810-neris.pcap
CTU-13-Dataset/12/
CTU-13-Dataset/12/botnet-capture-20110819-bot.pcap


root@tensorflow: ~/Chapter5/CTU-13-Dataset/8
File Edit View Search Terminal Help
root@tensorflow:~/Chapter5/CTU-13-Dataset# ls
1 10 12 4 5 7 8
root@tensorflow:~/Chapter5/CTU-13-Dataset# cd 8
root@tensorflow:~/Chapter5/CTU-13-Dataset/8# ls
botnet-capture-20110816-qvod.pcap capture20110816-3.binetflow QvodSetuPuls23.exe README
root@tensorflow:~/Chapter5/CTU-13-Dataset/8#

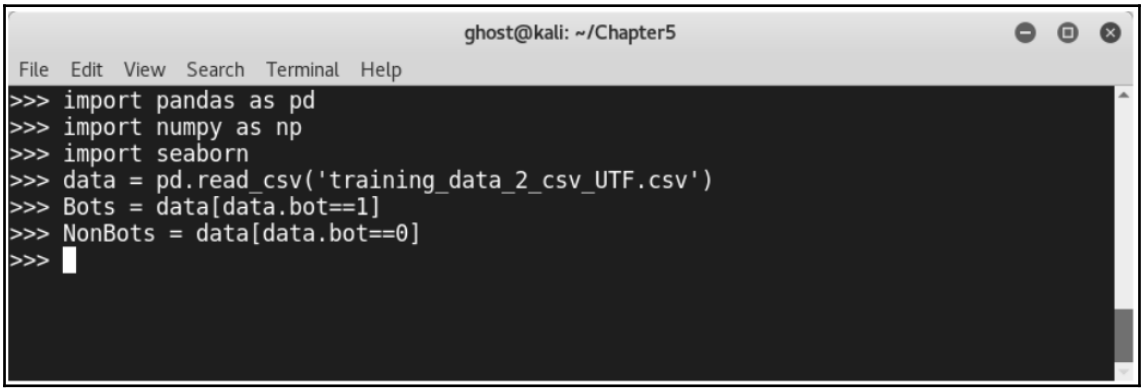
```
root@kali: /home/ghost/Chapter5/BotnetDetector
File  Edit  View  Search  Terminal  Help
>>> file = open('flowdata.pickle', 'rb')
>>> data  = pickle.load(file)
>>> Xdata = data[0]
>>> Ydata =  data[1]
>>> XdataT = data[2]
>>> YdataT = data[3]
>>> Xdata
array([[1.026539e+00, 0.000000e+00, 1.577000e+03, ..., 4.000000e+00,
        2.760000e+02, 6.110500e+04],
       [1.009595e+00, 0.000000e+00, 1.577000e+03, ..., 4.000000e+00,
        2.760000e+02, 6.110500e+04],
       [3.056586e+00, 0.000000e+00, 4.768000e+03, ..., 3.000000e+00,
        1.820000e+02, 3.400000e+01],
       ...,
       [1.031700e-01, 1.000000e+00, 2.079000e+03, ..., 2.000000e+00,
        3.740000e+02, 2.190507e+06],
       [1.004100e-01, 1.000000e+00, 2.079000e+03, ..., 2.000000e+00,
        2.320000e+02, 2.190507e+06],
       [1.031450e-01, 1.000000e+00, 2.079000e+03, ..., 2.000000e+00,
        2.700000e+02, 2.190507e+06]])
>>>
```

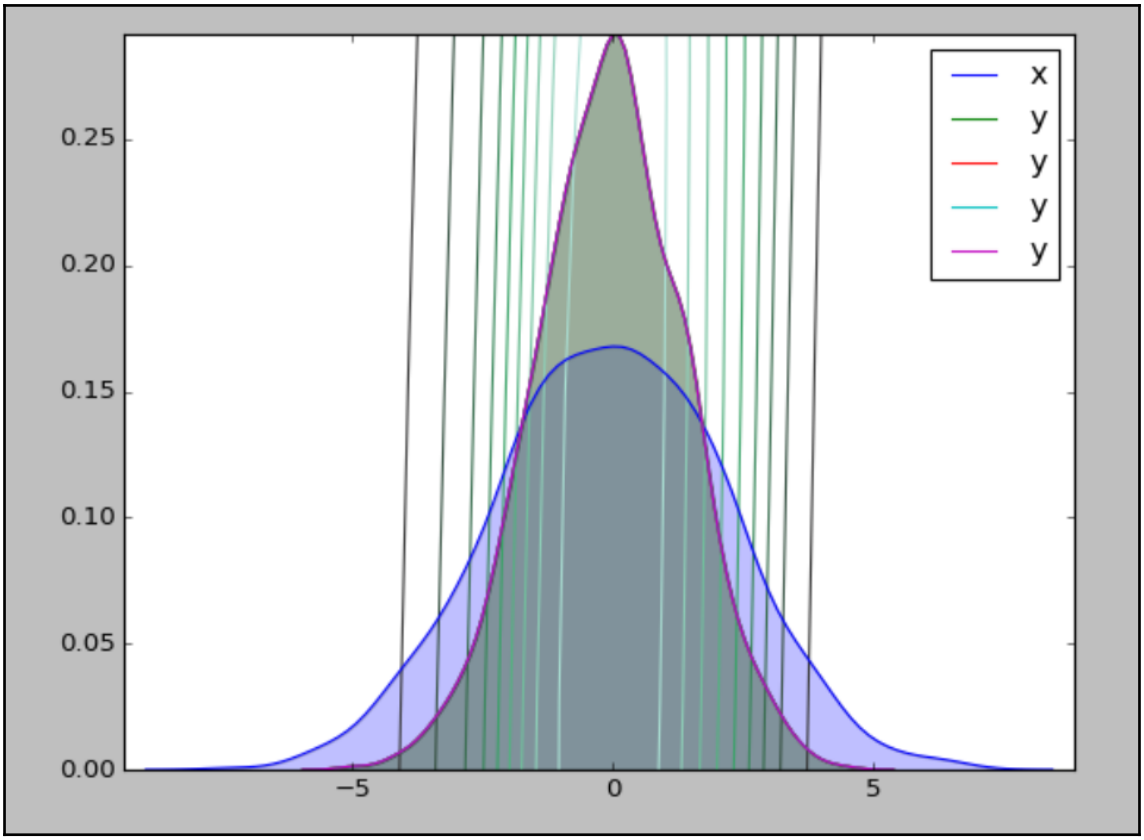
```
root@kali: /home/ghost/Chapter5/BotnetDetector
File  Edit  View  Search  Terminal  Help
>>> clf = DecisionTreeClassifier()
>>> clf.fit(Xdata,Ydata)
DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=None,
            max_features=None, max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=1, min_samples_split=2,
            min_weight_fraction_leaf=0.0, presort=False, random_state=None,
            splitter='best')
>>> Prediction = clf.predict(XdataT)
>>> Score = clf.score(XdataT,YdataT)
>>> print ("The Score of the Decision Tree Classifier is", Score * 100)
('The Score of the Decision Tree Classifier is', '99.91001799640073')
>>>
```


root@kali: /home/ghost/Chapter5/BotnetDetector
File Edit View Search Terminal Help
>>> clf = LogisticRegression(C=10000)
>>> clf.fit(Xdata,Ydata)
LogisticRegression(C=10000, class_weight=None, dual=False, fit_intercept=True,
intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,
penalty='l2', random_state=None, solver='liblinear', tol=0.0001,
verbose=0, warm_start=False)
>>> Prediction = clf.predict(XdataT)
>>> Score = clf.score(XdataT,YdataT)
>>> print ("The Score of the Logistic Regression Classifier is", Score * 100)
('The Score of the Logistic Regression Classifier is', '96.67066586682664')
>>>


ghost@kali: ~/Chapter5/BotnetDetector
File Edit View Search Terminal Help
>>> clf = GaussianNB()
>>> clf.fit(Xdata,Ydata)
GaussianNB(priors=None)
>>> Prediction = clf.predict(XdataT)
>>> Score = clf.score(XdataT,YdataT)
>>> print("The Score of the Gaussian Naive Bayes classifier is", Score * 100)
('The Score of the Gaussian Naive Bayes classifier is', '72.24555088982204')
>>>


ghost@kali: ~/Chapter5/BotnetDetector
File Edit View Search Terminal Help
>>> clf = KNeighborsClassifier()
>>> clf.fit(Xdata,Ydata)
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
metric_params=None, n_jobs=1, n_neighbors=5, p=2,
weights='uniform')
>>> Prediction = clf.predict(XdataT)
>>> Score = clf.score(XdataT,YdataT)
>>> print("The Score of the K-Nearest Neighbours classifier is", Score * 100)
('The Score of the K-Nearest Neighbours classifier is', '96.24075184963007')
>>>


ghost@kali: ~/Chapter5
File Edit View Search Terminal Help
>>> import pandas as pd
>>> import numpy as np
>>> import seaborn
>>> data = pd.read_csv('training_data_2_csv_UTF.csv')
>>> Bots = data[data.bot==1]
>>> NonBots = data[data.bot==0]
>>>


x
y
y
y
y
0.25
0.20
0.15
0.10
0.05
0.00
−5
0
5

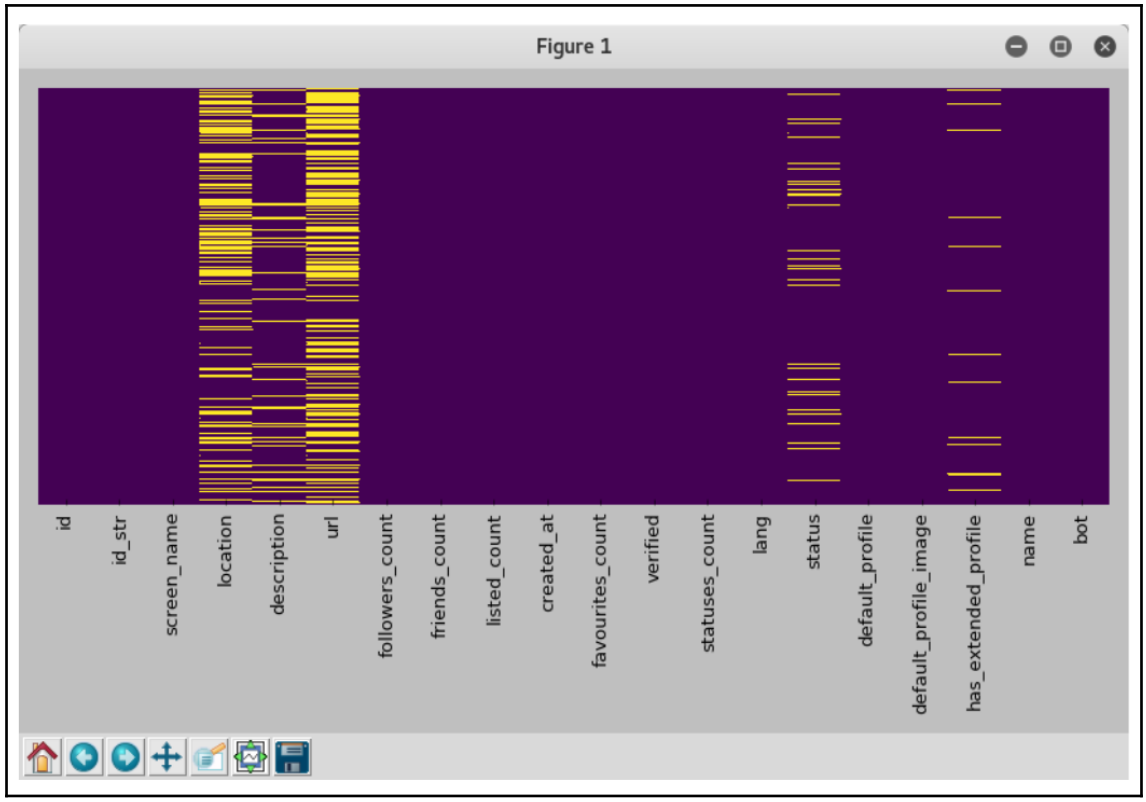
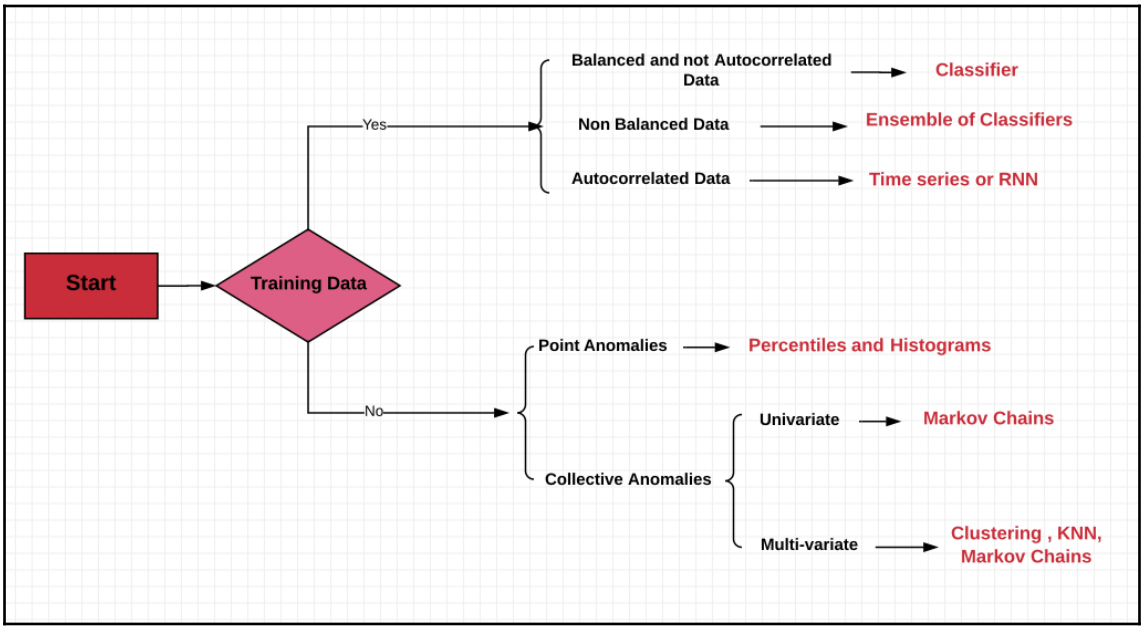
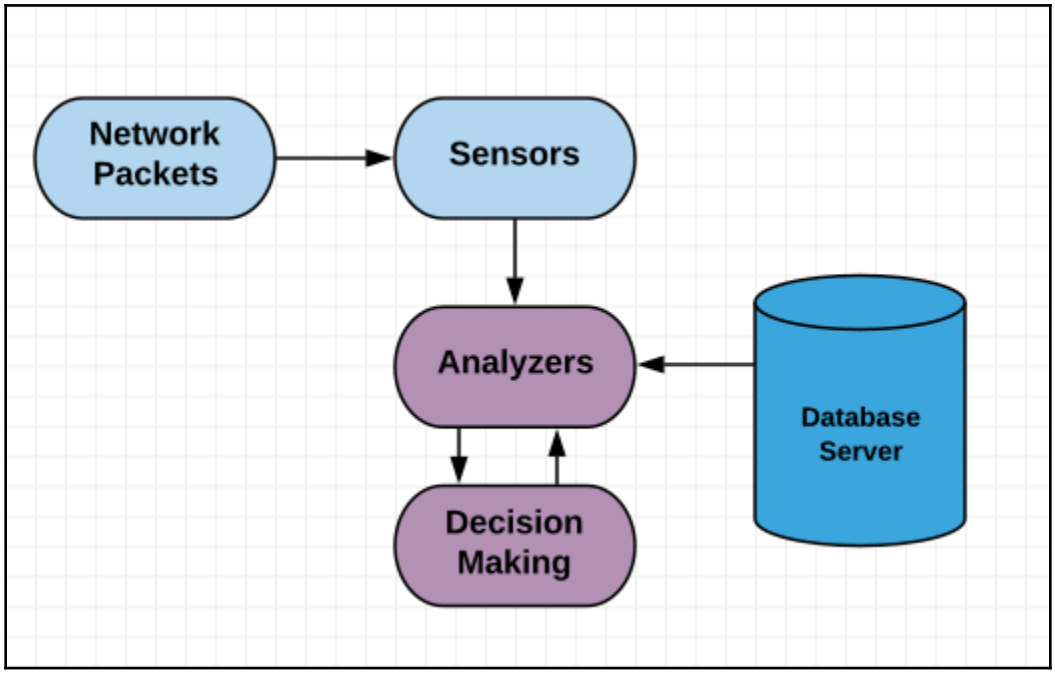
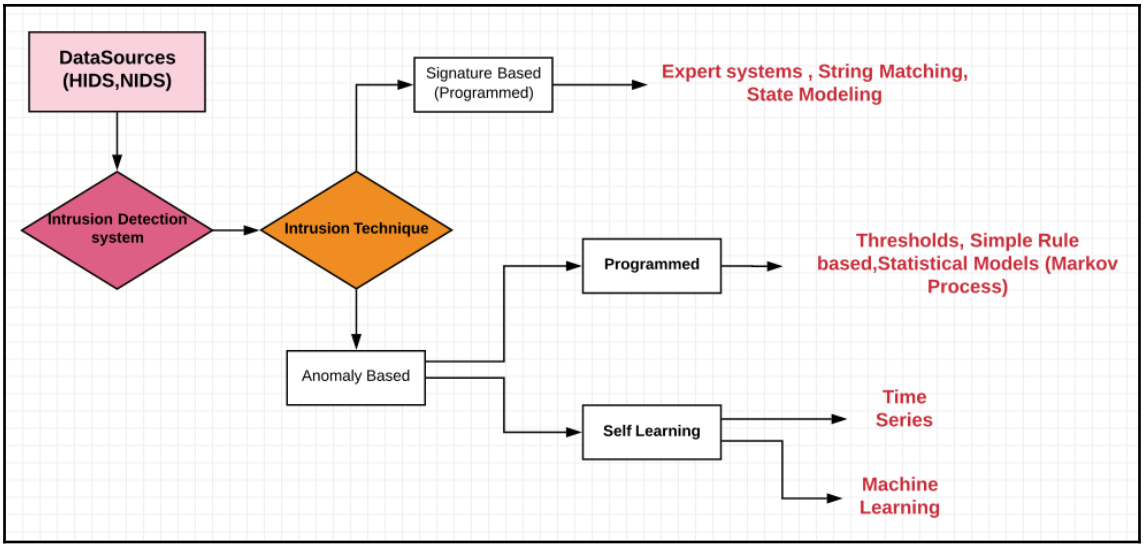


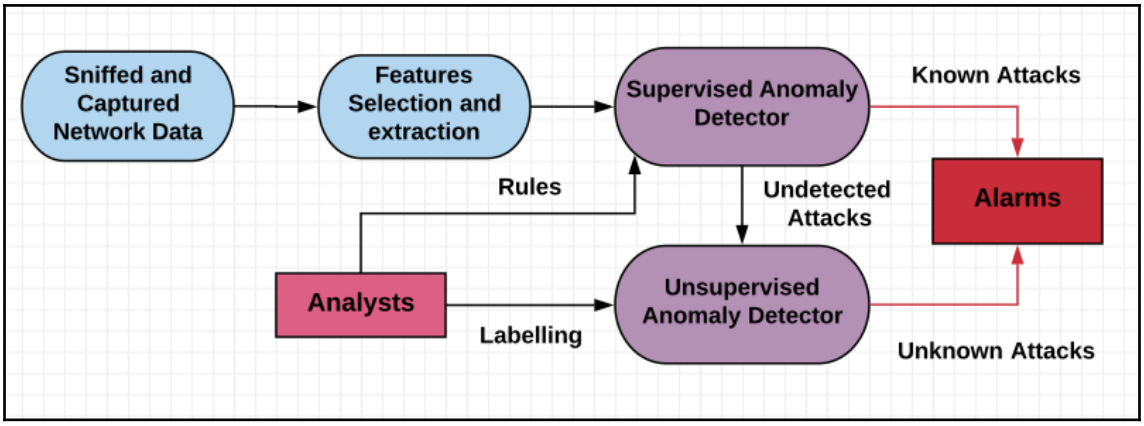
Figure 1
id
id_str
screen_name
location
description
url
followers_count
friends_count
listed_count
created_at
favourites_count
verified
statuses_count
lang
status
default_profile
default_profile_image
has_extended_profile
name
bot

Chapter 6: Machine Learning in Anomaly Detection Systems



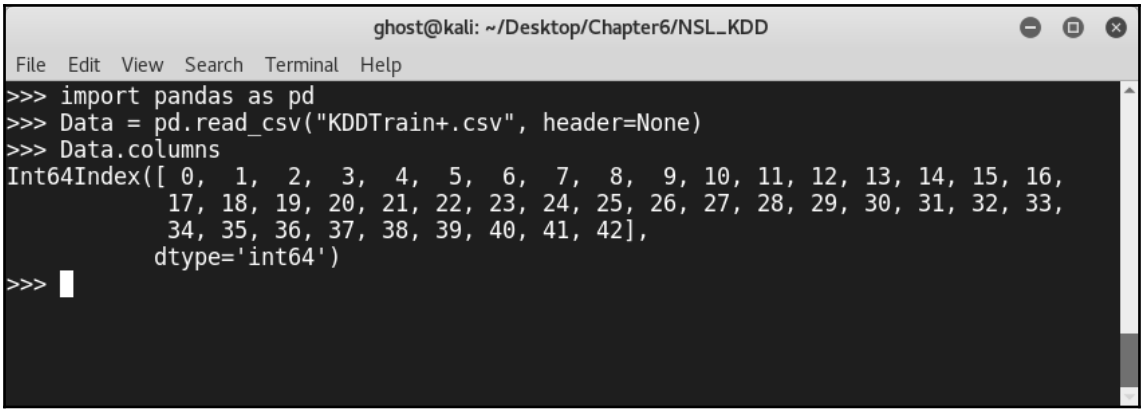

Network Packets
Sensors
Analyzers
Database Server
Decision Making

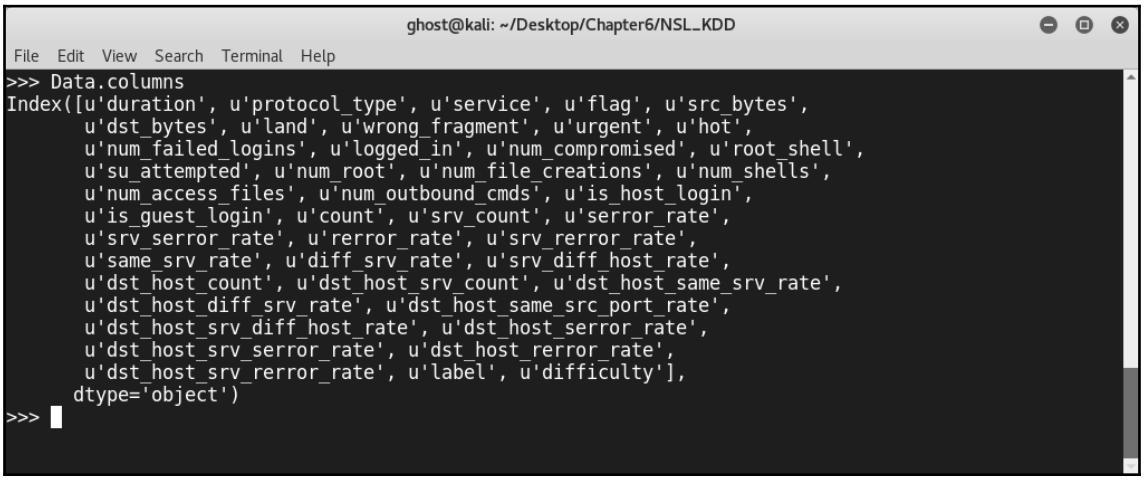

DataSources (HIDS,NIDS)
Intrusion Detection system
Intrusion Technique
Signature Based (Programmed)
Expert systems , String Matching, State Modeling
Anomaly Based
Programmed
Thresholds, Simple Rule based,Statistical Models (Markov Process)
Self Learning
Time Series
Machine Learning




```
ghost@kali: ~/Desktop/Chapter6
File  Edit  View  Search  Terminal  Help
ghost@kali:~/Desktop/Chapter6$ git clone https://github.com/defcom17/NSL_KDD
Cloning into 'NSL_KDD'...
remote: Counting objects: 15, done.
remote: Total 15 (delta 0), reused 0 (delta 0), pack-reused 15
Unpacking objects: 100% (15/15), done.
Checking connectivity... done.
ghost@kali:~/Desktop/Chapter6$ ls
NSL_KDD
ghost@kali:~/Desktop/Chapter6$ ls NSL_KDD
20 Percent Training Set.csv  KDDTest+.csv              Original NSL KDD Zip.zip
Attack Types.csv             KDDTest+.txt              ReadMe.txt
Field Names.csv              KDDTrain+_20Percent.txt   Small Training Set.csv
Field Names.docx             KDDTrain+.csv
KDDTest-21.txt               KDDTrain+.txt
ghost@kali:~/Desktop/Chapter6$
```

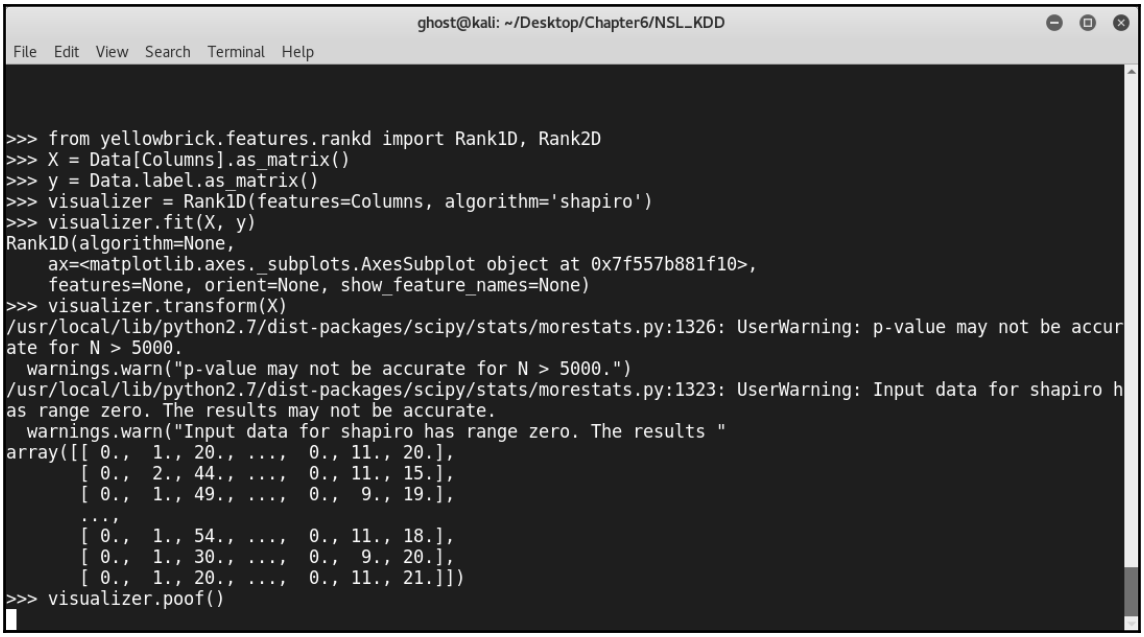
duration	continuous
protocol_type	symbolic
service	symbolic
flag	symbolic
src_bytes	continuous
dst_bytes	continuous
land	continuous
wrong_fragment	continuous
urgent	continuous
hot	continuous
num_failed_logins	continuous
logged_in	continuous
num_compromised	continuous
root_shell	continuous
su_attempted	continuous
num_root	continuous
num_file_creations	continuous
rerror_rate	continuous
srv_rerror_rate	continuous
same_srv_rate	continuous
diff_srv_rate	continuous
srv_diff_host_rate	continuous
dst_host_count	continuous
dst_host_srv_count	continuous
dst_host_same_srv_rate	continuous
dst_host_diff_srv_rate	continuous
dst_host_same_src_port_rate	continuous
dst_host_srv_diff_host_rate	continuous
dst_host_serror_rate	continuous
dst_host_srv_serror_rate	continuous
dst_host_rerror_rate	continuous
dst_host_srv_rerror_rate	continuous

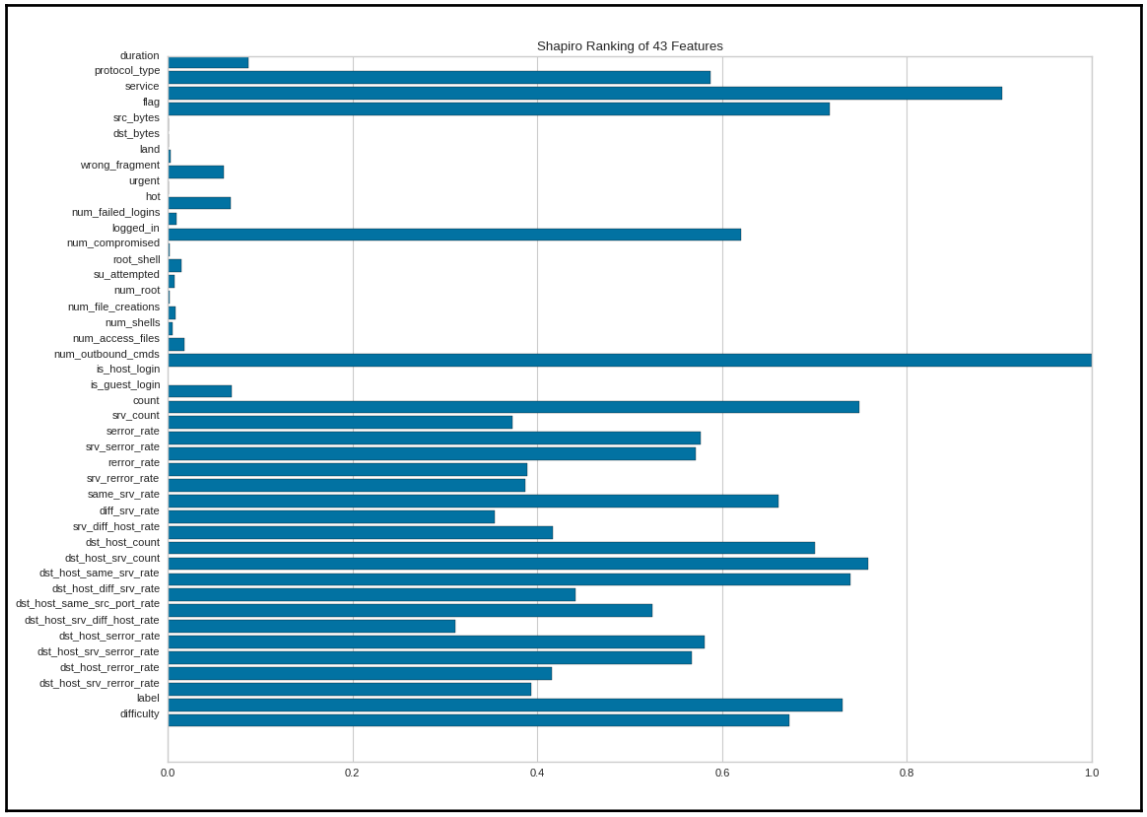

ghost@kali: ~/Desktop/Chapter6/NSL_KDD
File Edit View Search Terminal Help
>>> import pandas as pd
>>> Data = pd.read_csv("KDDTrain+.csv", header=None)
>>> Data.columns
Int64Index([0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
            17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33,
            34, 35, 36, 37, 38, 39, 40, 41, 42],
           dtype='int64')
>>>

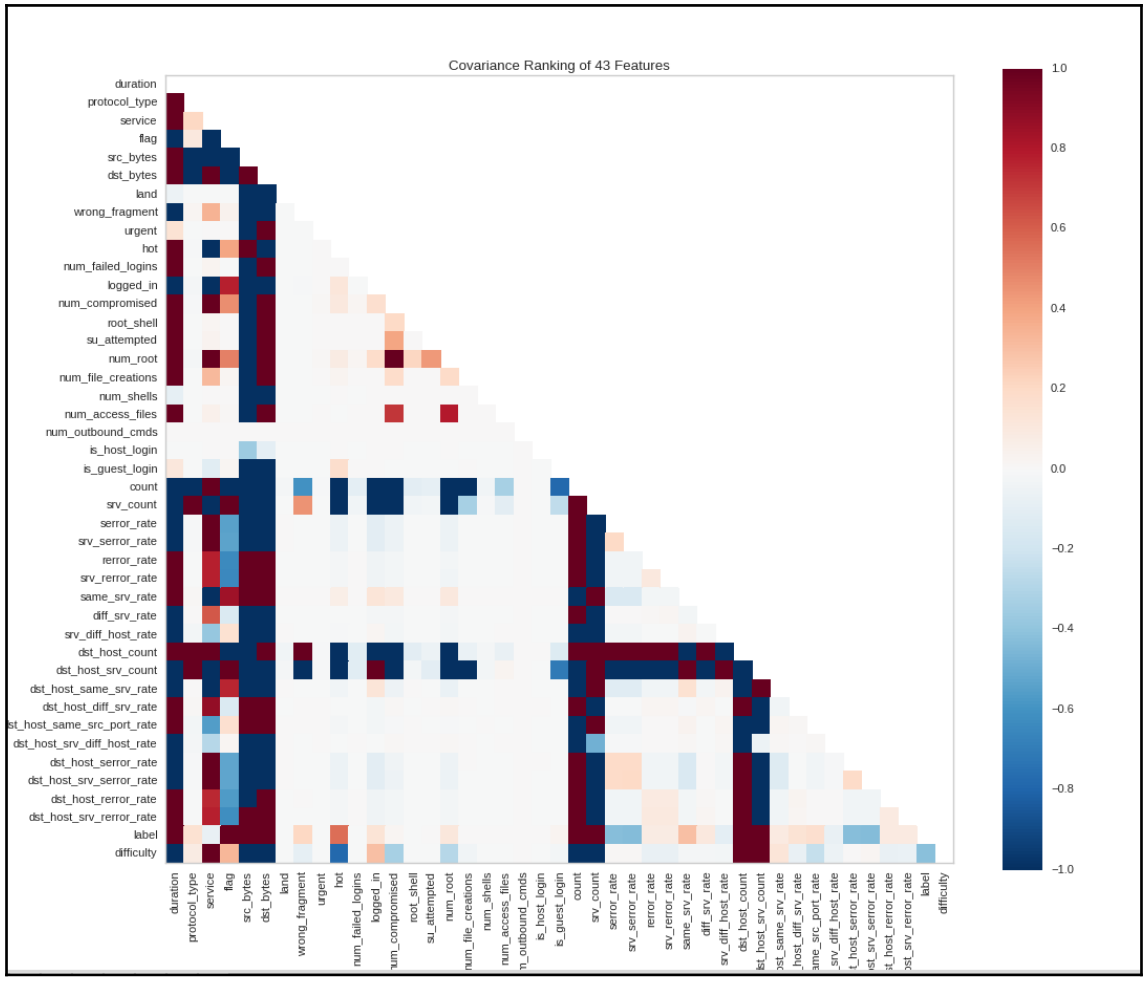

ghost@kali: ~/Desktop/Chapter6/NSL_KDD
File Edit View Search Terminal Help
>>> Data.columns
Index([u'duration', u'protocol_type', u'service', u'flag', u'src_bytes',
       u'dst_bytes', u'land', u'wrong_fragment', u'urgent', u'hot',
       u'num_failed_logins', u'logged_in', u'num_compromised', u'root_shell',
       u'su_attempted', u'num_root', u'num_file_creations', u'num_shells',
       u'num_access_files', u'num_outbound_cmds', u'is_host_login',
       u'is_guest_login', u'count', u'srv_count', u'serror_rate',
       u'srv_serror_rate', u'rerror_rate', u'srv_rerror_rate',
       u'same_srv_rate', u'diff_srv_rate', u'srv_diff_host_rate',
       u'dst_host_count', u'dst_host_srv_count', u'dst_host_same_srv_rate',
       u'dst_host_diff_srv_rate', u'dst_host_same_src_port_rate',
       u'dst_host_srv_diff_host_rate', u'dst_host_serror_rate',
       u'dst_host_srv_serror_rate', u'dst_host_rerror_rate',
       u'dst_host_srv_rerror_rate', u'label', u'difficulty'],
      dtype='object')
>>>

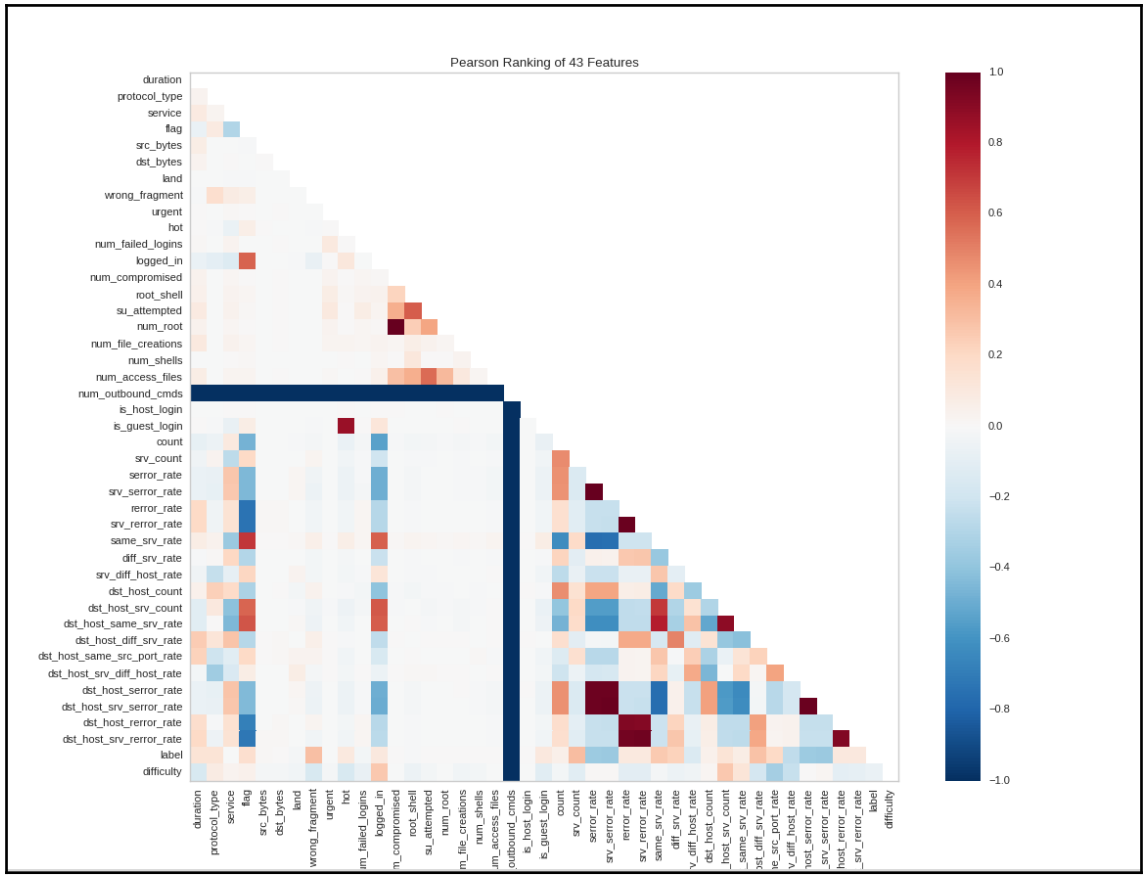
```
ghost@kali: ~/Desktop/Chapter6/NSL_KDD
File  Edit  View  Search  Terminal  Help
>>> Data.describe()
           duration     src_bytes     dst_bytes           land  \
count  125973.00000  1.259730e+05  1.259730e+05  125973.000000
mean      287.14465  4.556674e+04  1.977911e+04       0.000198
std      2604.51531  5.870331e+06  4.021269e+06       0.014086
min         0.00000  0.000000e+00  0.000000e+00       0.000000
25%         0.00000  0.000000e+00  0.000000e+00       0.000000
50%         0.00000  4.400000e+01  0.000000e+00       0.000000
75%         0.00000  2.760000e+02  5.160000e+02       0.000000
max     42908.00000  1.379964e+09  1.309937e+09       1.000000

       wrong_fragment         urgent            hot  num_failed_logins  \
count   125973.000000  125973.000000  125973.000000      125973.000000
mean         0.022687       0.000111       0.204409           0.001222
std          0.253530       0.014366       2.149968           0.045239
min          0.000000       0.000000       0.000000           0.000000
25%          0.000000       0.000000       0.000000           0.000000
50%          0.000000       0.000000       0.000000           0.000000
```

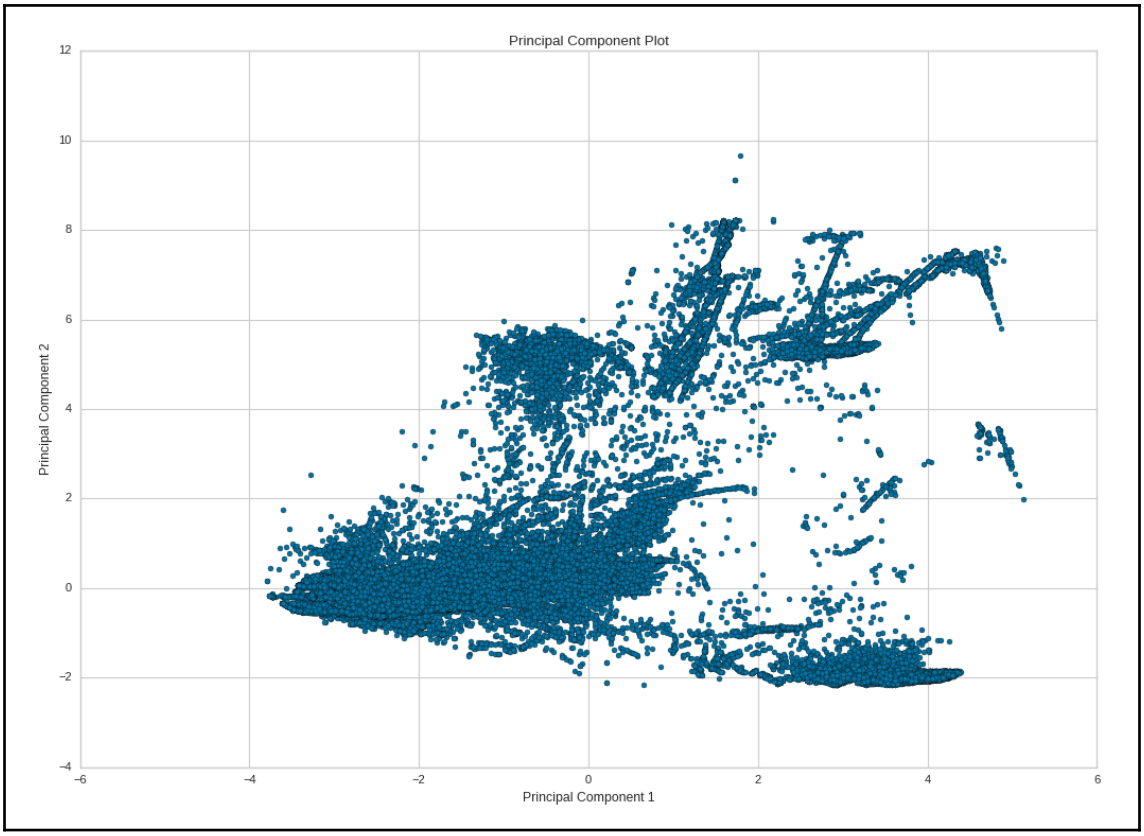
```
ghost@kali: ~/Desktop/Chapter6/NSL_KDD
File  Edit  View  Search  Terminal  Help
>>> Data.protocol_type = preprocessing.LabelEncoder().fit_transform(Data["protocol_type"])
>>> Data.service = preprocessing.LabelEncoder().fit_transform(Data["service"])
>>> Data.flag = preprocessing.LabelEncoder().fit_transform(Data["flag"])
>>> Data.label = preprocessing.LabelEncoder().fit_transform(Data["label"])
>>> Data.protocol_type
0         1
1         2
2         1
3         1
4         1
5         1
6         1
7         1
8         1
9         1
10        1
11        1
12        1
```

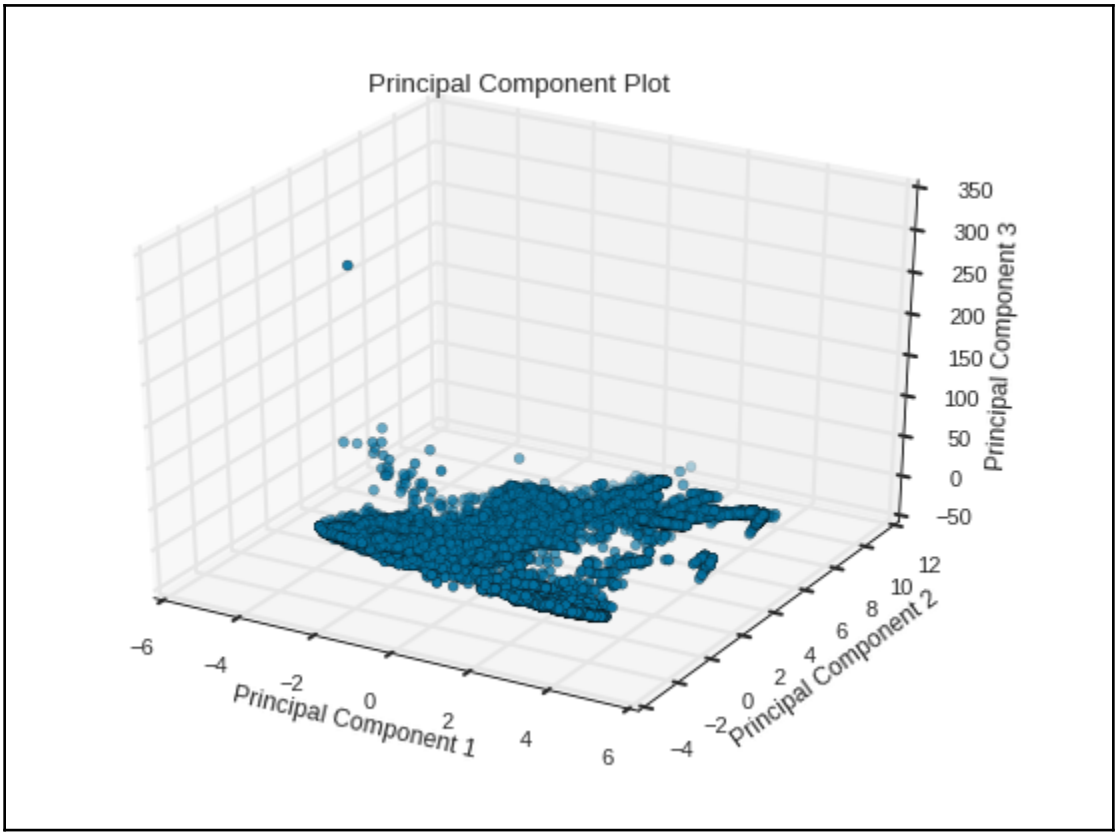

ghost@kali: ~/Desktop/Chapter6/NSL_KDD
File Edit View Search Terminal Help
>>> from yellowbrick.features.rankd import Rank1D, Rank2D
>>> X = Data[Columns].as_matrix()
>>> y = Data.label.as_matrix()
>>> visualizer = Rank1D(features=Columns, algorithm='shapiro')
>>> visualizer.fit(X, y)
Rank1D(algorithm=None,
ax=<matplotlib.axes._subplots.AxesSubplot object at 0x7f557b881f10>,
features=None, orient=None, show_feature_names=None)
>>> visualizer.transform(X)
/usr/local/lib/python2.7/dist-packages/scipy/stats/morestats.py:1326: UserWarning: p-value may not be accur
ate for N > 5000.
warnings.warn("p-value may not be accurate for N > 5000.")
/usr/local/lib/python2.7/dist-packages/scipy/stats/morestats.py:1323: UserWarning: Input data for shapiro h
as range zero. The results may not be accurate.
warnings.warn("Input data for shapiro has range zero. The results "
array([[0., 1., 20., ..., 0., 11., 20.],
[0., 2., 44., ..., 0., 11., 15.],
[0., 1., 49., ..., 0., 9., 19.],
...,
[0., 1., 54., ..., 0., 11., 18.],
[0., 1., 30., ..., 0., 9., 20.],
[0., 1., 20., ..., 0., 11., 21.]])
>>> visualizer.poof()


Shapiro Ranking of 43 Features
duration
protocol_type
service
flag
src_bytes
dst_bytes
land
wrong_fragment
urgent
hot
num_failed_logins
logged_in
num_compromised
root_shell
su_attempted
num_root
num_file_creations
num_shells
num_access_files
num_outbound_cmds
is_host_login
is_guest_login
count
srv_count
serror_rate
srv_serror_rate
rerror_rate
srv_rerror_rate
same_srv_rate
diff_srv_rate
srv_diff_host_rate
dst_host_count
dst_host_srv_count
dst_host_same_srv_rate
dst_host_diff_srv_rate
dst_host_same_src_port_rate
dst_host_srv_diff_host_rate
dst_host_serror_rate
dst_host_srv_serror_rate
dst_host_rerror_rate
dst_host_srv_rerror_rate
label
difficulty
0.0
0.2
0.4
0.6
0.8
1.0


Covariance Ranking of 43 Features
duration
protocol_type
service
flag
src_bytes
dst_bytes
land
wrong_fragment
urgent
hot
num_failed_logins
logged_in
num_compromised
root_shell
su_attempted
num_root
num_file_creations
num_shells
num_access_files
num_outbound_cmds
is_host_login
is_guest_login
count
srv_count
serror_rate
srv_serror_rate
rerror_rate
srv_rerror_rate
same_srv_rate
diff_srv_rate
srv_diff_host_rate
dst_host_count
dst_host_srv_count
dst_host_same_srv_rate
dst_host_diff_srv_rate
dst_host_same_src_port_rate
dst_host_srv_diff_host_rate
dst_host_serror_rate
dst_host_srv_serror_rate
dst_host_rerror_rate
dst_host_srv_rerror_rate
label
difficulty
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0

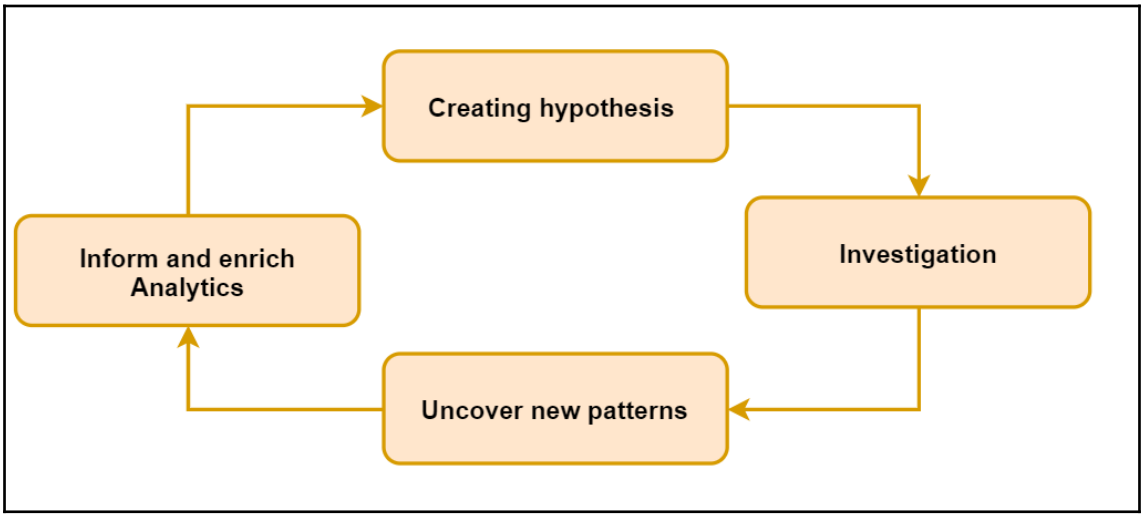

Pearson Ranking of 43 Features
duration
protocol_type
service
flag
src_bytes
dst_bytes
land
wrong_fragment
urgent
hot
num_failed_logins
logged_in
num_compromised
root_shell
su_attempted
num_root
num_file_creations
num_shells
num_access_files
num_outbound_cmds
is_host_login
is_guest_login
count
srv_count
serror_rate
srv_serror_rate
rerror_rate
srv_rerror_rate
same_srv_rate
diff_srv_rate
srv_diff_host_rate
dst_host_count
dst_host_srv_count
dst_host_same_srv_rate
dst_host_diff_srv_rate
dst_host_same_src_port_rate
dst_host_srv_diff_host_rate
dst_host_serror_rate
dst_host_srv_serror_rate
dst_host_rerror_rate
dst_host_srv_rerror_rate
label
difficulty
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0

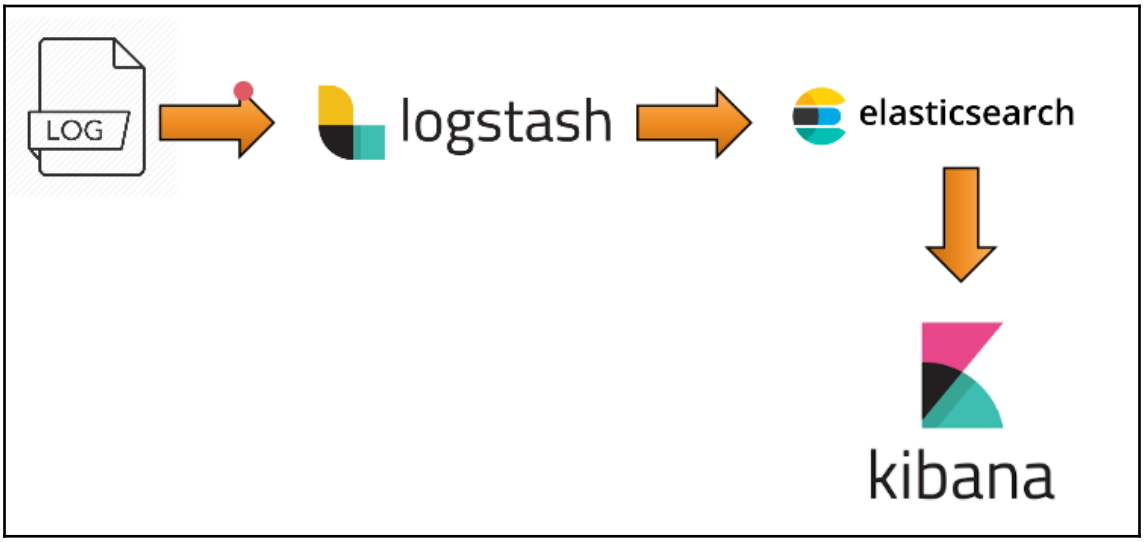

Principal Component Plot
Principal Component 2
Principal Component 1
12
10
8
6
4
2
0
−2
−4
−6
−4
−2
0
2
4
6

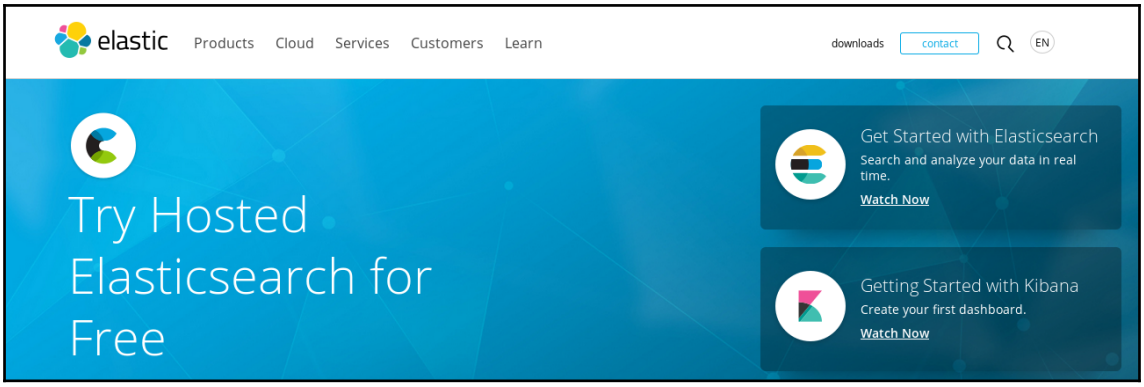


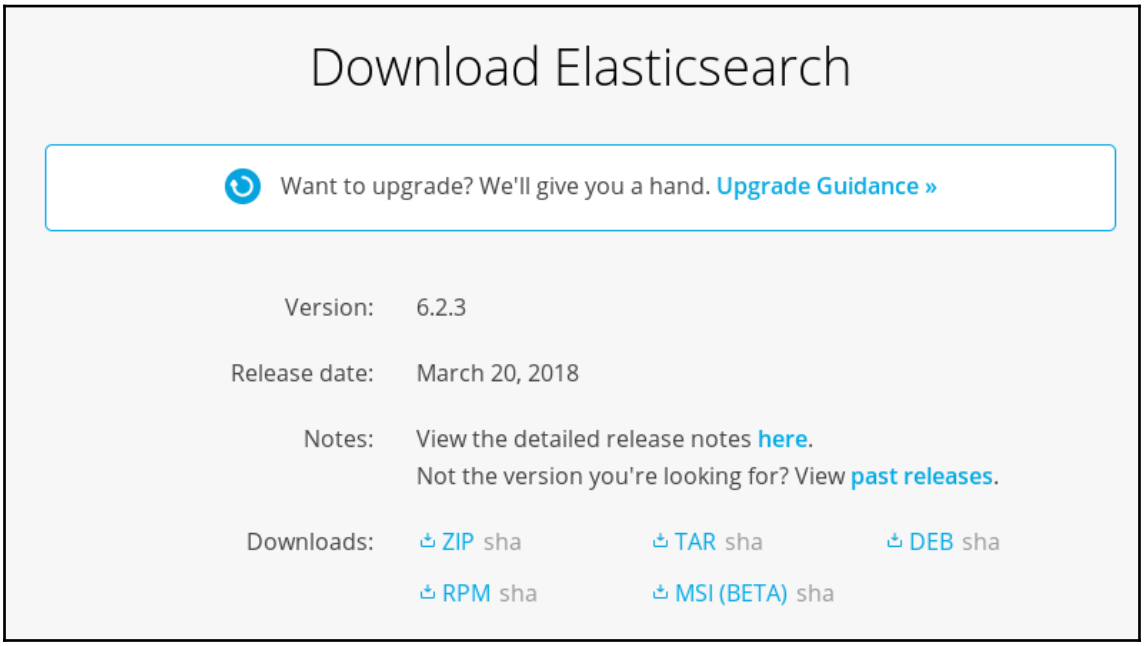

```
ghost@kali: ~/Desktop/Chapter6/NSL_KDD
File  Edit  View  Search  Terminal  Help
>>> X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
>>> clf.fit(X, y)
RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',
            max_depth=2, max_features='auto', max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=1, min_samples_split=2,
            min_weight_fraction_leaf=0.0, n_estimators=10, n_jobs=1,
            oob_score=False, random_state=0, verbose=0, warm_start=False)
>>> Score = clf.score(X_test,y_test)
>>> print (Score*100)
85.75511014090097
>>>
```

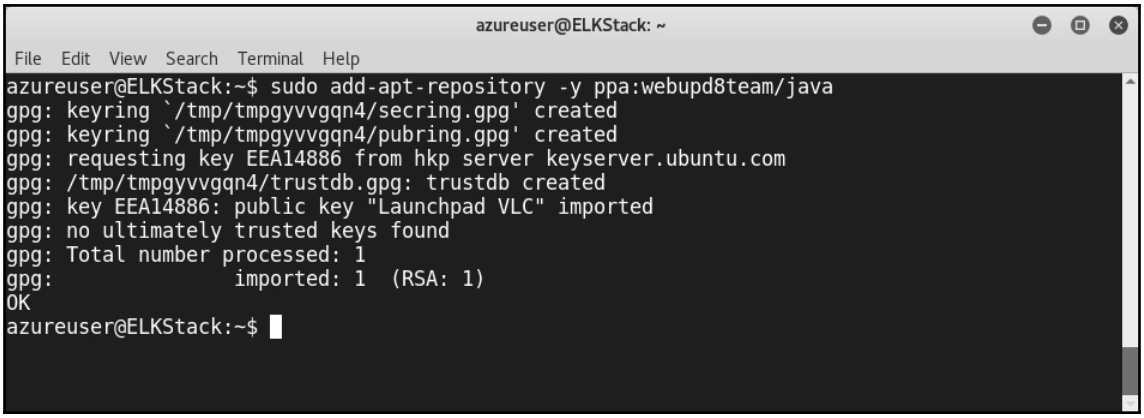
Chapter 7: Detecting Advanced Persistent Threats

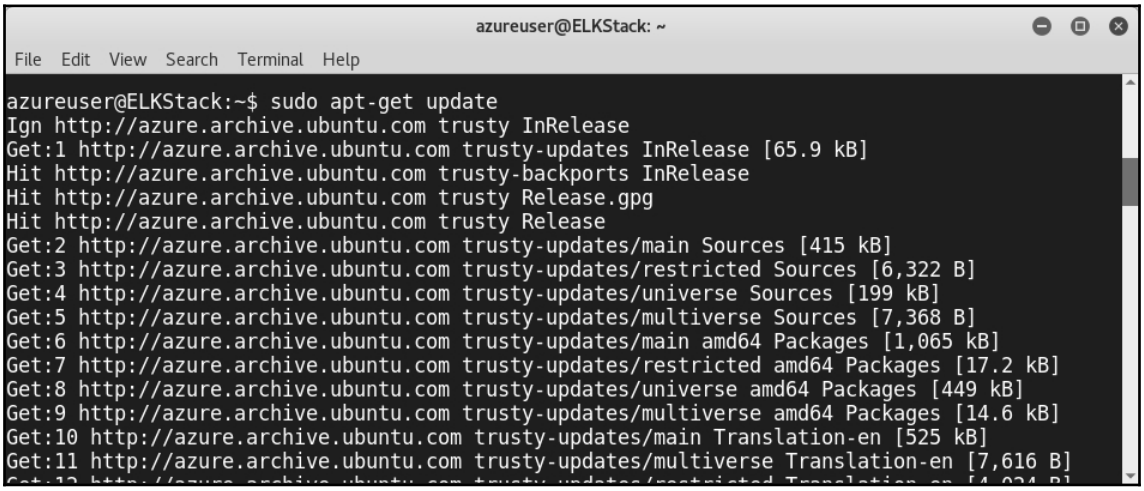


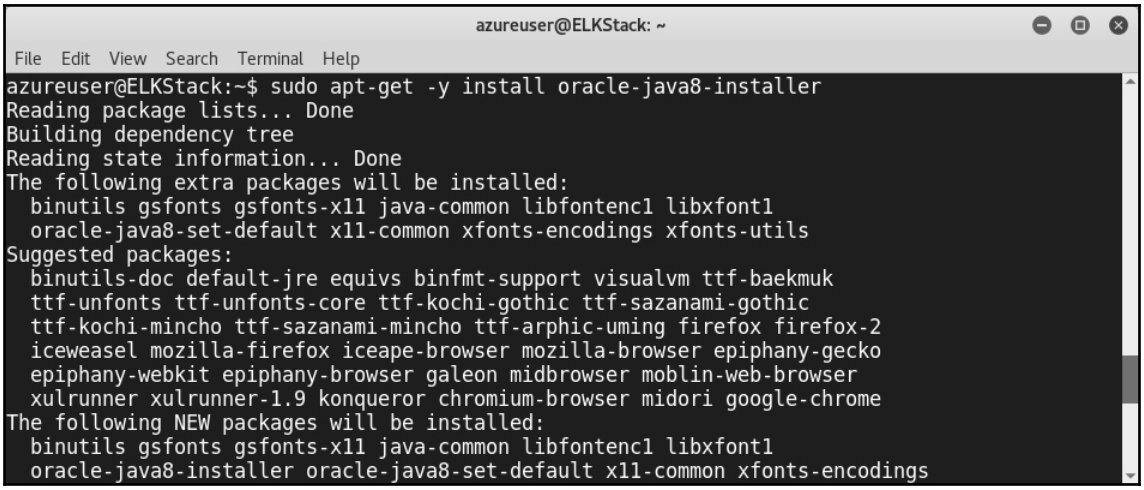



elastic
Products
Cloud
Services
Customers
Learn
downloads
contact
EN
Try Hosted Elasticsearch for Free
Get Started with Elasticsearch
Search and analyze your data in real time.
Watch Now
Getting Started with Kibana
Create your first dashboard.
Watch Now

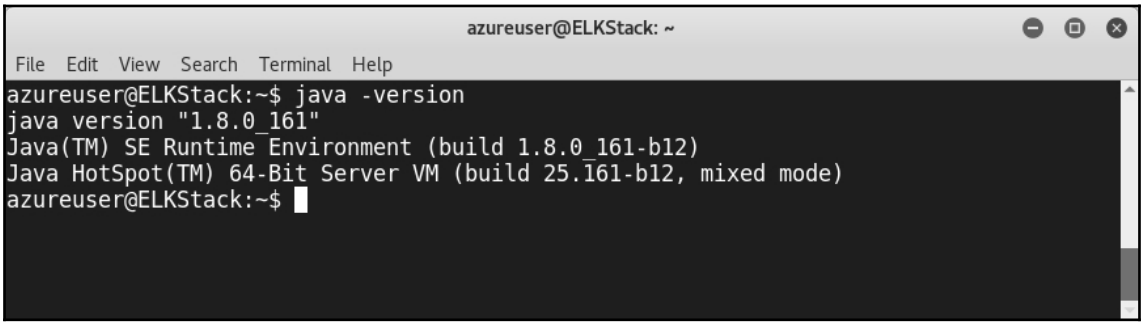

Download Elasticsearch
Want to upgrade? We'll give you a hand. Upgrade Guidance »
Version: 6.2.3
Release date: March 20, 2018
Notes: View the detailed release notes here.
Not the version you're looking for? View past releases.
Downloads: ZIP sha TAR sha DEB sha
RPM sha MSI (BETA) sha

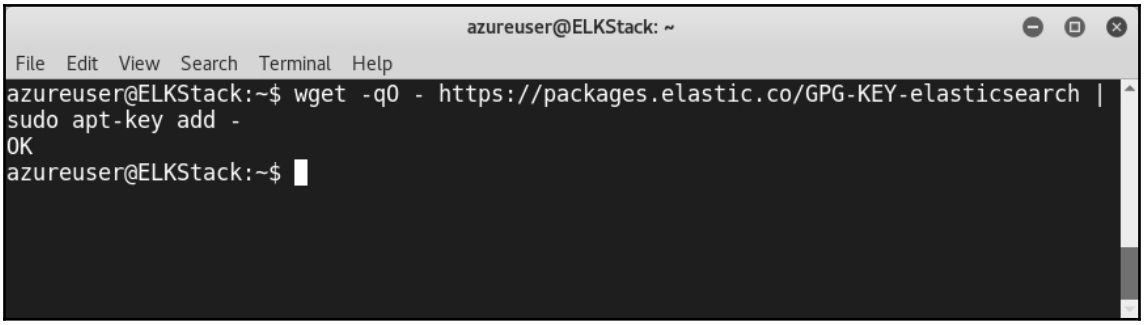

azureuser@ELKStack: ~
File Edit View Search Terminal Help
azureuser@ELKStack:~$ sudo add-apt-repository -y ppa:webupd8team/java
gpg: keyring `/tmp/tmpgyvvgqn4/secring.gpg' created
gpg: keyring `/tmp/tmpgyvvgqn4/pubring.gpg' created
gpg: requesting key EEA14886 from hkp server keyserver.ubuntu.com
gpg: /tmp/tmpgyvvgqn4/trustdb.gpg: trustdb created
gpg: key EEA14886: public key "Launchpad VLC" imported
gpg: no ultimately trusted keys found
gpg: Total number processed: 1
gpg: imported: 1 (RSA: 1)
OK
azureuser@ELKStack:~$


azureuser@ELKStack: ~
File Edit View Search Terminal Help
azureuser@ELKStack:~$ sudo apt-get update
Ign http://azure.archive.ubuntu.com trusty InRelease
Get:1 http://azure.archive.ubuntu.com trusty-updates InRelease [65.9 kB]
Hit http://azure.archive.ubuntu.com trusty-backports InRelease
Hit http://azure.archive.ubuntu.com trusty Release.gpg
Hit http://azure.archive.ubuntu.com trusty Release
Get:2 http://azure.archive.ubuntu.com trusty-updates/main Sources [415 kB]
Get:3 http://azure.archive.ubuntu.com trusty-updates/restricted Sources [6,322 B]
Get:4 http://azure.archive.ubuntu.com trusty-updates/universe Sources [199 kB]
Get:5 http://azure.archive.ubuntu.com trusty-updates/multiverse Sources [7,368 B]
Get:6 http://azure.archive.ubuntu.com trusty-updates/main amd64 Packages [1,065 kB]
Get:7 http://azure.archive.ubuntu.com trusty-updates/restricted amd64 Packages [17.2 kB]
Get:8 http://azure.archive.ubuntu.com trusty-updates/universe amd64 Packages [449 kB]
Get:9 http://azure.archive.ubuntu.com trusty-updates/multiverse amd64 Packages [14.6 kB]
Get:10 http://azure.archive.ubuntu.com trusty-updates/main Translation-en [525 kB]
Get:11 http://azure.archive.ubuntu.com trusty-updates/multiverse Translation-en [7,616 B]

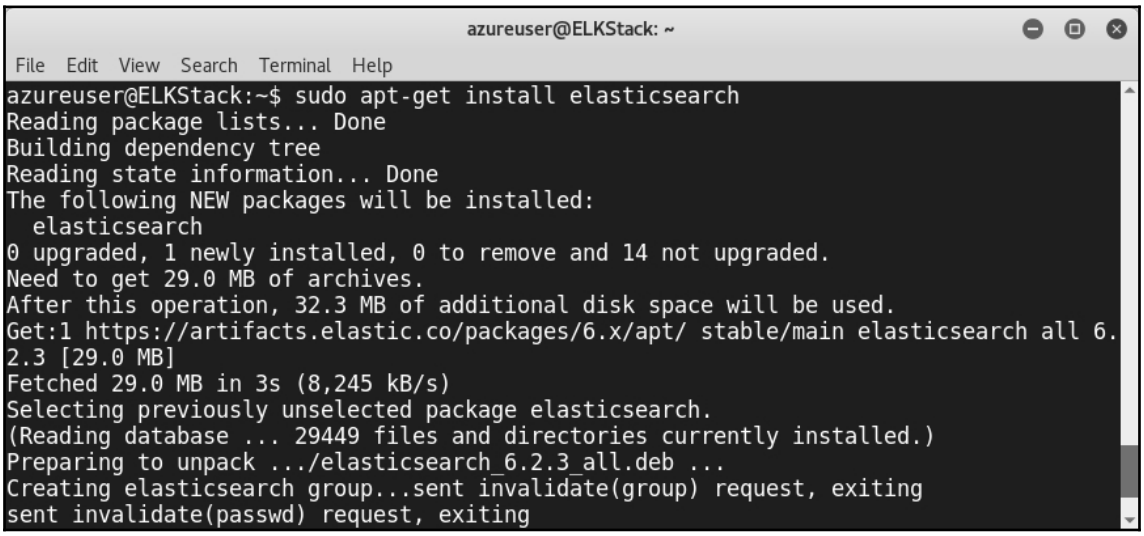

azureuser@ELKStack: ~
File Edit View Search Terminal Help
azureuser@ELKStack:~$ sudo apt-get -y install oracle-java8-installer
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following extra packages will be installed:
binutils gsfonts gsfonts-x11 java-common libfontenc1 libxfont1
oracle-java8-set-default x11-common xfonts-encodings xfonts-utils
Suggested packages:
binutils-doc default-jre equivs binfmt-support visualvm ttf-baekmuk
ttf-unfonts ttf-unfonts-core ttf-kochi-gothic ttf-sazanami-gothic
ttf-kochi-mincho ttf-sazanami-mincho ttf-arphic-uming firefox firefox-2
iceweasel mozilla-firefox iceape-browser mozilla-browser epiphany-gecko
epiphany-webkit epiphany-browser galeon midbrowser moblin-web-browser
xulrunner xulrunner-1.9 konqueror chromium-browser midori google-chrome
The following NEW packages will be installed:
binutils gsfonts gsfonts-x11 java-common libfontenc1 libxfont1
oracle-java8-installer oracle-java8-set-default x11-common xfonts-encodings

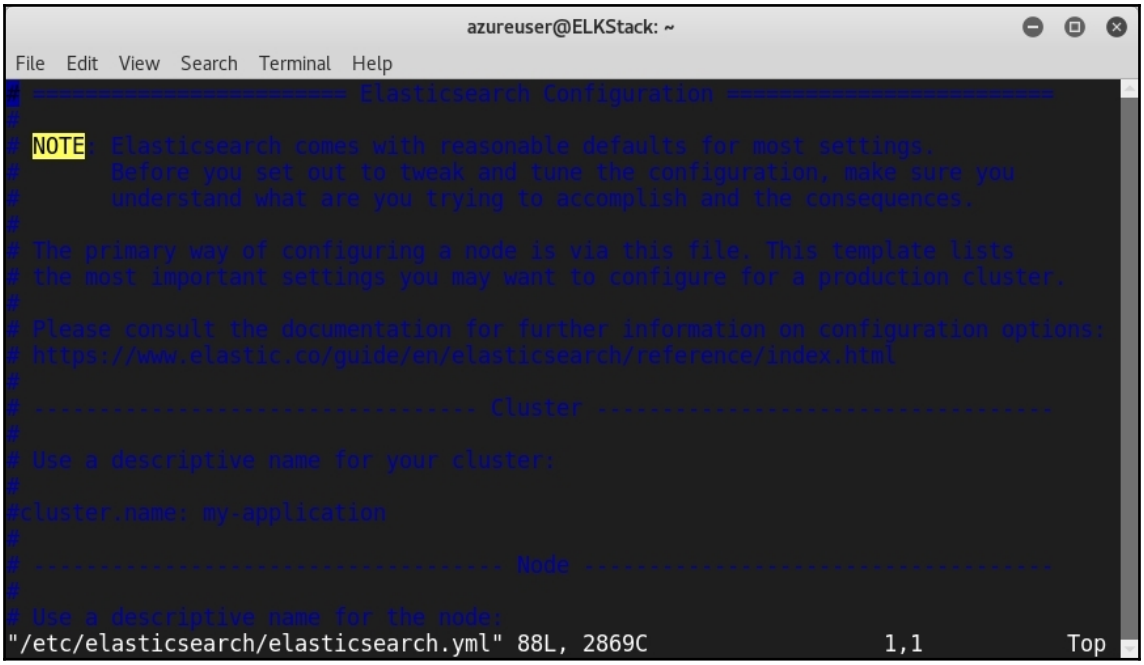

azureuser@ELKStack: ~
File Edit View Search Terminal Help
Package configuration
Configuring oracle-java8-installer
Oracle Binary Code License Agreement for the Java SE Platform Products and JavaFX
You MUST agree to the license available in http://java.com/license if you want to use Oracle JDK.
<Ok>

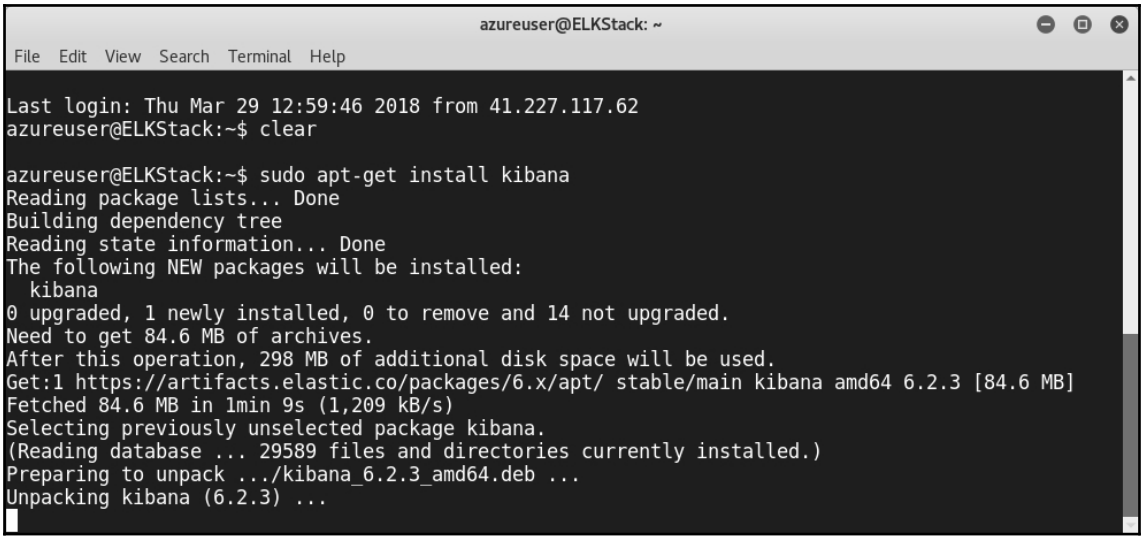

azureuser@ELKStack: ~
File Edit View Search Terminal Help
azureuser@ELKStack:~$ java -version
java version "1.8.0_161"
Java(TM) SE Runtime Environment (build 1.8.0_161-b12)
Java HotSpot(TM) 64-Bit Server VM (build 25.161-b12, mixed mode)
azureuser@ELKStack:~$

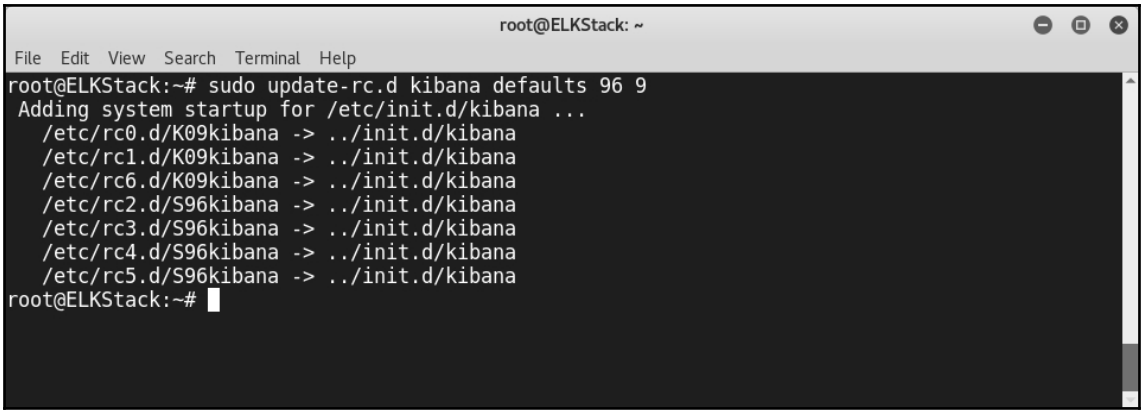

azureuser@ELKStack: ~
File Edit View Search Terminal Help
azureuser@ELKStack:~$ wget -qO - https://packages.elastic.co/GPG-KEY-elasticsearch | sudo apt-key add -
OK
azureuser@ELKStack:~$

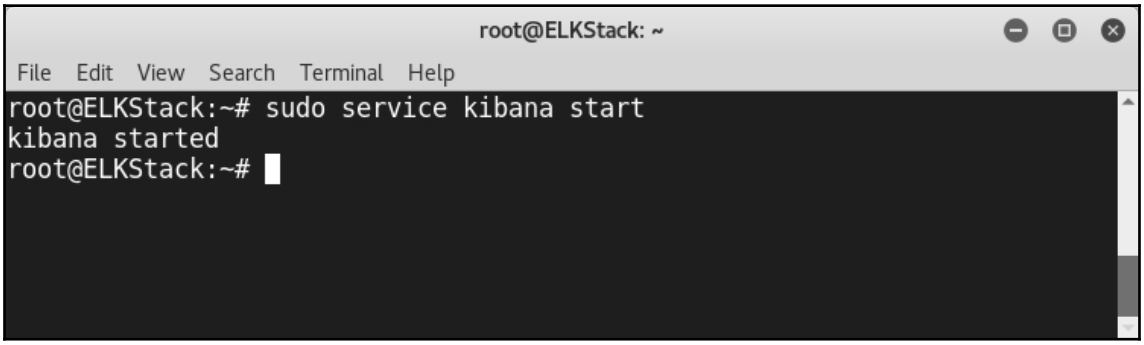

azureuser@ELKStack: ~
File Edit View Search Terminal Help
azureuser@ELKStack:~$ echo "deb https://artifacts.elastic.co/packages/6.x/apt stable main" | sudo tee -a /etc/apt/sources.list.d/elastic-6.x.list
deb https://artifacts.elastic.co/packages/6.x/apt stable main
azureuser@ELKStack:~$

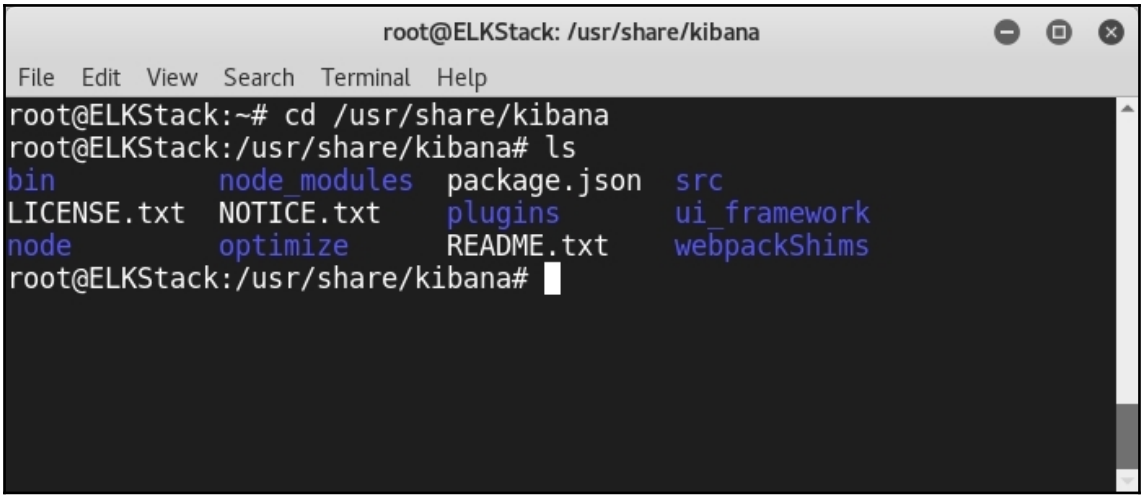

azureuser@ELKStack: ~
File Edit View Search Terminal Help
azureuser@ELKStack:~$ sudo apt-get install elasticsearch
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
elasticsearch
0 upgraded, 1 newly installed, 0 to remove and 14 not upgraded.
Need to get 29.0 MB of archives.
After this operation, 32.3 MB of additional disk space will be used.
Get:1 https://artifacts.elastic.co/packages/6.x/apt/ stable/main elasticsearch all 6.2.3 [29.0 MB]
Fetched 29.0 MB in 3s (8,245 kB/s)
Selecting previously unselected package elasticsearch.
(Reading database ... 29449 files and directories currently installed.)
Preparing to unpack .../elasticsearch_6.2.3_all.deb ...
Creating elasticsearch group...sent invalidate(group) request, exiting
sent invalidate(passwd) request, exiting

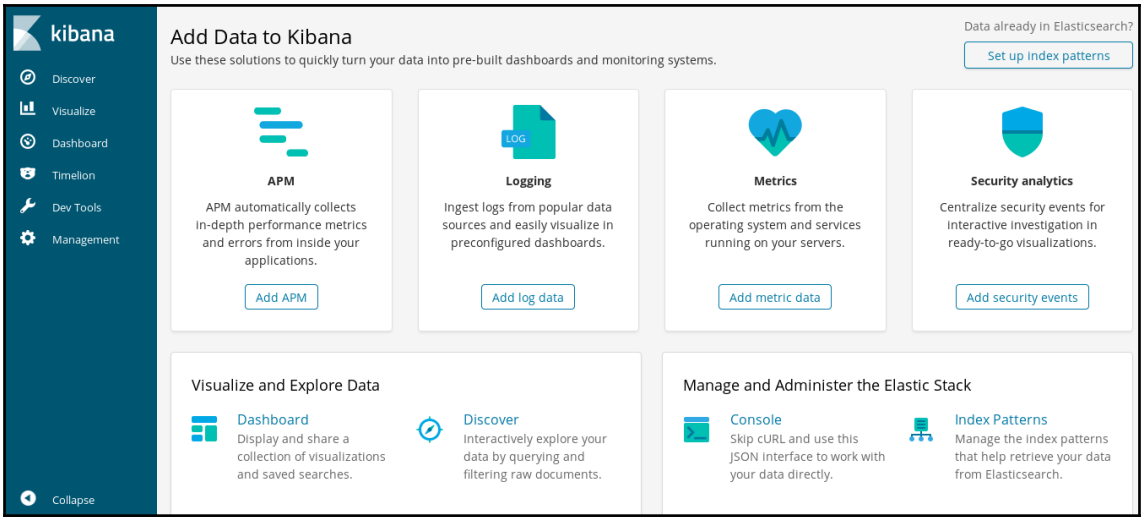

azureuser@ELKStack: ~
File Edit View Search Terminal Help
========================== Elasticsearch Configuration =========================
#
NOTE: Elasticsearch comes with reasonable defaults for most settings.
Before you set out to tweak and tune the configuration, make sure you
understand what are you trying to accomplish and the consequences.
#
The primary way of configuring a node is via this file. This template lists
the most important settings you may want to configure for a production cluster.
#
Please consult the documentation for further information on configuration options:
https://www.elastic.co/guide/en/elasticsearch/reference/index.html
#
---------------------------------- Cluster -----------------------------------
#
Use a descriptive name for your cluster:
#
#cluster.name: my-application
#
------------------------------------ Node ------------------------------------
#
Use a descriptive name for the node:
"/etc/elasticsearch/elasticsearch.yml" 88L, 2869C 1,1 Top

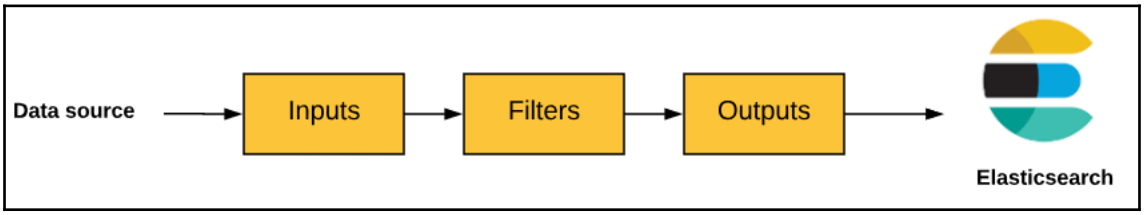

azureuser@ELKStack: ~
File Edit View Search Terminal Help
Last login: Thu Mar 29 12:59:46 2018 from 41.227.117.62
azureuser@ELKStack:~$ clear
azureuser@ELKStack:~$ sudo apt-get install kibana
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
kibana
0 upgraded, 1 newly installed, 0 to remove and 14 not upgraded.
Need to get 84.6 MB of archives.
After this operation, 298 MB of additional disk space will be used.
Get:1 https://artifacts.elastic.co/packages/6.x/apt/ stable/main kibana amd64 6.2.3 [84.6 MB]
Fetched 84.6 MB in 1min 9s (1,209 kB/s)
Selecting previously unselected package kibana.
(Reading database ... 29589 files and directories currently installed.)
Preparing to unpack .../kibana_6.2.3_amd64.deb ...
Unpacking kibana (6.2.3) ...

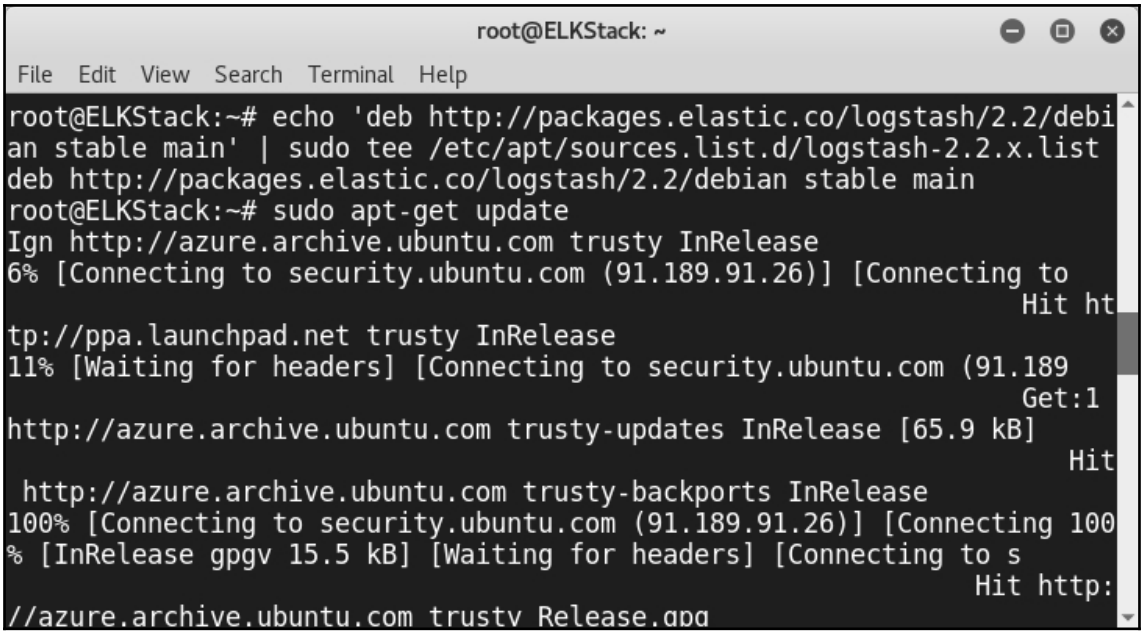

root@ELKStack: ~
File Edit View Search Terminal Help
root@ELKStack:~# sudo update-rc.d kibana defaults 96 9
Adding system startup for /etc/init.d/kibana ...
/etc/rc0.d/K09kibana -> ../init.d/kibana
/etc/rc1.d/K09kibana -> ../init.d/kibana
/etc/rc6.d/K09kibana -> ../init.d/kibana
/etc/rc2.d/S96kibana -> ../init.d/kibana
/etc/rc3.d/S96kibana -> ../init.d/kibana
/etc/rc4.d/S96kibana -> ../init.d/kibana
/etc/rc5.d/S96kibana -> ../init.d/kibana
root@ELKStack:~#

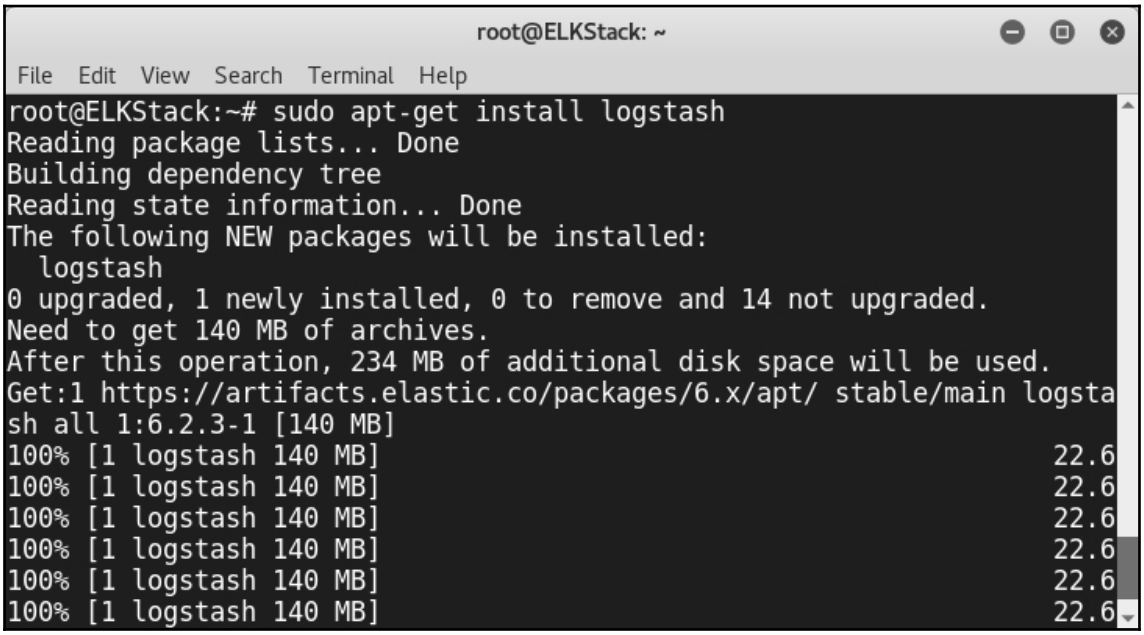

root@ELKStack: ~
File Edit View Search Terminal Help
root@ELKStack:~# sudo service kibana start
kibana started
root@ELKStack:~#

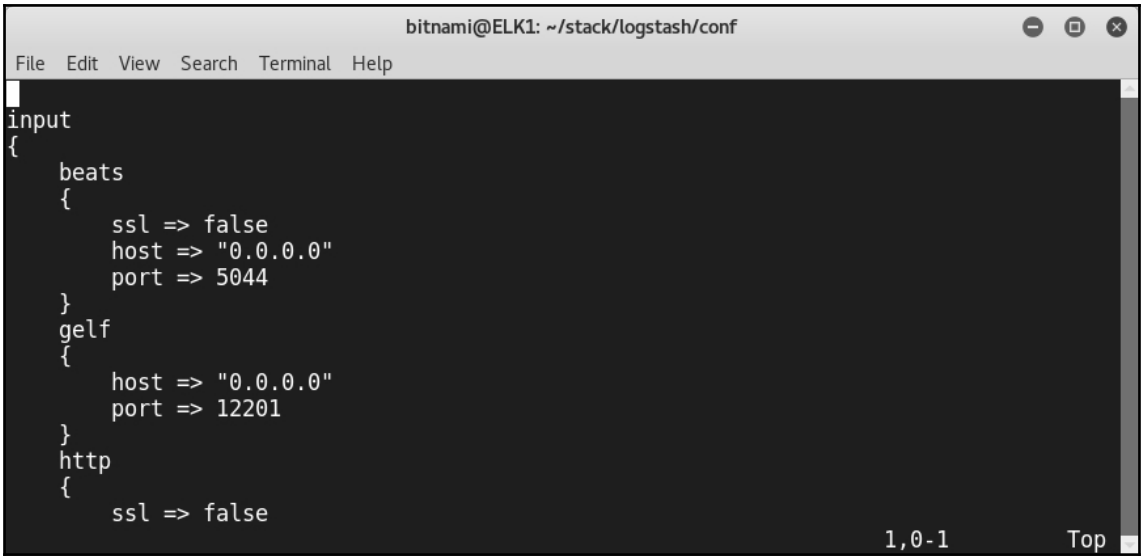

root@ELKStack: /usr/share/kibana
File Edit View Search Terminal Help
root@ELKStack:~# cd /usr/share/kibana
root@ELKStack:/usr/share/kibana# ls
bin node_modules package.json src
LICENSE.txt NOTICE.txt plugins ui_framework
node optimize README.txt webpackShims
root@ELKStack:/usr/share/kibana#

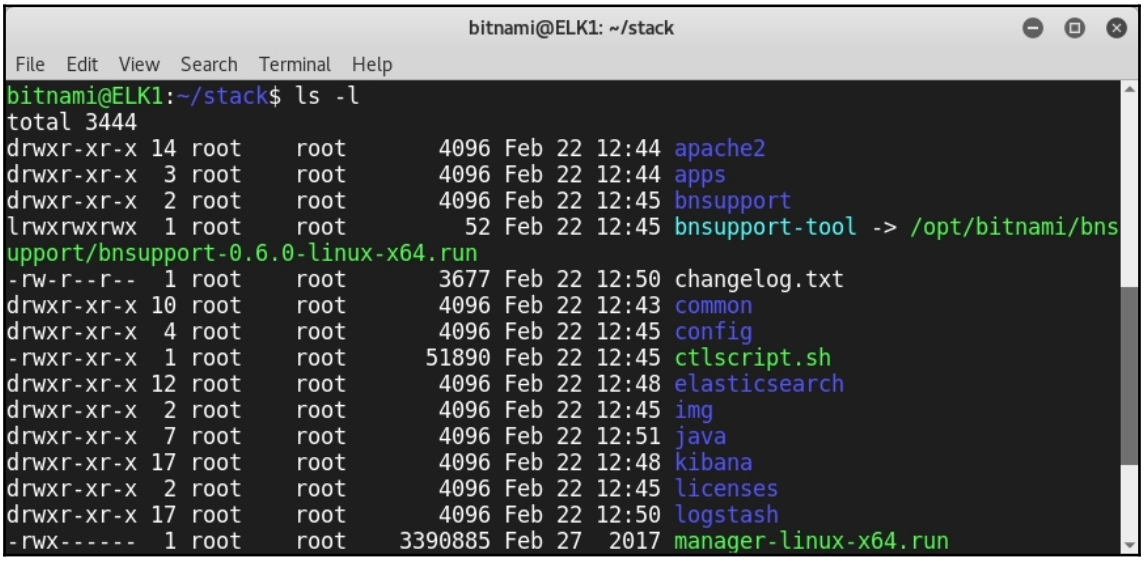

kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Collapse
Add Data to Kibana
Use these solutions to quickly turn your data into pre-built dashboards and monitoring systems.
Data already in Elasticsearch?
Set up index patterns
APM
APM automatically collects in-depth performance metrics and errors from inside your applications.
Add APM
LOG
Logging
Ingest logs from popular data sources and easily visualize in preconfigured dashboards.
Add log data
Metrics
Collect metrics from the operating system and services running on your servers.
Add metric data
Security analytics
Centralize security events for interactive investigation in ready-to-go visualizations.
Add security events
Visualize and Explore Data
Dashboard
Display and share a collection of visualizations and saved searches.
Discover
Interactively explore your data by querying and filtering raw documents.
Manage and Administer the Elastic Stack
Console
Skip cURL and use this JSON interface to work with your data directly.
Index Patterns
Manage the index patterns that help retrieve your data from Elasticsearch.

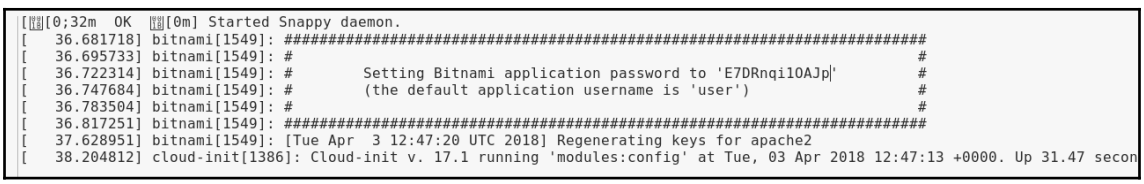

Data source
Inputs
Filters
Outputs
Elasticsearch

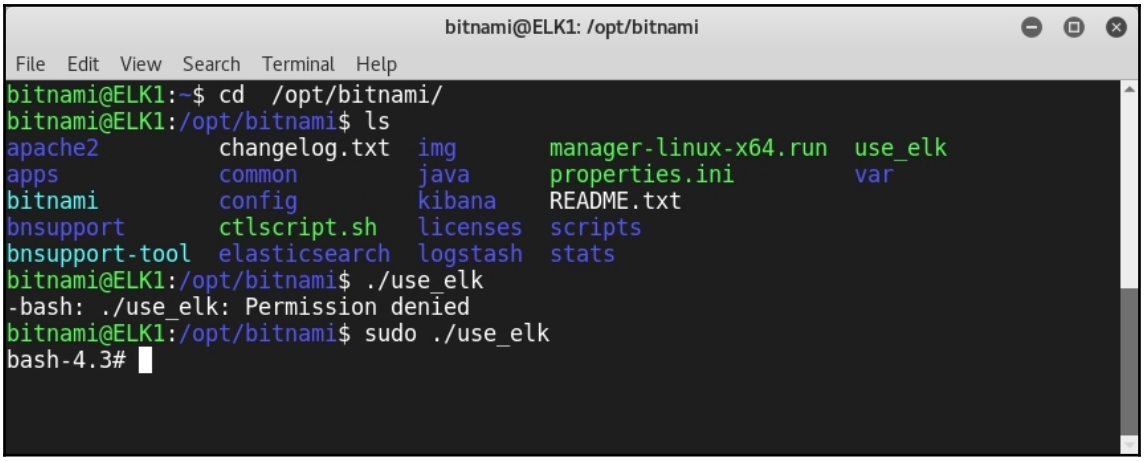

root@ELKStack: ~
File Edit View Search Terminal Help
root@ELKStack:~# echo 'deb http://packages.elastic.co/logstash/2.2/debi
an stable main' | sudo tee /etc/apt/sources.list.d/logstash-2.2.x.list
deb http://packages.elastic.co/logstash/2.2/debian stable main
root@ELKStack:~# sudo apt-get update
Ign http://azure.archive.ubuntu.com trusty InRelease
6% [Connecting to security.ubuntu.com (91.189.91.26)] [Connecting to
Hit ht
tp://ppa.launchpad.net trusty InRelease
11% [Waiting for headers] [Connecting to security.ubuntu.com (91.189
Get:1
http://azure.archive.ubuntu.com trusty-updates InRelease [65.9 kB]
Hit
http://azure.archive.ubuntu.com trusty-backports InRelease
100% [Connecting to security.ubuntu.com (91.189.91.26)] [Connecting 100
% [InRelease gpgv 15.5 kB] [Waiting for headers] [Connecting to s
Hit http:
//azure.archive.ubuntu.com trusty Release.gpg

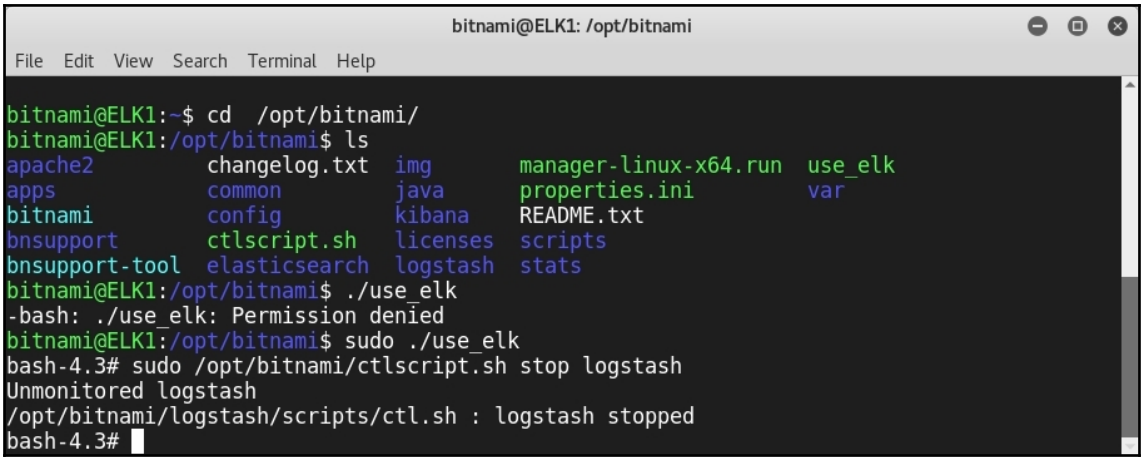

root@ELKStack: ~
File Edit View Search Terminal Help
root@ELKStack:~# sudo apt-get install logstash
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
logstash
0 upgraded, 1 newly installed, 0 to remove and 14 not upgraded.
Need to get 140 MB of archives.
After this operation, 234 MB of additional disk space will be used.
Get:1 https://artifacts.elastic.co/packages/6.x/apt/ stable/main logsta
sh all 1:6.2.3-1 [140 MB]
100% [1 logstash 140 MB] 22.6
100% [1 logstash 140 MB] 22.6
100% [1 logstash 140 MB] 22.6
100% [1 logstash 140 MB] 22.6
100% [1 logstash 140 MB] 22.6
100% [1 logstash 140 MB] 22.6

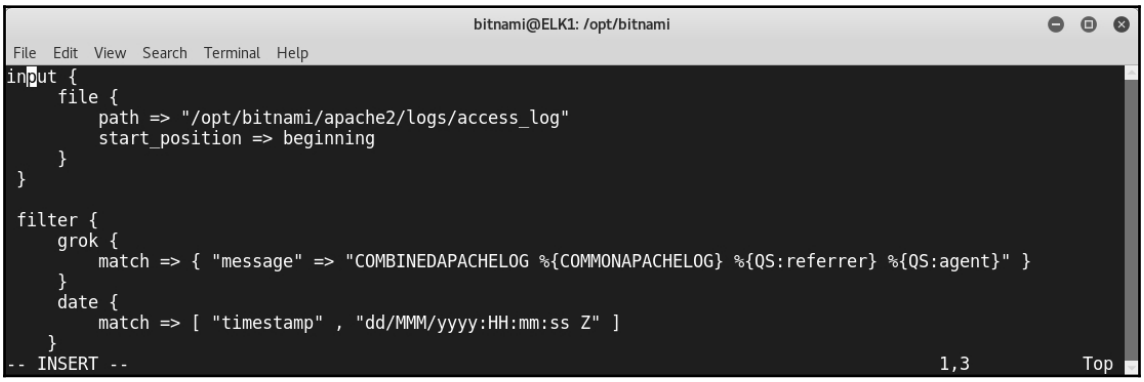

bitnami@ELK1: ~/stack/logstash/conf
File Edit View Search Terminal Help
input
{
beats
{
ssl => false
host => "0.0.0.0"
port => 5044
}
gelf
{
host => "0.0.0.0"
port => 12201
}
http
{
ssl => false
1,0-1 Top


bitnami@ELK1: ~/stack
File Edit View Search Terminal Help
bitnami@ELK1:~/stack$ ls -l
total 3444
drwxr-xr-x 14 root root 4096 Feb 22 12:44 apache2
drwxr-xr-x 3 root root 4096 Feb 22 12:44 apps
drwxr-xr-x 2 root root 4096 Feb 22 12:45 bnsupport
lrwxrwxrwx 1 root root 52 Feb 22 12:45 bnsupport-tool -> /opt/bitnami/bns
upport/bnsupport-0.6.0-linux-x64.run
-rw-r--r-- 1 root root 3677 Feb 22 12:50 changelog.txt
drwxr-xr-x 10 root root 4096 Feb 22 12:43 common
drwxr-xr-x 4 root root 4096 Feb 22 12:45 config
-rwxr-xr-x 1 root root 51890 Feb 22 12:45 ctlscript.sh
drwxr-xr-x 12 root root 4096 Feb 22 12:48 elasticsearch
drwxr-xr-x 2 root root 4096 Feb 22 12:45 img
drwxr-xr-x 7 root root 4096 Feb 22 12:51 java
drwxr-xr-x 17 root root 4096 Feb 22 12:48 kibana
drwxr-xr-x 2 root root 4096 Feb 22 12:45 licenses
drwxr-xr-x 17 root root 4096 Feb 22 12:50 logstash
-rwx------ 1 root root 3390885 Feb 27 2017 manager-linux-x64.run

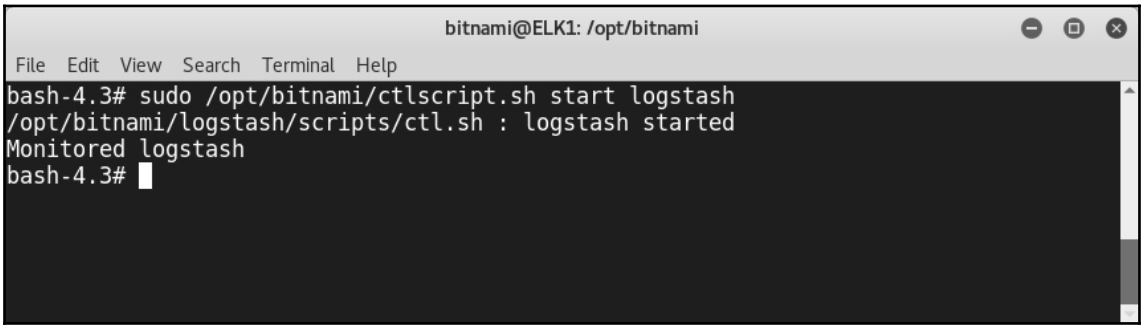

[OK] Started Snappy daemon.
[36.681718] bitnami[1549]: ##
[36.695733] bitnami[1549]: # #
[36.722314] bitnami[1549]: # Setting Bitnami application password to 'E7DRnqi1OAJp' #
[36.747684] bitnami[1549]: # (the default application username is 'user') #
[36.783504] bitnami[1549]: # #
[36.817251] bitnami[1549]: ##
[37.628951] bitnami[1549]: [Tue Apr 3 12:47:20 UTC 2018] Regenerating keys for apache2
[38.204812] cloud-init[1386]: Cloud-init v. 17.1 running 'modules:config' at Tue, 03 Apr 2018 12:47:13 +0000. Up 31.47 secon


bitnami@ELK1: /opt/bitnami
File Edit View Search Terminal Help
bitnami@ELK1:~$ cd /opt/bitnami/
bitnami@ELK1:/opt/bitnami$ ls
apache2 changelog.txt img manager-linux-x64.run use_elk
apps common java properties.ini var
bitnami config kibana README.txt
bnsupport ctlscript.sh licenses scripts
bnsupport-tool elasticsearch logstash stats
bitnami@ELK1:/opt/bitnami$ ./use_elk
-bash: ./use_elk: Permission denied
bitnami@ELK1:/opt/bitnami$ sudo ./use_elk
bash-4.3#

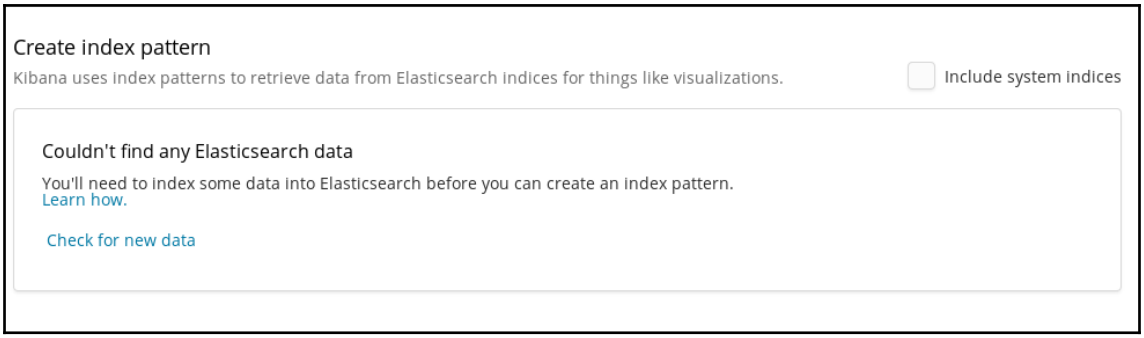

bitnami@ELK1: /opt/bitnami
File Edit View Search Terminal Help
bitnami@ELK1:~$ cd /opt/bitnami/
bitnami@ELK1:/opt/bitnami$ ls
apache2 changelog.txt img manager-linux-x64.run use_elk
apps common java properties.ini var
bitnami config kibana README.txt
bnsupport ctlscript.sh licenses scripts
bnsupport-tool elasticsearch logstash stats
bitnami@ELK1:/opt/bitnami$ ./use_elk
-bash: ./use_elk: Permission denied
bitnami@ELK1:/opt/bitnami$ sudo ./use_elk
bash-4.3# sudo /opt/bitnami/ctlscript.sh stop logstash
Unmonitored logstash
/opt/bitnami/logstash/scripts/ctl.sh : logstash stopped
bash-4.3#

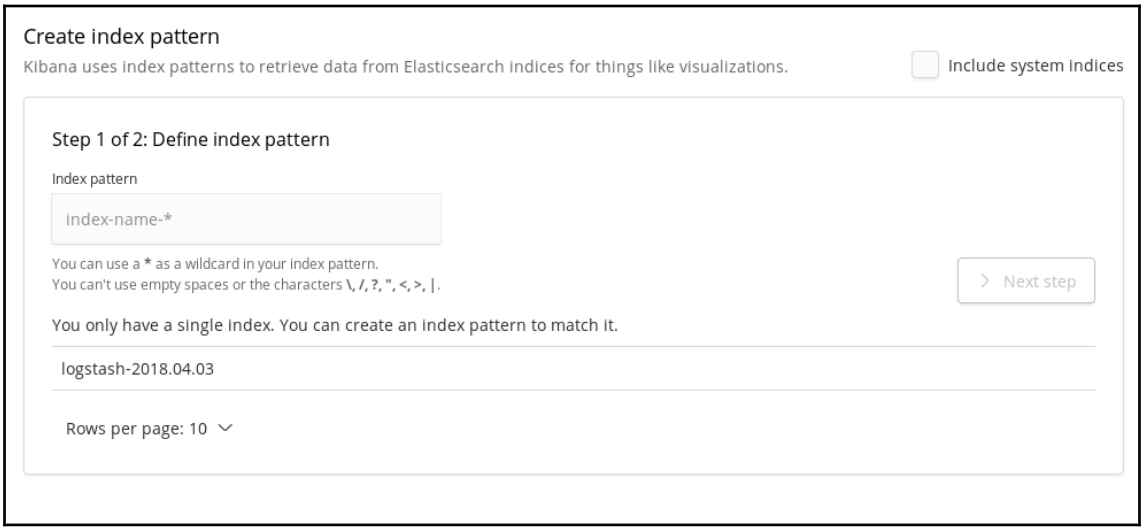

bitnami@ELK1: /opt/bitnami
File Edit View Search Terminal Help
input {
file {
path => "/opt/bitnami/apache2/logs/access_log"
start_position => beginning
}
}
filter {
grok {
match => { "message" => "COMBINEDAPACHELOG %{COMMONAPACHELOG} %{QS:referrer} %{QS:agent}" }
}
date {
match => ["timestamp" , "dd/MMM/yyyy:HH:mm:ss Z"]
}
-- INSERT -- 1,3 Top


bitnami@ELK1: /opt/bitnami
File Edit View Search Terminal Help
bash-4.3# vi access-log.conf
bash-4.3# ls
access-log.conf logstash.conf
_and_exit /opt/bitnami/logstash/bin/logstash -f /opt/bitnami/logstash/conf/ --config.test
Sending Logstash's logs to /opt/bitnami/logstash/logs which is now configured via log4j2.properties
[2018-04-03T15:36:56,558][INFO][logstash.modules.scaffold] Initializing module {:module_name=>"fb_apache", :directory=>"
/opt/bitnami/logstash/modules/fb_apache/configuration"}
[2018-04-03T15:36:56,583][INFO][logstash.modules.scaffold] Initializing module {:module_name=>"netflow", :directory=>"/o
pt/bitnami/logstash/modules/netflow/configuration"}
[2018-04-03T15:36:57,461][WARN][logstash.config.source.multilocal] Ignoring the 'pipelines.yml' file because modules or
command line options are specified
Configuration OK
[2018-04-03T15:37:03,712][INFO][logstash.runner   ] Using config.test_and_exit mode. Config Validation Result: OK
. Exiting Logstash
bash-4.3#

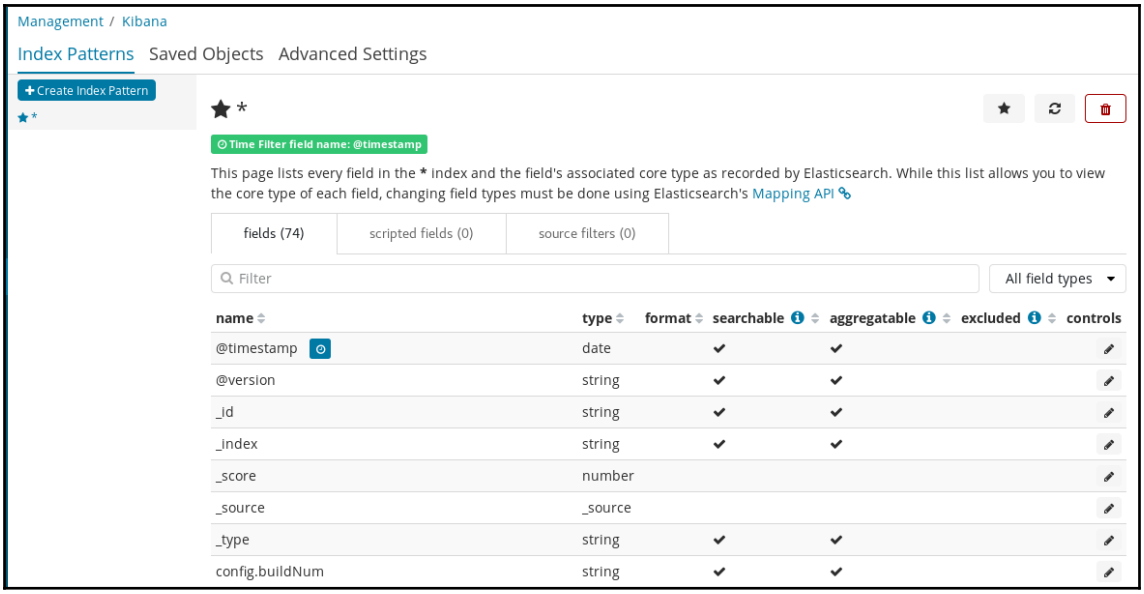

bitnami@ELK1: /opt/bitnami
File Edit View Search Terminal Help
bash-4.3# sudo /opt/bitnami/ctlscript.sh start logstash
/opt/bitnami/logstash/scripts/ctl.sh : logstash started
Monitored logstash
bash-4.3#


bitnami@ELK1: /opt/bitnami
File Edit View Search Terminal Help
bash-4.3# curl 'localhost:9200/_cat/indices?v'
health status index   uuid   pri rep docs.count docs.deleted store.
size pri.store.size
yellow open   logstash-2018.04.03 llvYcU5jTXCFeWq12c-XWw   5   1   295   270   191
.7kb   191.7kb
bash-4.3#

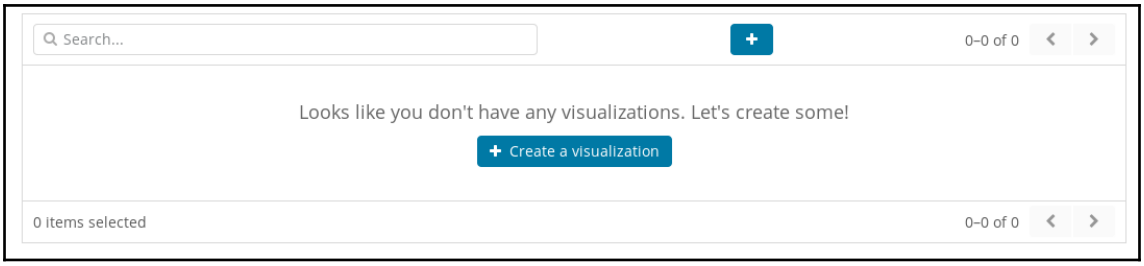

Create index pattern
Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.
Include system indices
Couldn't find any Elasticsearch data
You'll need to index some data into Elasticsearch before you can create an index pattern.
Learn how.
Check for new data

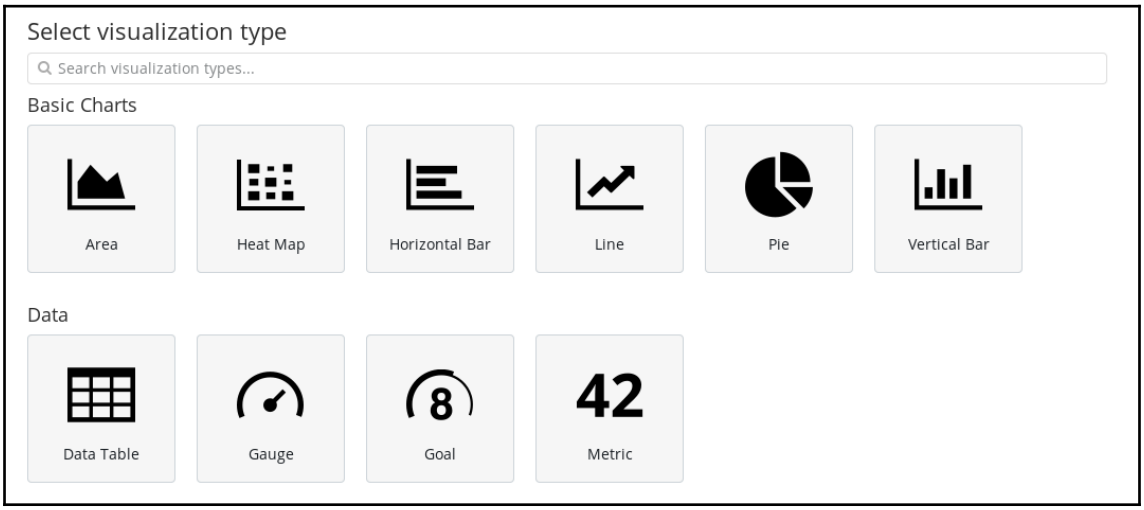

Create index pattern
Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.
Include system indices
Step 1 of 2: Define index pattern
Index pattern
index-name-*
You can use a * as a wildcard in your index pattern.
You can't use empty spaces or the characters \, /, ?, ", <, >, |.
Next step
You only have a single index. You can create an index pattern to match it.
logstash-2018.04.03
Rows per page: 10

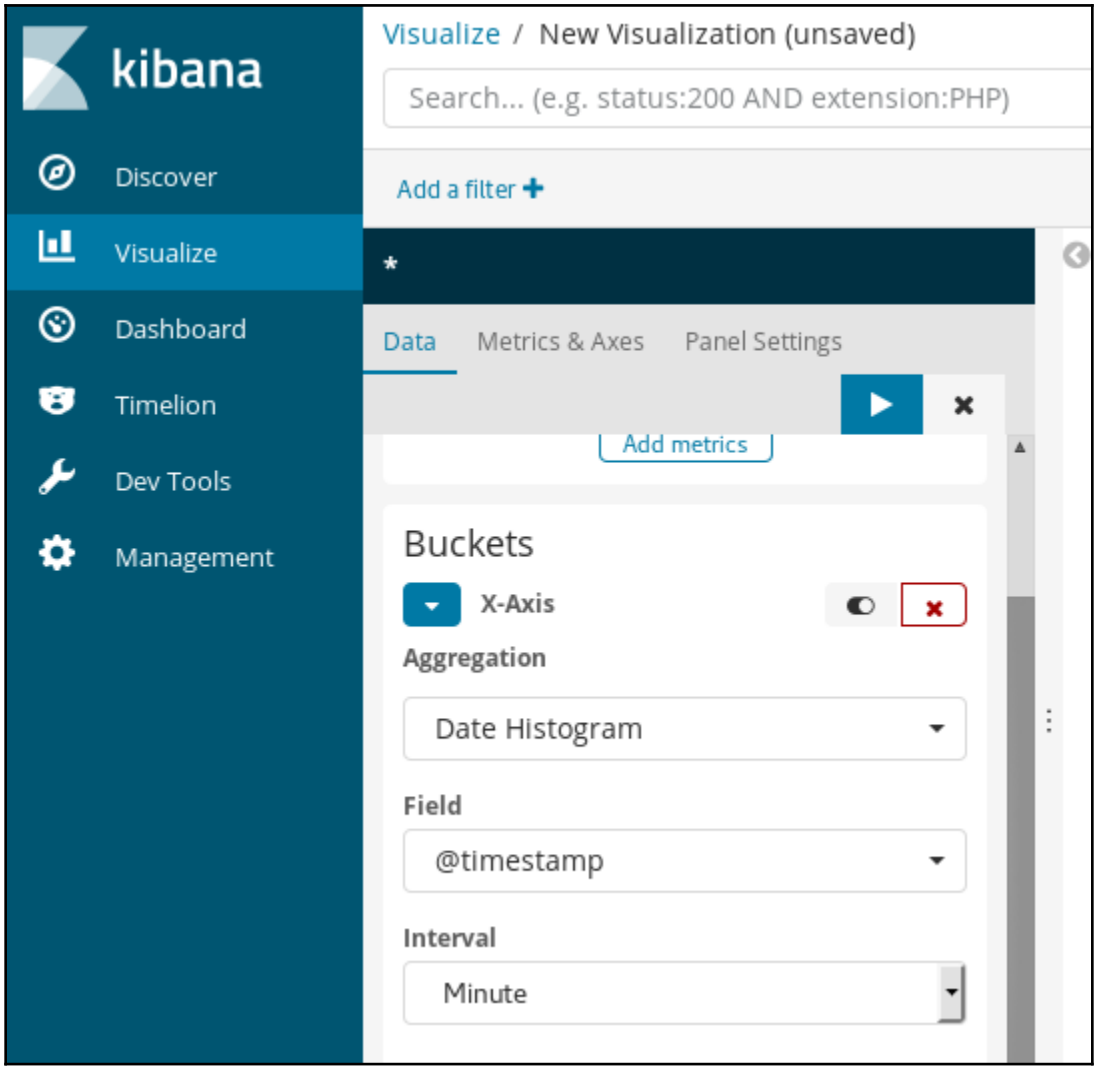

Step 1 of 2: Define index pattern
Index pattern
*
You can use a * as a wildcard in your index pattern.
You can't use empty spaces or the characters \, /, ?, ", <, >, |.
Next step
Success! Your index pattern matches 1 index.
logstash-2018.04.03
Rows per page: 10

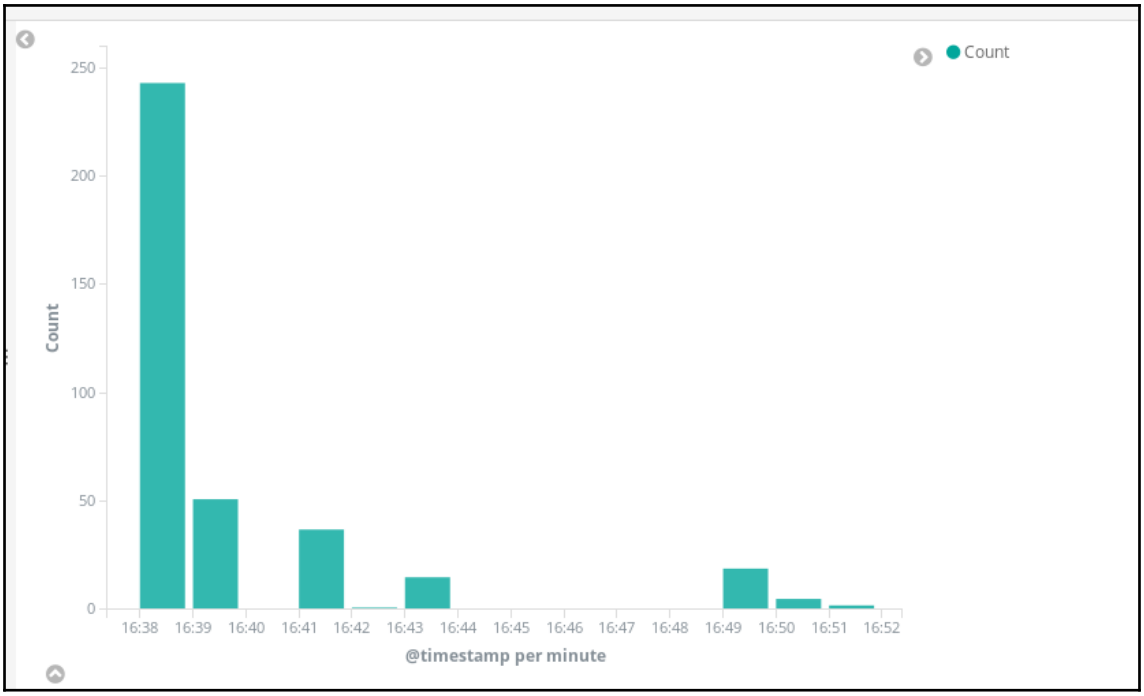

Management / Kibana
Index Patterns
Saved Objects
Advanced Settings
+ Create Index Pattern
* *
Time Filter field name: @timestamp
This page lists every field in the * index and the field's associated core type as recorded by Elasticsearch. While this list allows you to view the core type of each field, changing field types must be done using Elasticsearch's Mapping API
fields (74)
scripted fields (0)
source filters (0)
Filter
All field types
name
type
format
searchable
aggregatable
excluded
controls
@timestamp
date
@version
string
_id
string
_index
string
_score
number
_source
_source
_type
string
config.buildNum
string

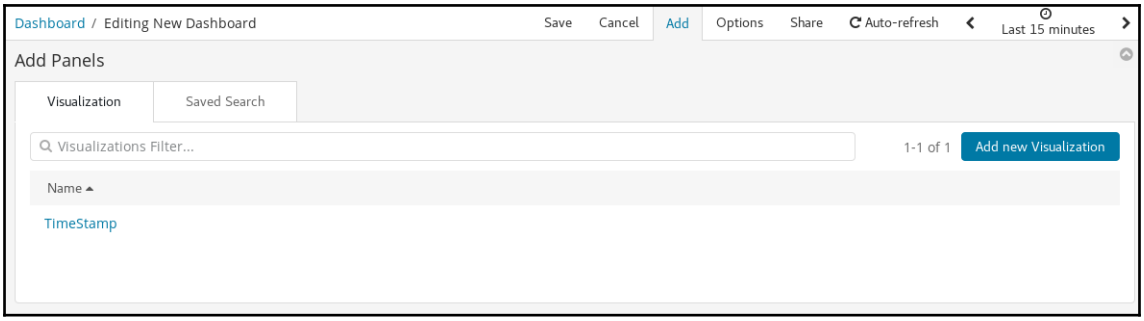

kibana
343 hits
New
Save
Open
Share
Auto-refresh
Last 15 minutes
Search... (e.g. status:200 AND extension:PHP)
Uses lucene query syntax
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Add a filter
Selected Fields
source
Available Fields
@timestamp
@version
id
index
score
type
host
message
path
tags
April 3rd 2018, 16:28:37.271 - April 3rd 2018, 16:43:37.272 —
Auto
Count
@timestamp per 30 seconds
Time
_source
April 3rd 2018, 16:43:07.255
@timestamp: April 3rd 2018, 16:43:07.255 path: /opt/bitnami/apache2/logs/access_log @version: 1 host: ELK1 message: 41.227.170.142 - user [03/Apr/2018:15:43:06 +0000] "GET /elk/ui/favicons/favicon.ico HTTP/1.1" 304 - tags: _grokparsefailure _id: %{logstash_checksum} _type: doc _index: logstash-2018.04.03 _score: -
April 3rd 2018, 16:43:07.255
@timestamp: April 3rd 2018, 16:43:07.255 path: /opt/bitnami/apache2/logs/access_log @version: 1 host: ELK1 message: 41.227.170.142 - user [03/Apr/2018:15:43:06 +0000] "GET /elk/ui/favicons/


Search...
0–0 of 0
Looks like you don't have any visualizations. Let's create some!
Create a visualization
0 items selected
0–0 of 0

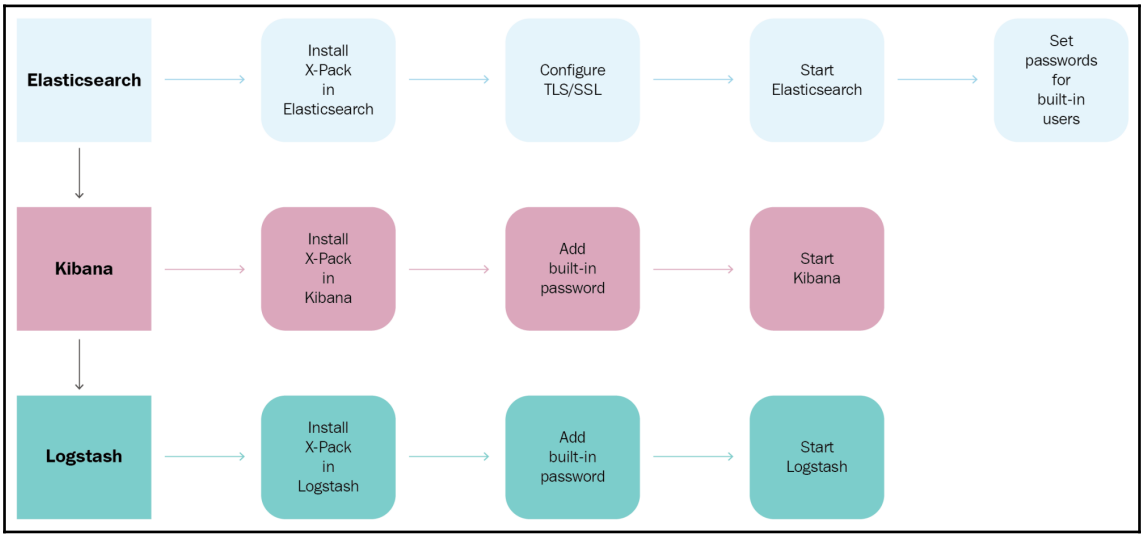

Select visualization type
Search visualization types...
Basic Charts
Area
Heat Map
Horizontal Bar
Line
Pie
Vertical Bar
Data
Data Table
Gauge
Goal
42
Metric

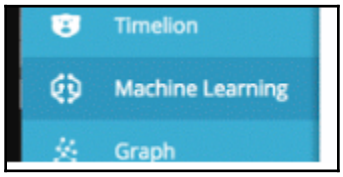

kibana
Discover
Visualize
Dashboard
Timelion
Dev Tools
Management
Visualize / New Visualization (unsaved)
Search... (e.g. status:200 AND extension:PHP)
Add a filter +
*
Data
Metrics & Axes
Panel Settings
Add metrics
Buckets
X-Axis
Aggregation
Date Histogram
Field
@timestamp
Interval
Minute

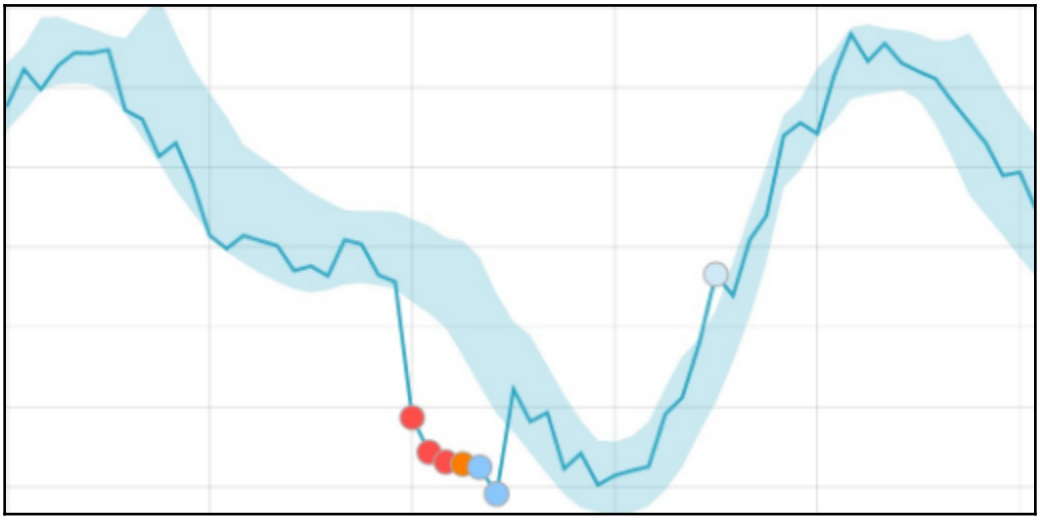

Count
250
200
150
100
50
0
Count
16:38
16:39
16:40
16:41
16:42
16:43
16:44
16:45
16:46
16:47
16:48
16:49
16:50
16:51
16:52
@timestamp per minute


Dashboard / Editing New Dashboard
Save
Cancel
Add
Options
Share
Auto-refresh
Last 15 minutes
Add Panels
Visualization
Saved Search
Visualizations Filter...
1-1 of 1
Add new Visualization
Name
TimeStamp

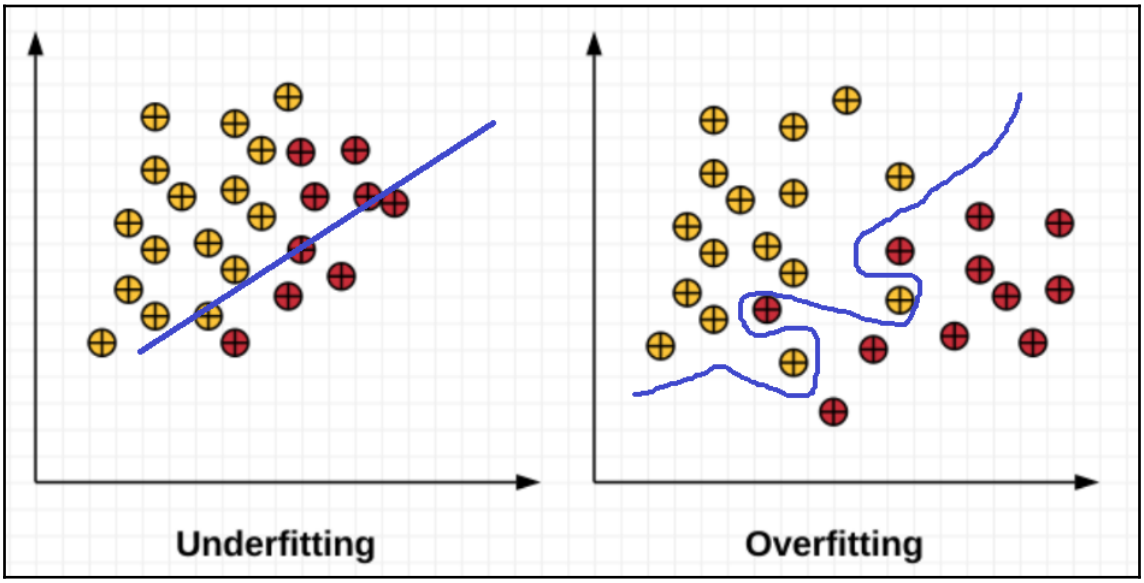

Search... (e.g. status:200 AND extension:PHP)
Add a filter +
TimeStamp
Count
40
30
20
10
0
Count
16:45
16:50
@timestamp per minute

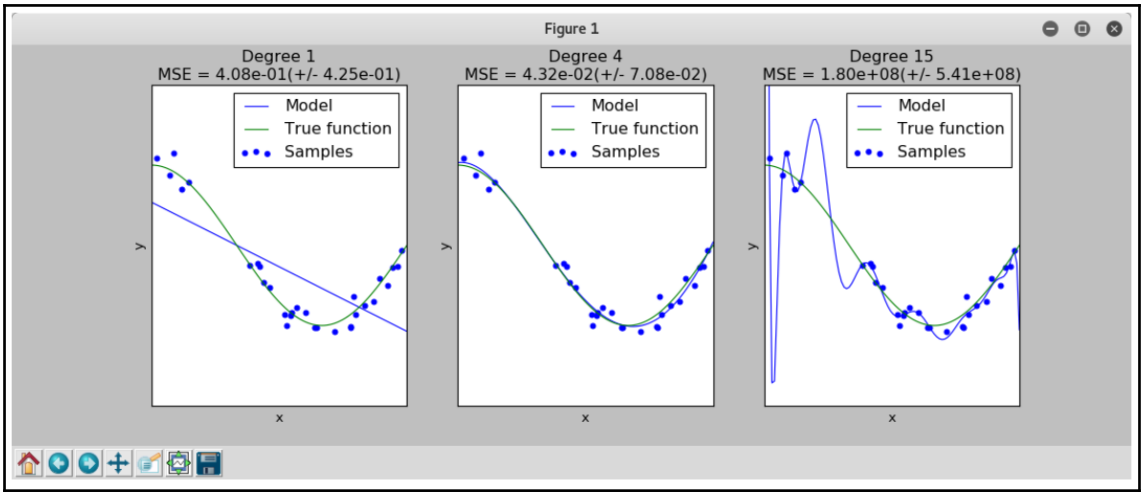

Elasticsearch
Install X-Pack in Elasticsearch
Configure TLS/SSL
Start Elasticsearch
Set passwords for built-in users
Kibana
Install X-Pack in Kibana
Add built-in password
Start Kibana
Logstash
Install X-Pack in Logstash
Add built-in password
Start Logstash

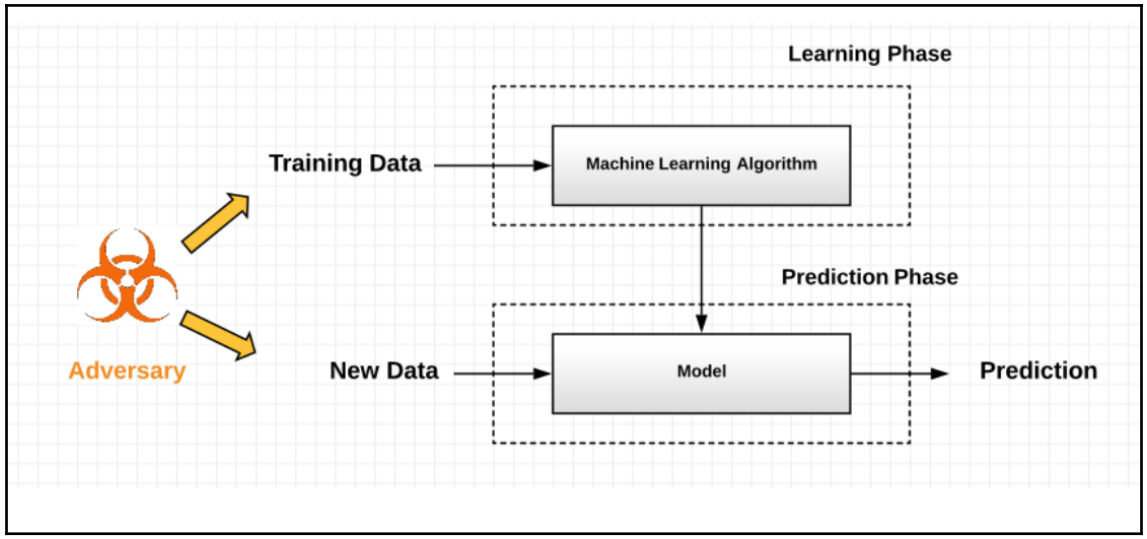

Timelion
Machine Learning
Graph

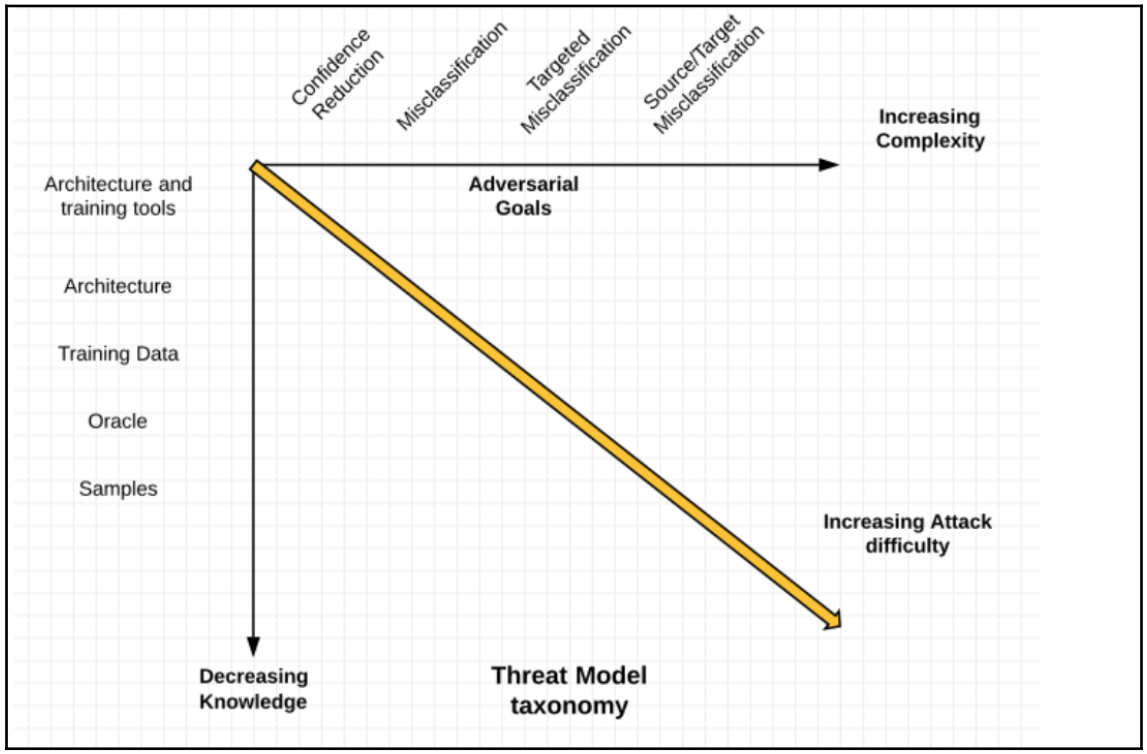


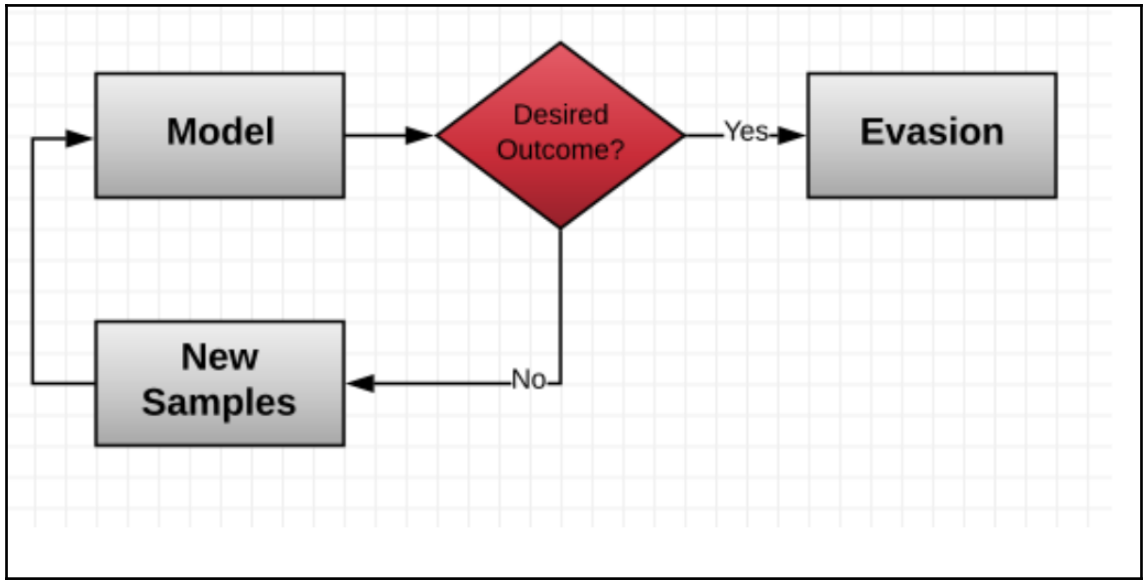
Chapter 8: Evading Intrusion Detection Systems

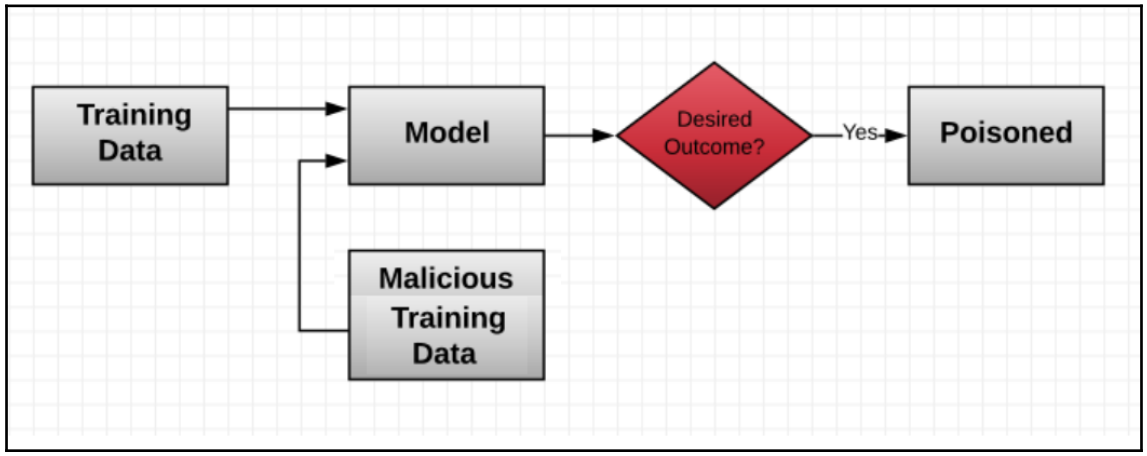


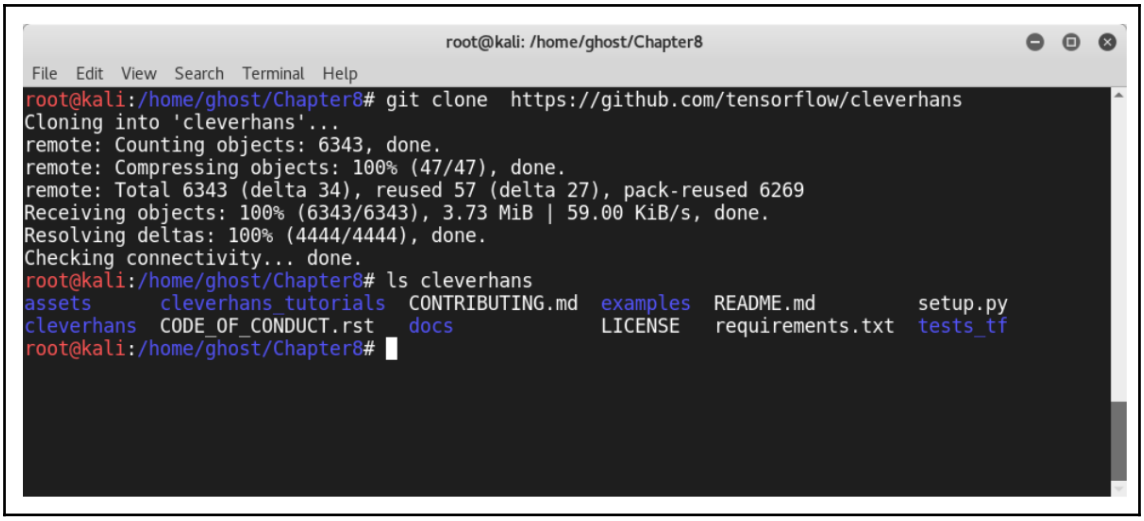


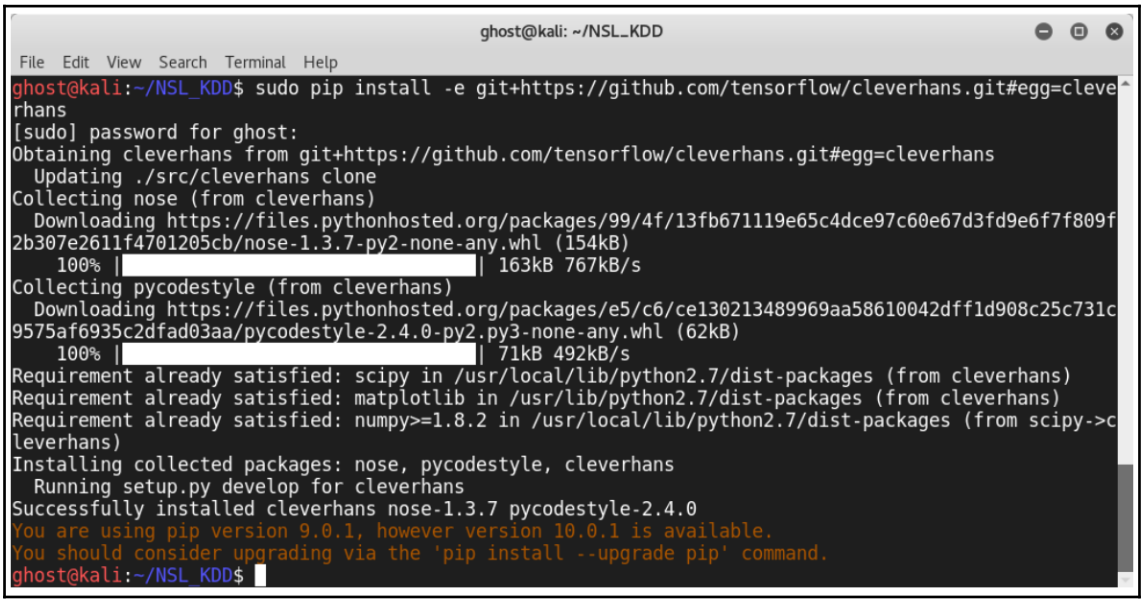

Learning Phase
Training Data
Machine Learning Algorithm
Prediction Phase
Adversary
New Data
Model
Prediction

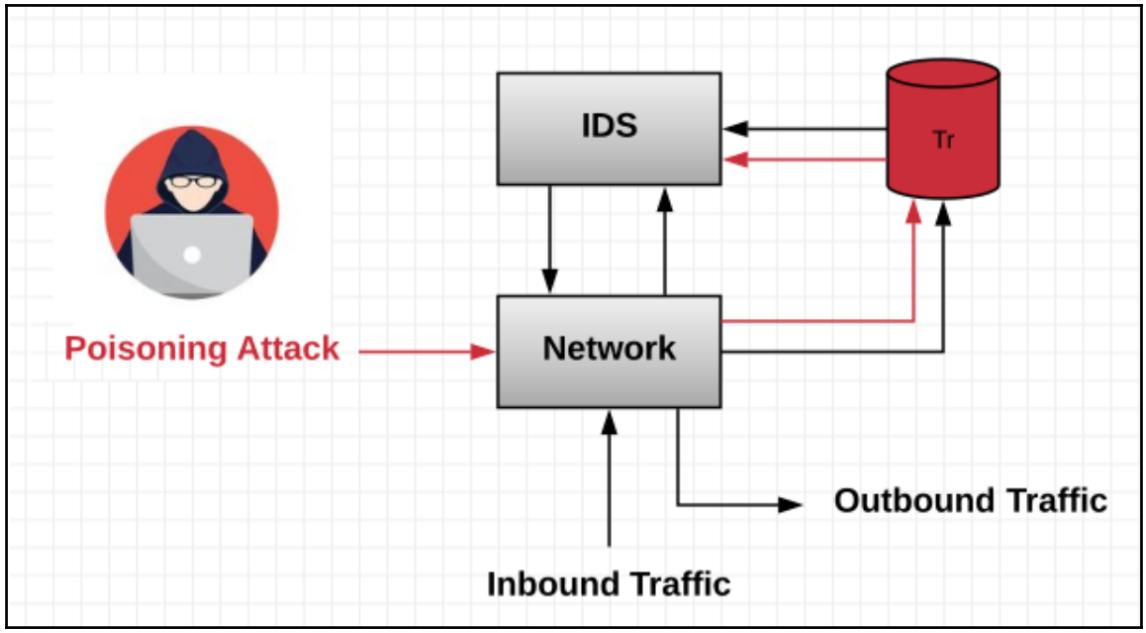

Confidence Reduction
Misclassification
Targeted Misclassification
Source/Target Misclassification
Increasing Complexity
Architecture and training tools
Adversarial Goals
Architecture
Training Data
Oracle
Samples
Increasing Attack difficulty
Decreasing Knowledge
Threat Model taxonomy

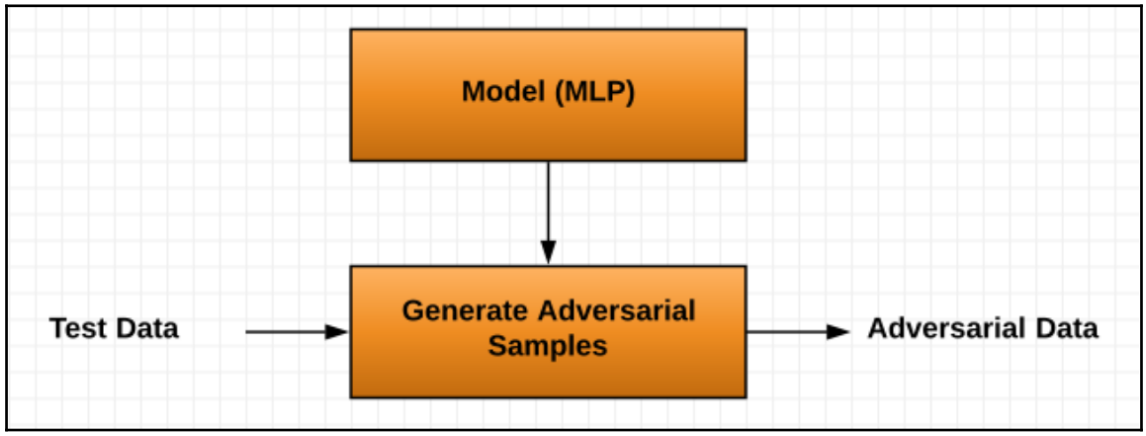

Model
Desired Outcome?
Yes
Evasion
New Samples
No

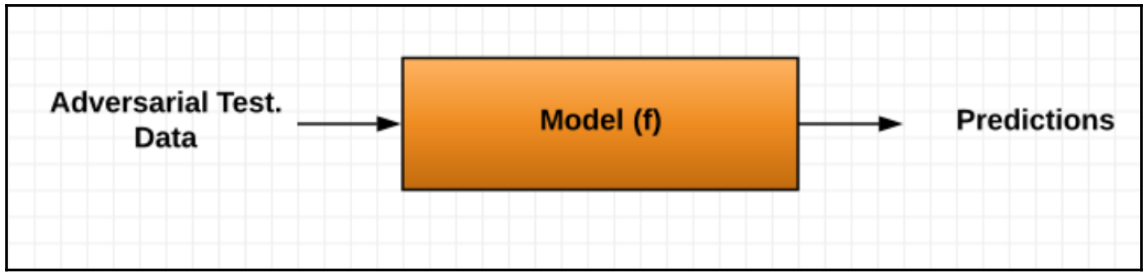

Training Data
Model
Desired Outcome?
Yes
Poisoned
Malicious Training Data


root@kali: /home/ghost/Chapter8
File Edit View Search Terminal Help
root@kali:/home/ghost/Chapter8# git clone https://github.com/tensorflow/cleverhans
Cloning into 'cleverhans'...
remote: Counting objects: 6343, done.
remote: Compressing objects: 100% (47/47), done.
remote: Total 6343 (delta 34), reused 57 (delta 27), pack-reused 6269
Receiving objects: 100% (6343/6343), 3.73 MiB | 59.00 KiB/s, done.
Resolving deltas: 100% (4444/4444), done.
Checking connectivity... done.
root@kali:/home/ghost/Chapter8# ls cleverhans
assets cleverhans_tutorials CONTRIBUTING.md examples README.md setup.py
cleverhans CODE_OF_CONDUCT.rst docs LICENSE requirements.txt tests_tf
root@kali:/home/ghost/Chapter8#

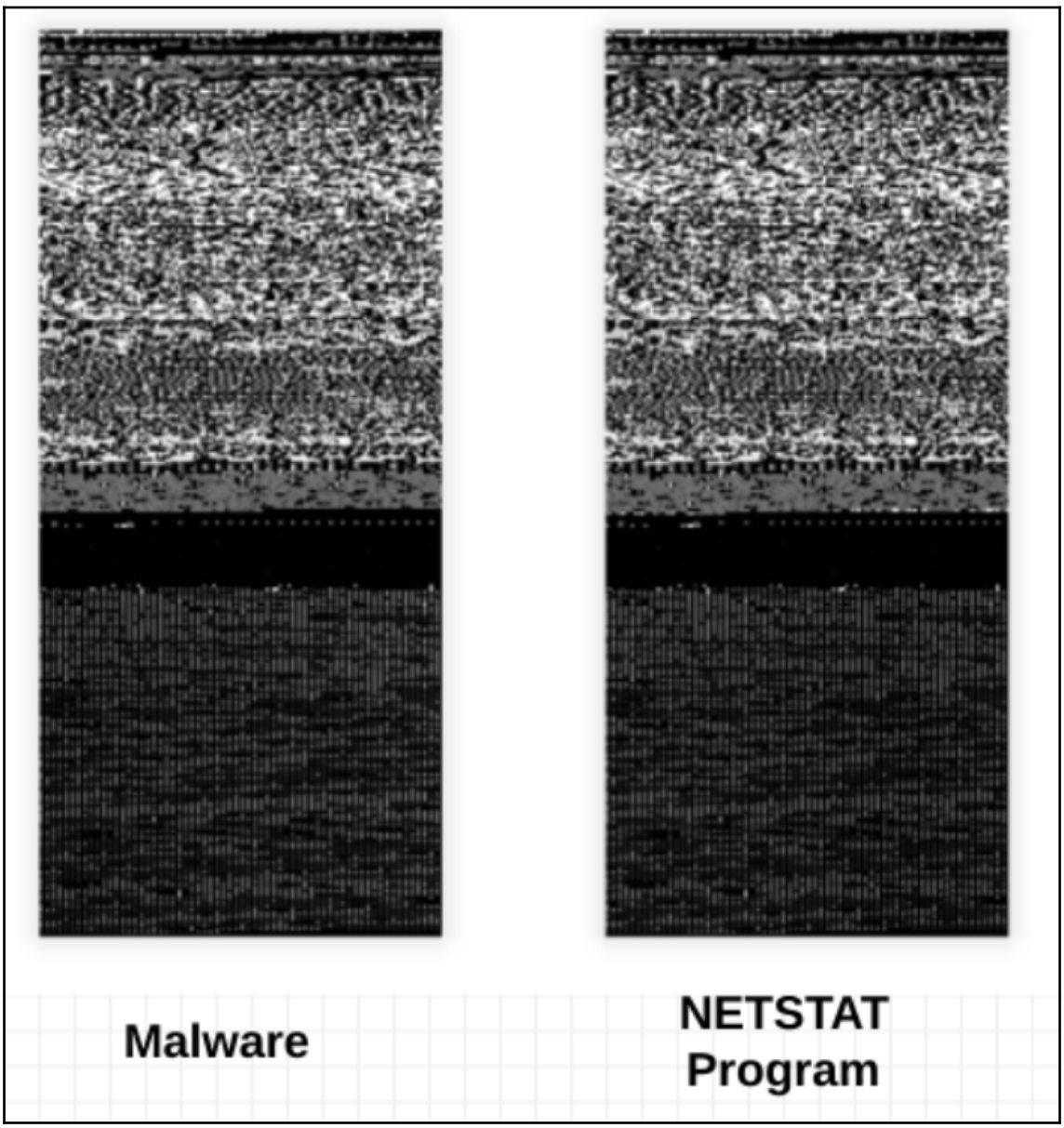

ghost@kali: ~/NSL_KDD
File Edit View Search Terminal Help
ghost@kali:~/NSL_KDD$ sudo pip install -e git+https://github.com/tensorflow/cleverhans.git#egg=cleverhans
[sudo] password for ghost:
Obtaining cleverhans from git+https://github.com/tensorflow/cleverhans.git#egg=cleverhans
Updating ./src/cleverhans clone
Collecting nose (from cleverhans)
Downloading https://files.pythonhosted.org/packages/99/4f/13fb671119e65c4dce97c60e67d3fd9e6f7f809f2b307e2611f4701205cb/nose-1.3.7-py2-none-any.whl (154kB)
100% | 163kB 767kB/s
Collecting pycodestyle (from cleverhans)
Downloading https://files.pythonhosted.org/packages/e5/c6/ce130213489969aa58610042dff1d908c25c731c9575af6935c2dfad03aa/pycodestyle-2.4.0-py2.py3-none-any.whl (62kB)
100% | 71kB 492kB/s
Requirement already satisfied: scipy in /usr/local/lib/python2.7/dist-packages (from cleverhans)
Requirement already satisfied: matplotlib in /usr/lib/python2.7/dist-packages (from cleverhans)
Requirement already satisfied: numpy>=1.8.2 in /usr/local/lib/python2.7/dist-packages (from scipy->cleverhans)
Installing collected packages: nose, pycodestyle, cleverhans
Running setup.py develop for cleverhans
Successfully installed cleverhans nose-1.3.7 pycodestyle-2.4.0
You are using pip version 9.0.1, however version 10.0.1 is available.
You should consider upgrading via the 'pip install --upgrade pip' command.
ghost@kali:~/NSL_KDD$

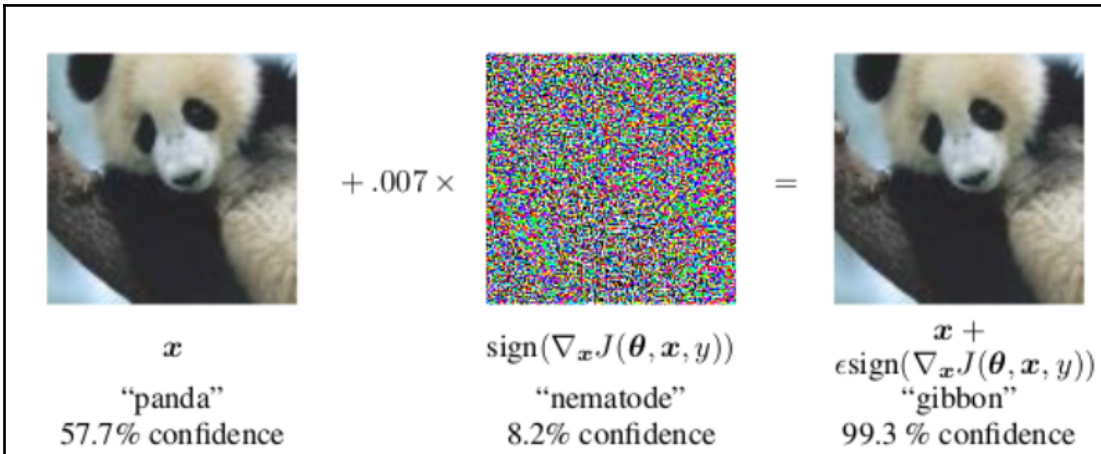

IDS
Tr
Poisoning Attack
Network
Outbound Traffic
Inbound Traffic


Model (MLP)
Test Data
Generate Adversarial Samples
Adversarial Data


Adversarial Test.
Data
Model (f)
Predictions

Chapter 9: Bypassing Machine Learning Malware Detectors






```
root@kali: /home/ghost/Chapter9
File  Edit  View  Search  Terminal  Help

>>> data = test_data[1][0]
>>> activations = Model.feedforward(data)
>>> prediction = np.argmax(activations)
>>> activations
array([[1.92687103e-05],
       [6.88997097e-04],
       [9.99999834e-01],
       [1.20186152e-06],
       [1.15908489e-07],
       [2.82359808e-08],
       [1.41317517e-03],
       [3.08833421e-16],
       [4.22214376e-10],
       [7.63893968e-14]])
>>> prediction
2
>>>
```

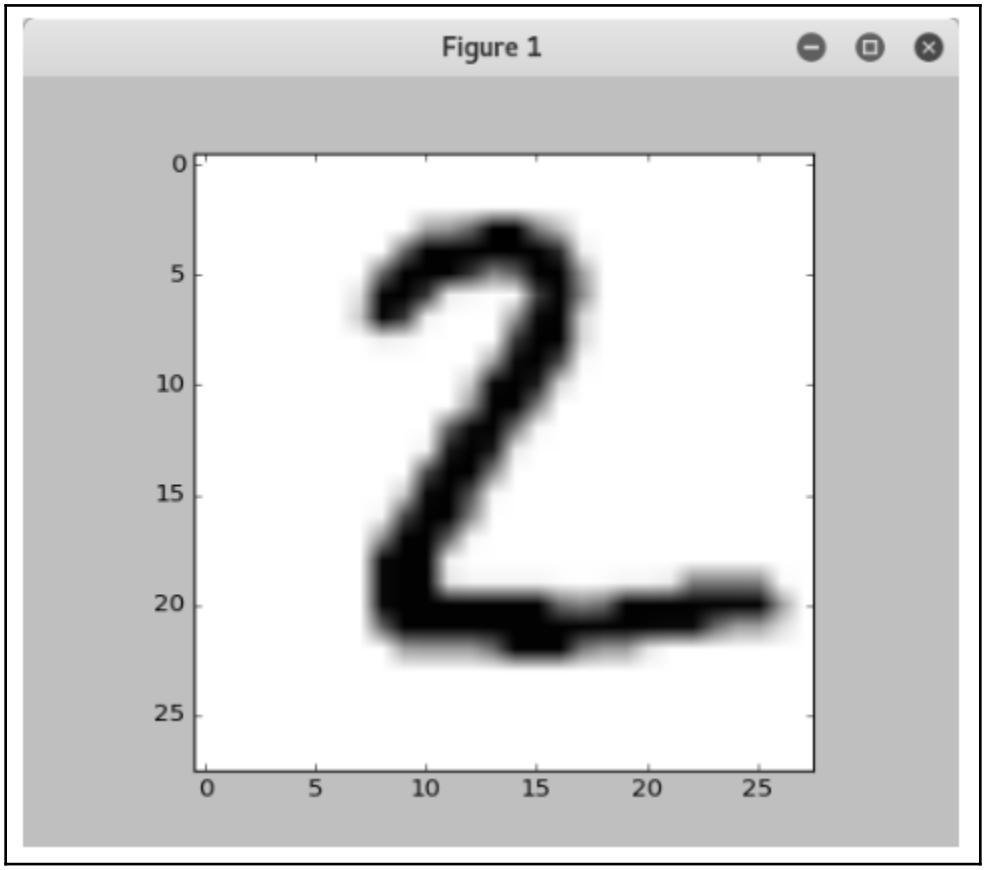
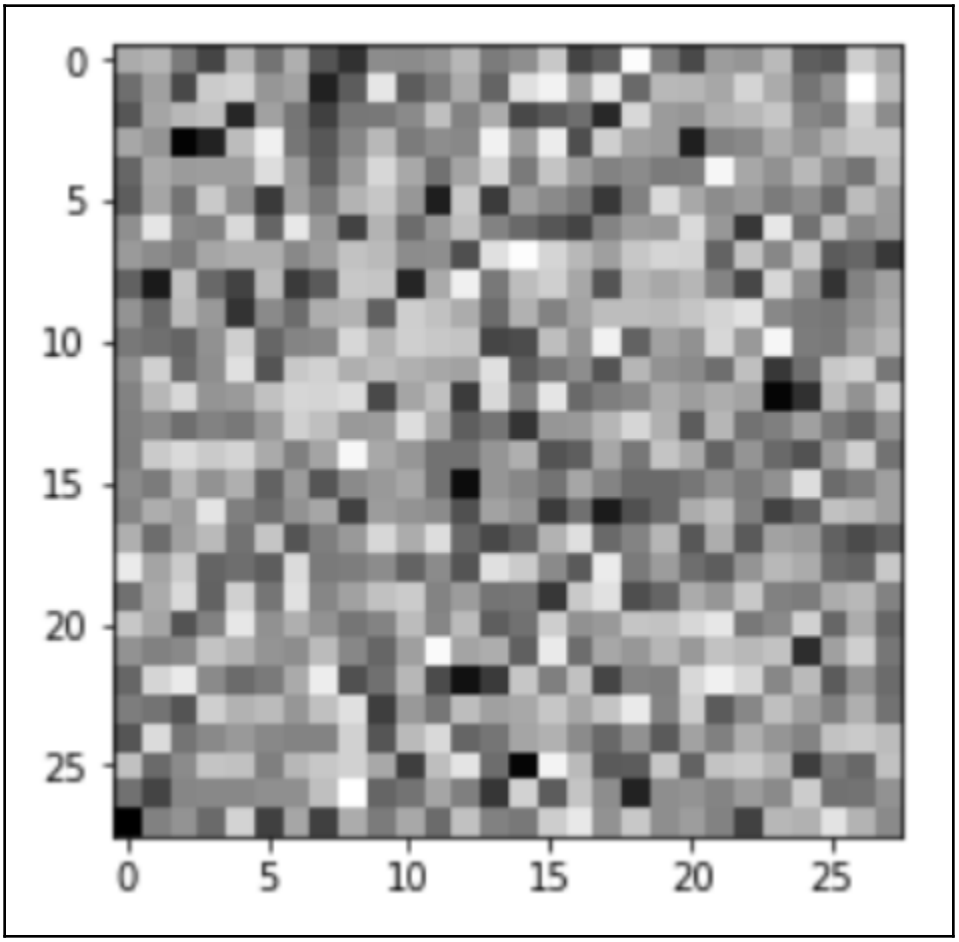


Figure 1
0
5
10
15
20
25
0
5
10
15
20
25

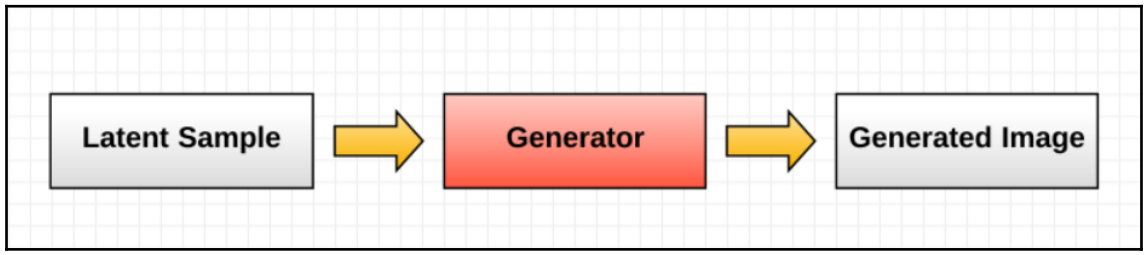
$$Y = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

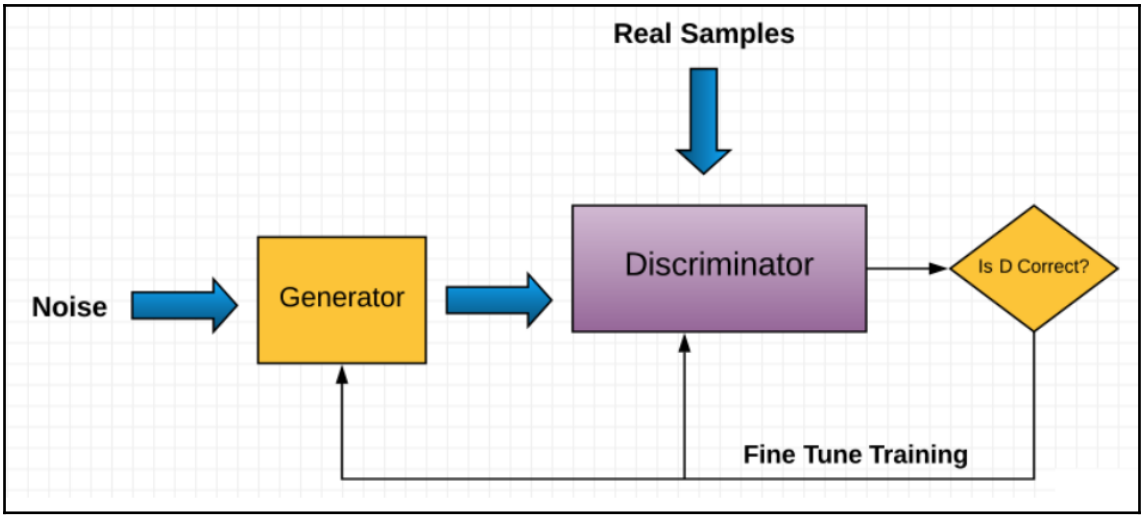
$$C = \frac{1}{2} \times \|Y_{goal} - \hat{y}(\vec{x})\|_2^2$$

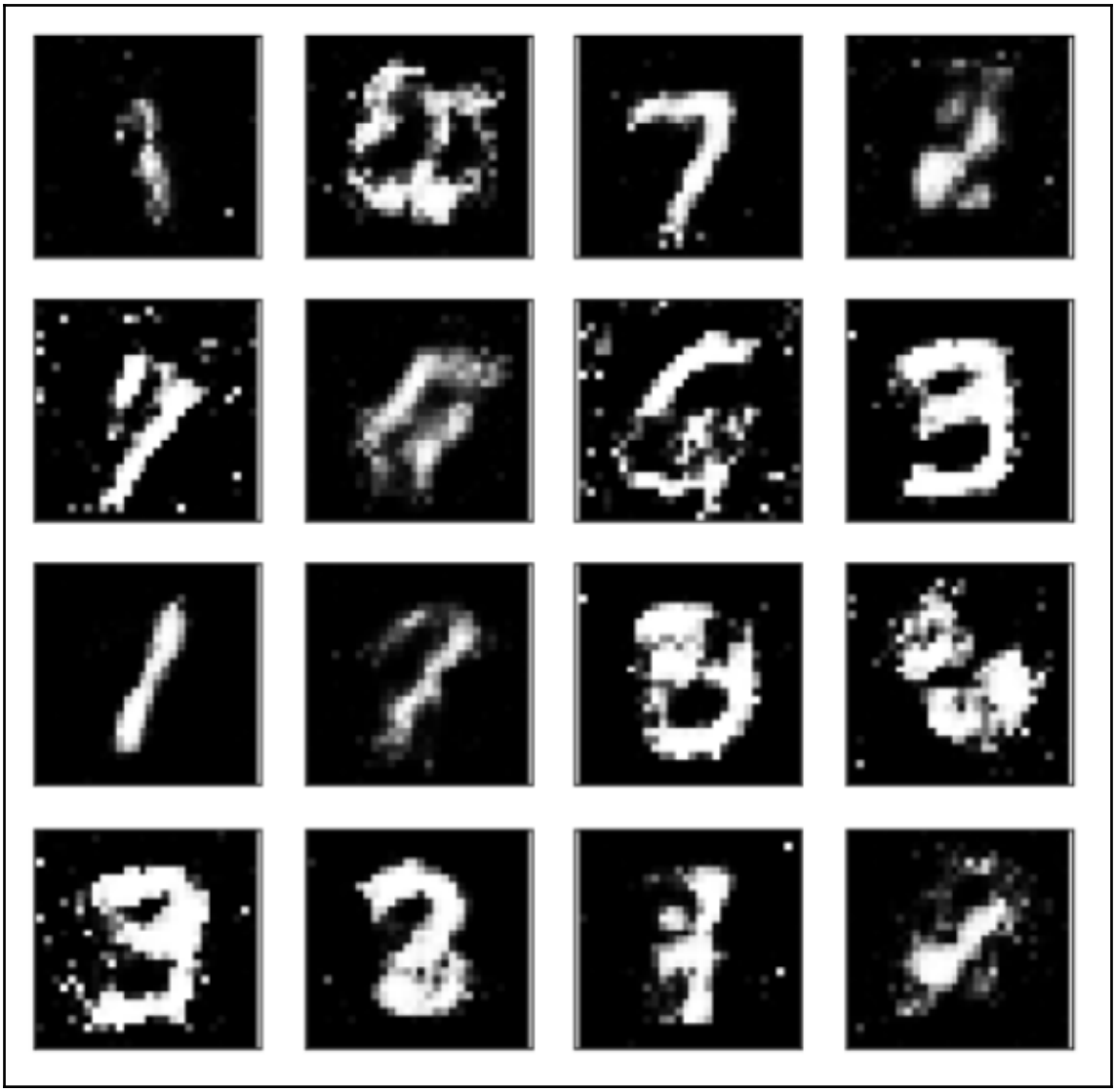


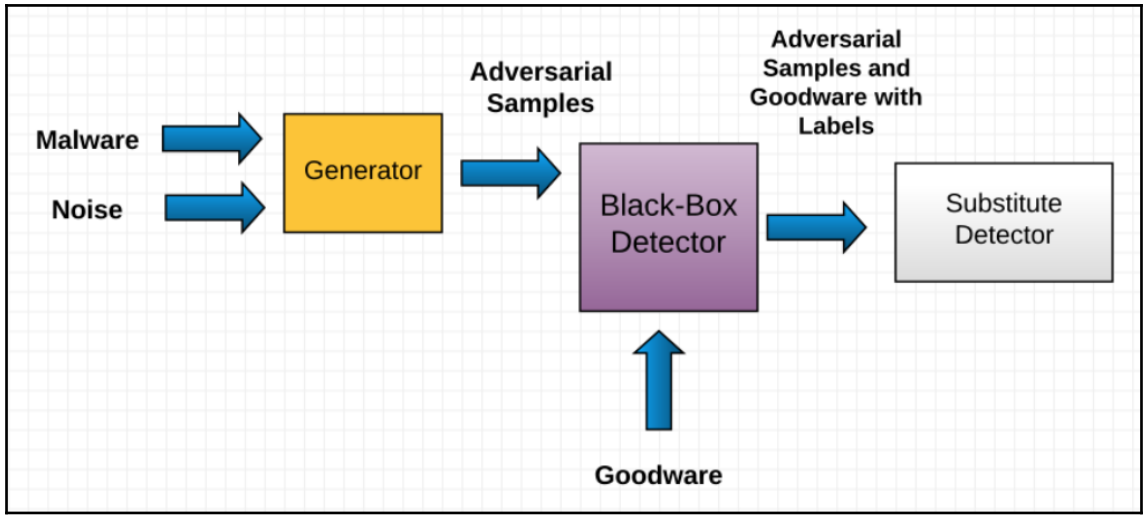
$$C = \frac{1}{2} \times \|Y_{goal} - \hat{y}(\vec{x})\|_2^2 + \lambda \|\vec{x} - X_{target}\|_2^2$$

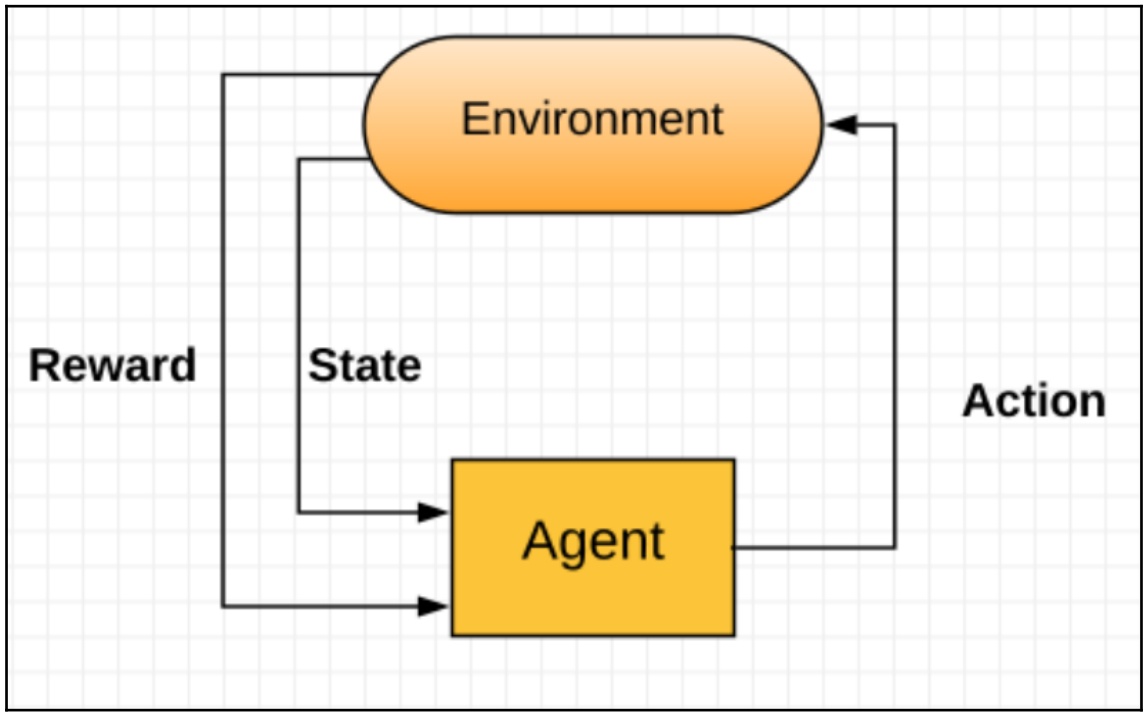
```
root@kali: /home/ghost/Chapter9
File  Edit  View  Search  Terminal  Help
root@kali:/home/ghost/Chapter9# pip install foolbox
Collecting foolbox
  Downloading https://files.pythonhosted.org/packages/52/7e/e9e1ad78b7f3240eaed2c408d9d998a9f6149b433c95f54c2117ef15fa9f/foolbox-1.1.0.tar.gz (218kB)
    100% |████████████████████████████████| 225kB 491kB/s
Requirement already satisfied: numpy in /usr/local/lib/python2.7/dist-packages (from foolbox)
Requirement already satisfied: scipy in /usr/local/lib/python2.7/dist-packages (from foolbox)
Requirement already satisfied: setuptools in /usr/local/lib/python2.7/dist-packages (from foolbox)
Requirement already satisfied: future in /usr/lib/python2.7/dist-packages (from foolbox)
Requirement already satisfied: futures in /usr/lib/python2.7/dist-packages (from foolbox)
Requirement already satisfied: appdirs>=1.4.0 in /usr/local/lib/python2.7/dist-packages (from setuptools->foolbox)
Requirement already satisfied: packaging>=16.8 in /usr/local/lib/python2.7/dist-packages (from setuptools->foolbox)
Requirement already satisfied: six>=1.6.0 in /usr/local/lib/python2.7/dist-packages (from setuptools->foolbox)
Requirement already satisfied: pyparsing in /usr/local/lib/python2.7/dist-packages (from packaging>=16.8->setuptools->foolbox)
Building wheels for collected packages: foolbox
  Running setup.py bdist_wheel for foolbox ... done
  Stored in directory: /root/.cache/pip/wheels/33/6f/94/dacdaeaad3becfcb062161dd399074dbd453ae35b6ca35bbe7
```

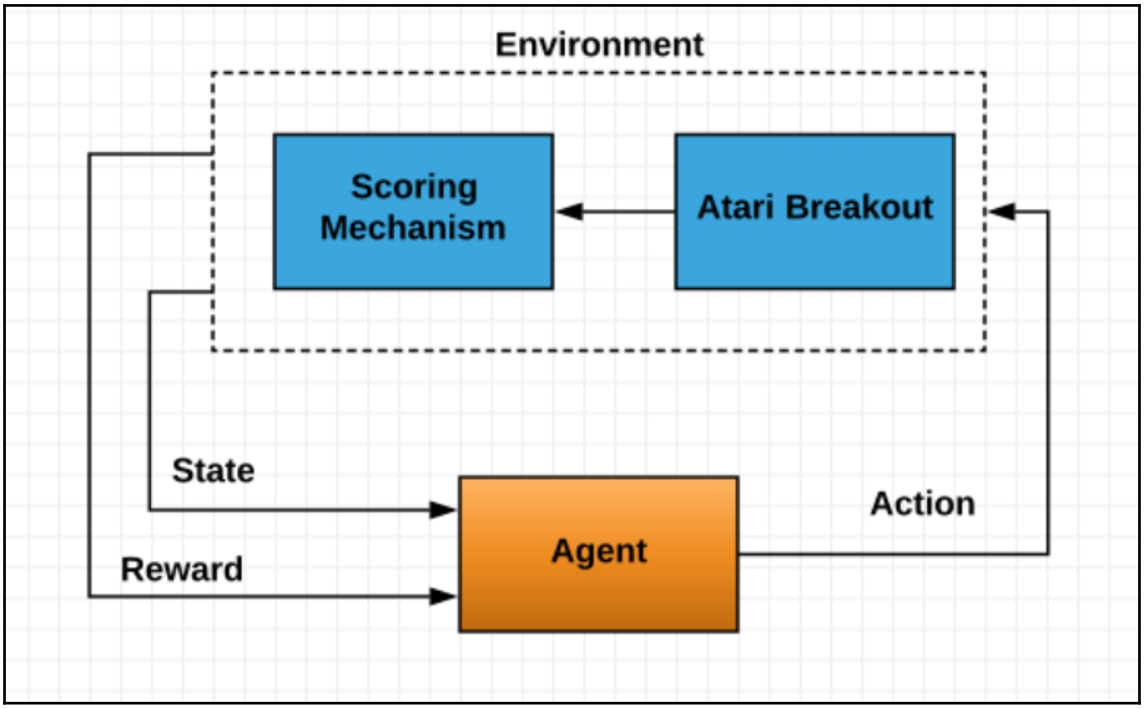


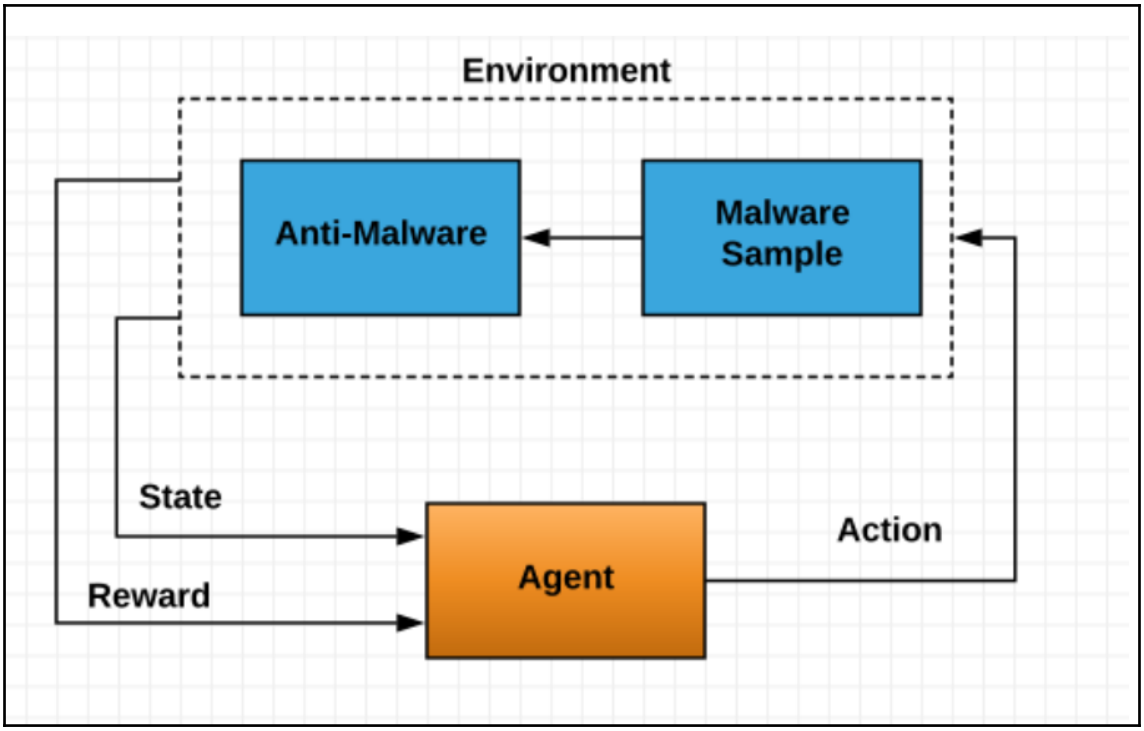

Real Samples
Noise
Generator
Discriminator
Is D Correct?
Fine Tune Training

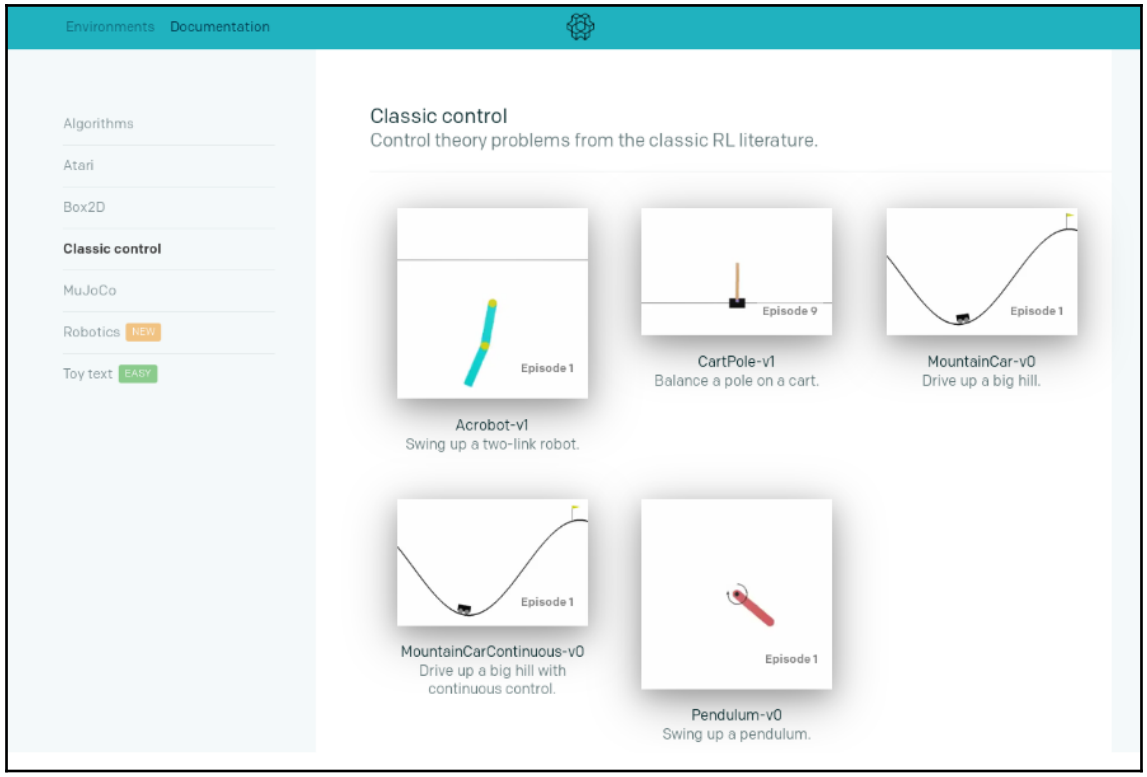



Malware
Noise
Generator
Adversarial Samples
Black-Box Detector
Adversarial Samples and Goodware with Labels
Substitute Detector
Goodware

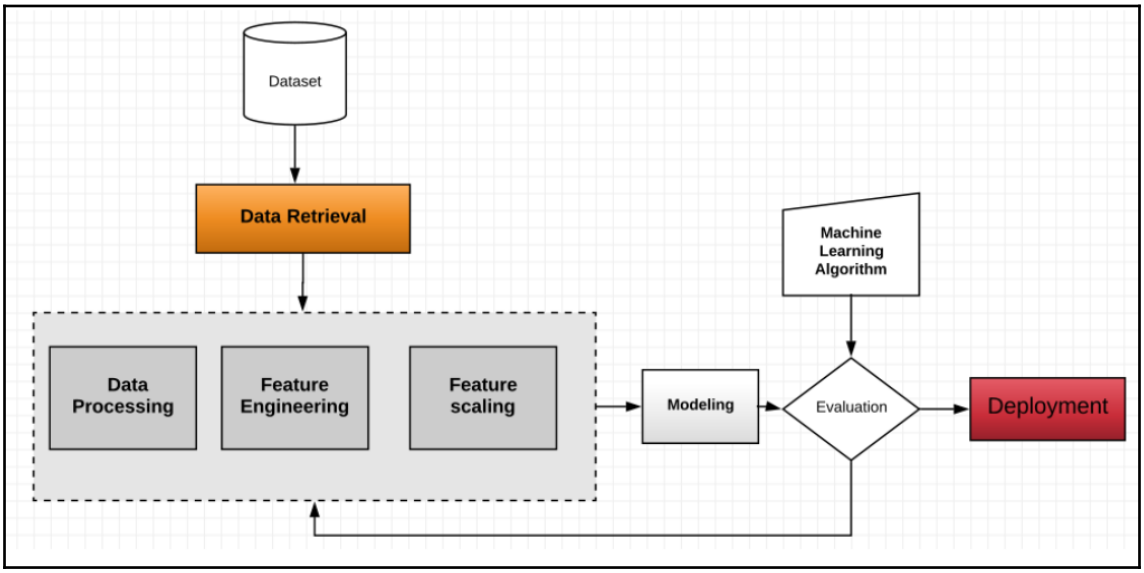

Environment
Reward
State
Action
Agent

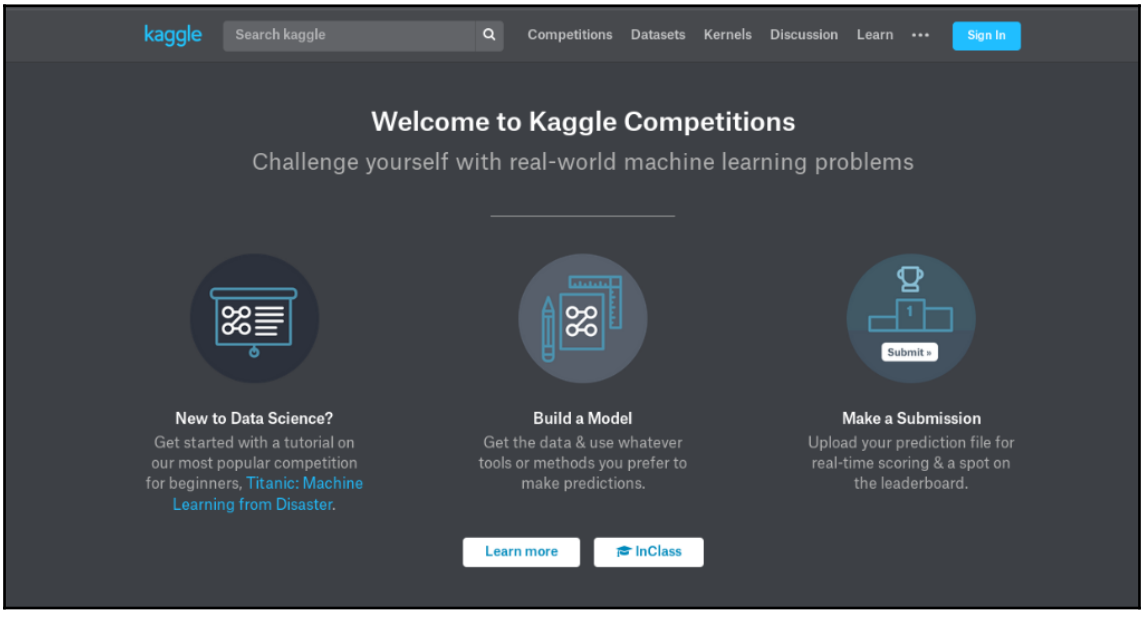

Environment
Scoring Mechanism
Atari Breakout
State
Reward
Agent
Action

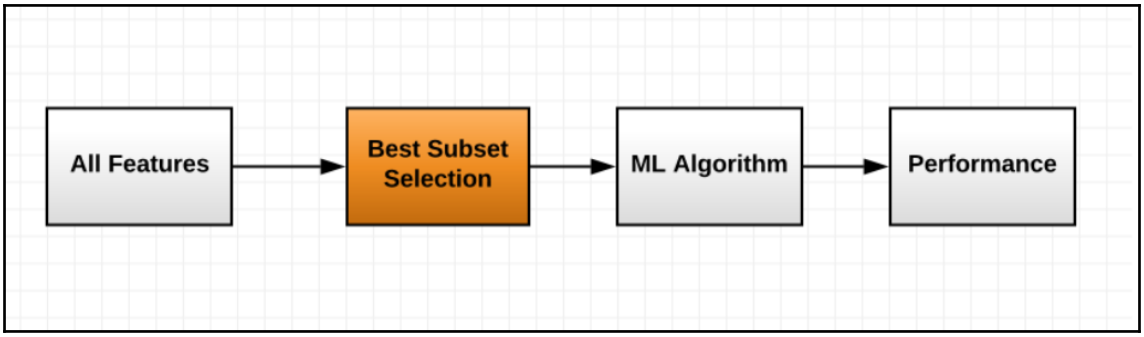

Environment
Anti-Malware
Malware Sample
State
Reward
Agent
Action


Environments
Documentation
Algorithms
Atari
Box2D
Classic control
MuJoCo
Robotics NEW
Toy text EASY
Classic control
Control theory problems from the classic RL literature.
Episode 1
Acrobot-v1
Swing up a two-link robot.
Episode 9
CartPole-v1
Balance a pole on a cart.
Episode 1
MountainCar-v0
Drive up a big hill.
Episode 1
MountainCarContinuous-v0
Drive up a big hill with continuous control.
Episode 1
Pendulum-v0
Swing up a pendulum.

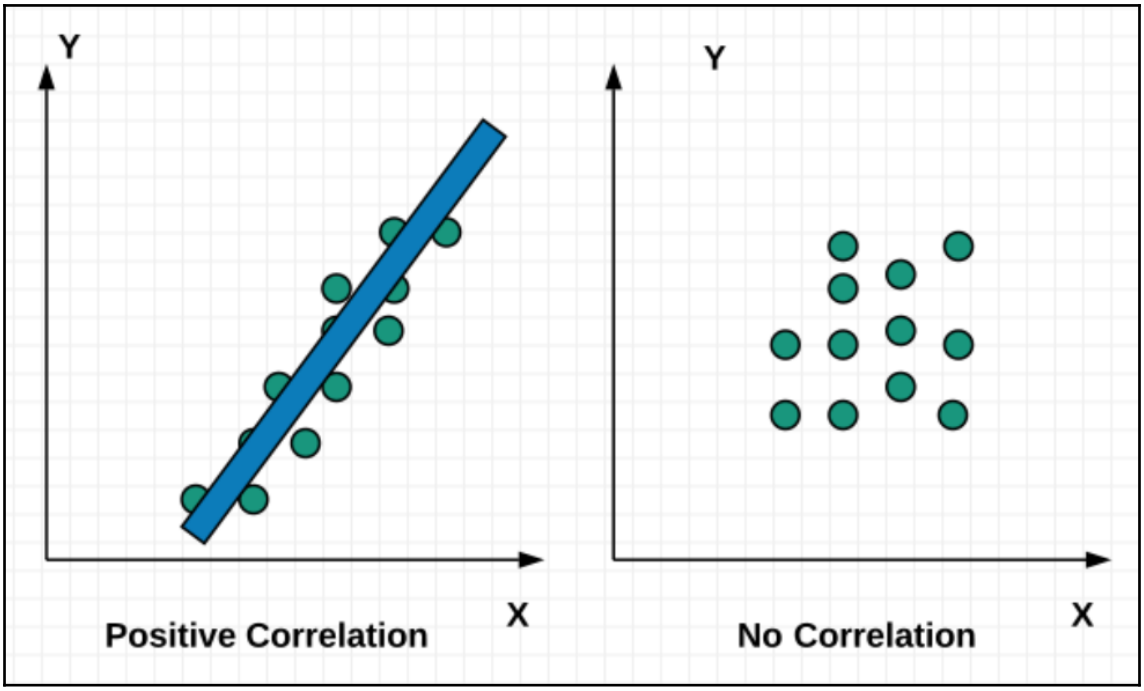
Chapter 10: Best Practices for Machine Learning and Feature Engineering

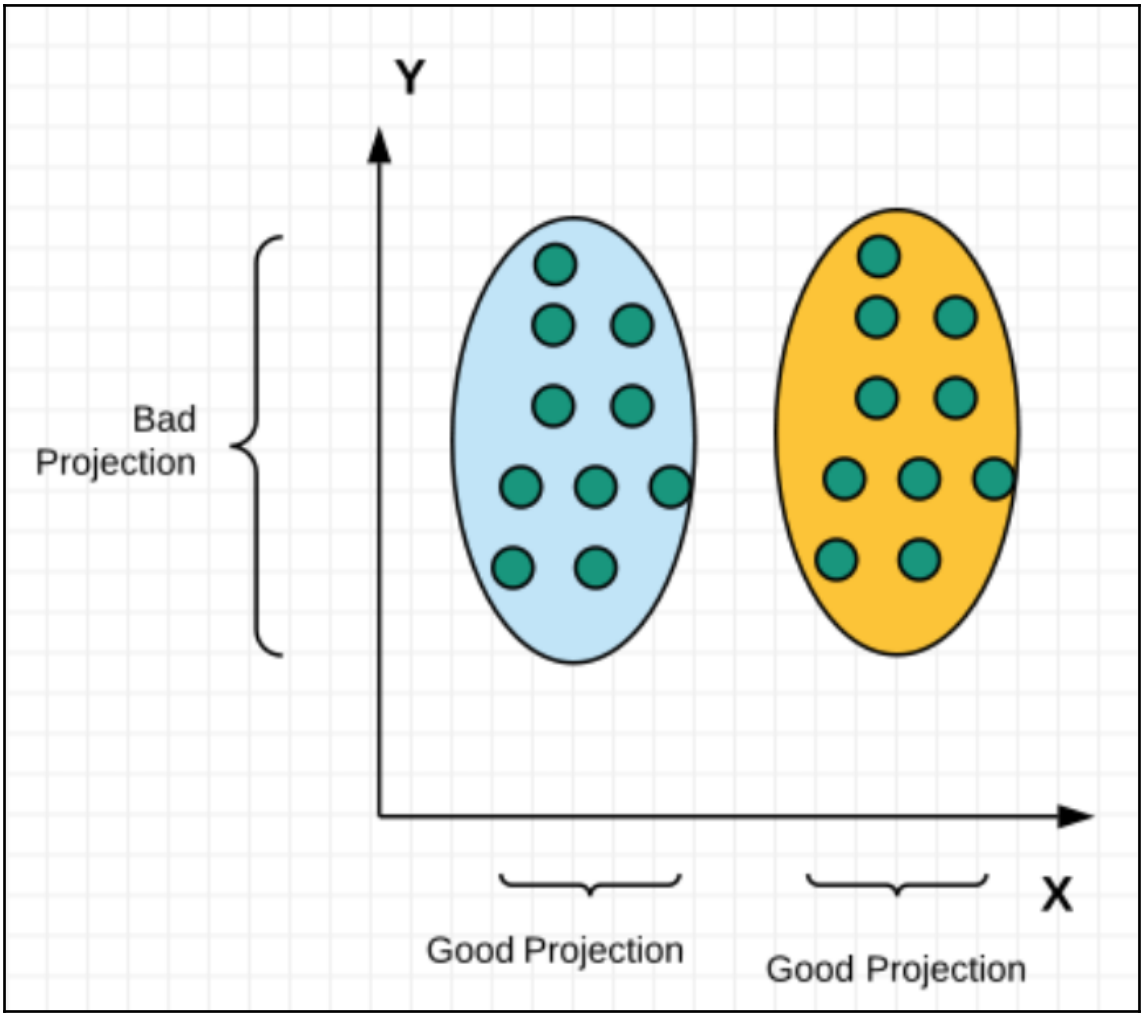




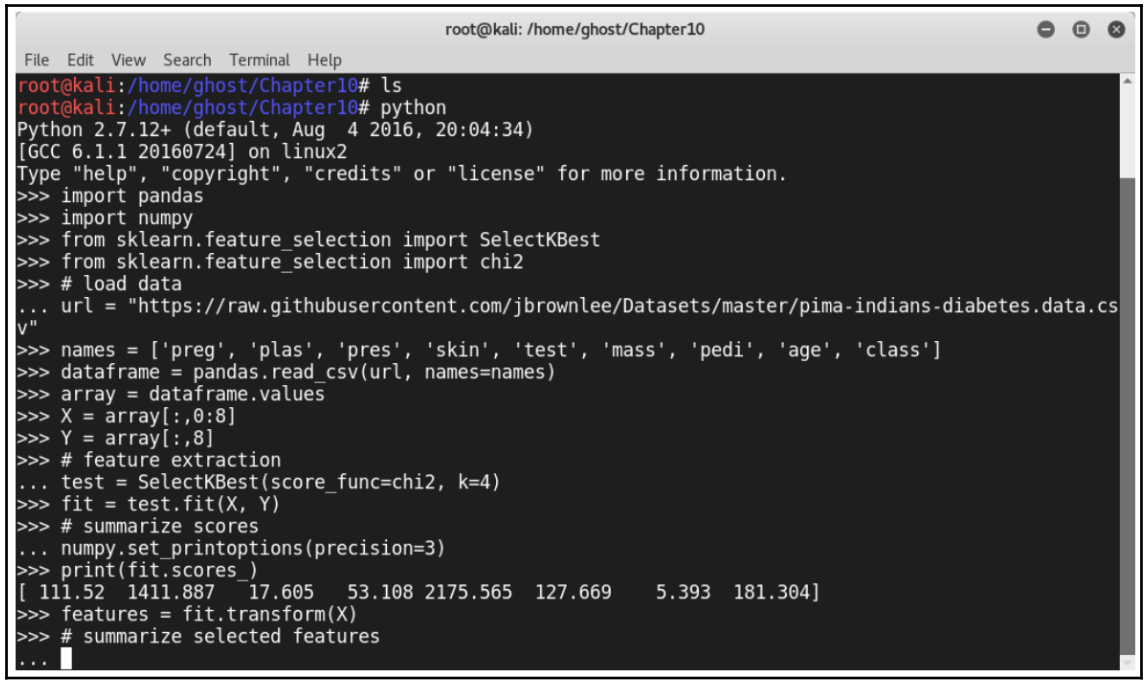


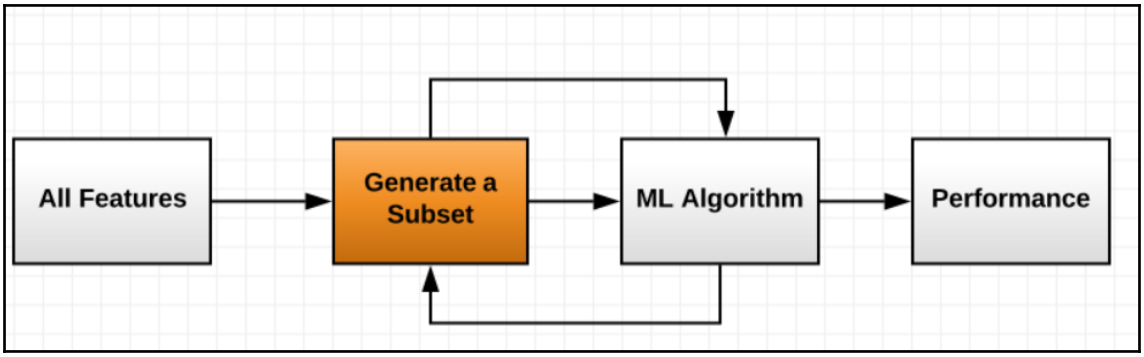
$$P_{x,y} = \frac{Cov(X,Y)}{\mathrm{d}x\mathrm{d}y}$$

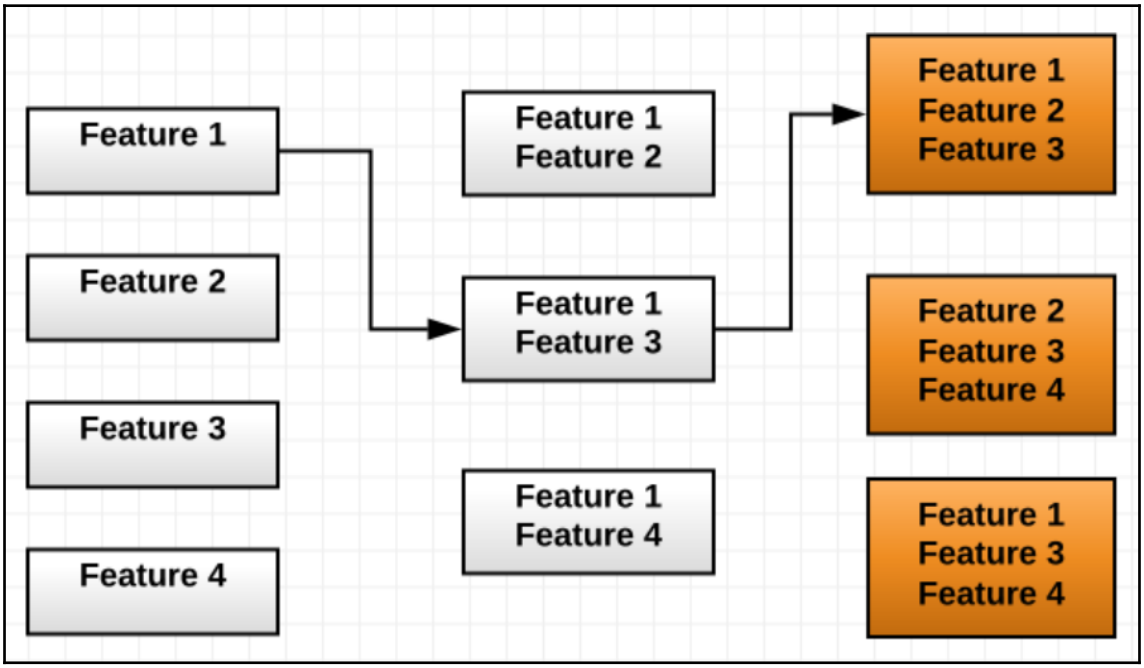

Y
X
Positive Correlation
Y
X
No Correlation

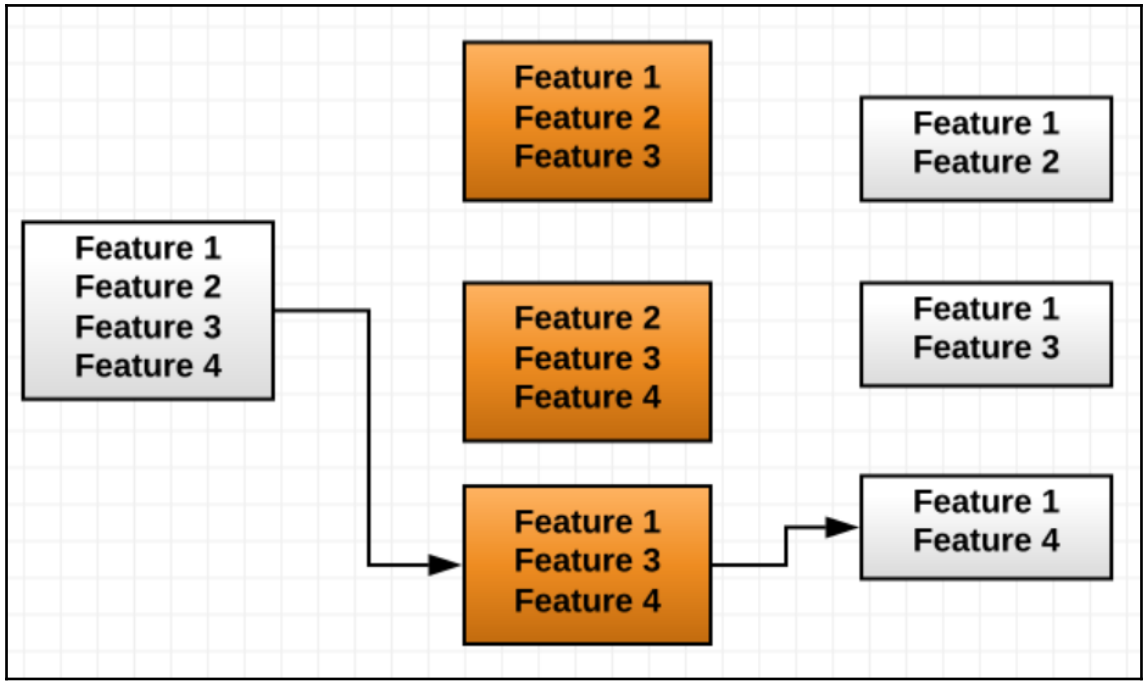

Y
Bad
Projection
X
Good Projection
Good Projection

$$X_c^2 = \frac{(O_i - E_i)^2}{E_i}$$

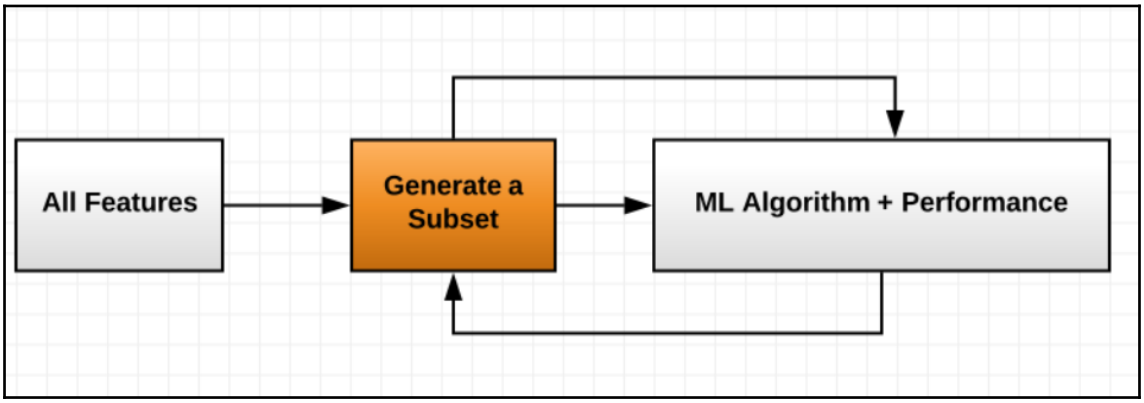


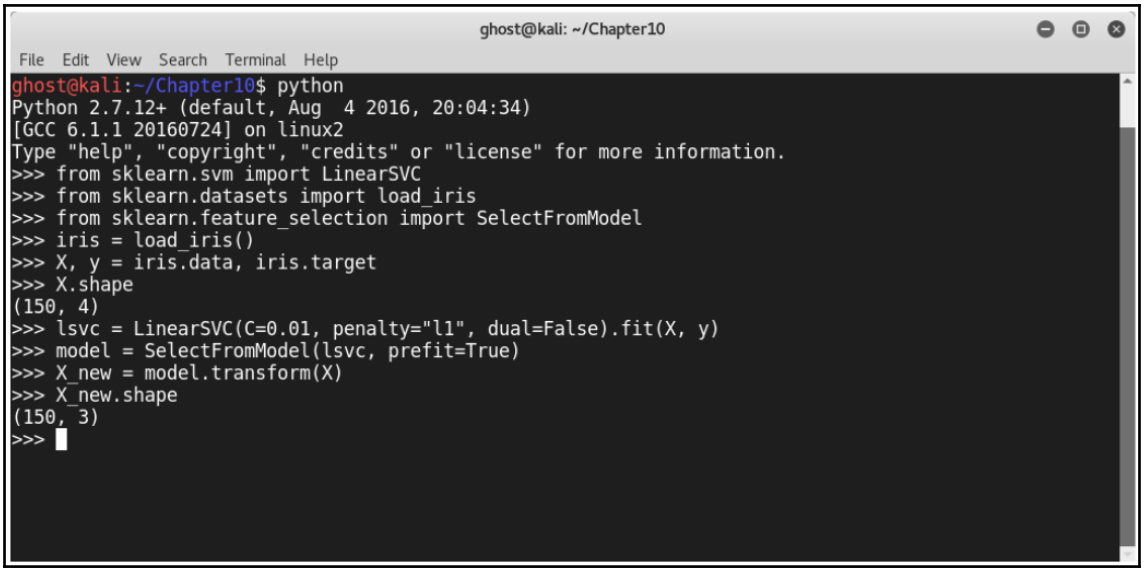

All Features
Generate a Subset
ML Algorithm
Performance

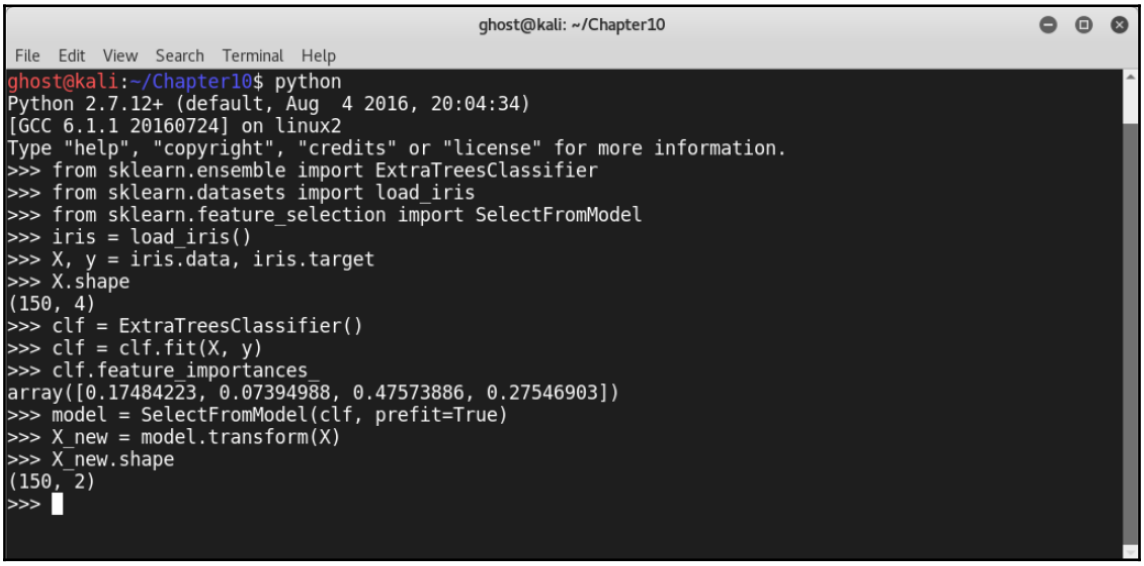

Feature 1
Feature 2
Feature 3
Feature 4
Feature 1
Feature 2
Feature 1
Feature 3
Feature 1
Feature 4
Feature 1
Feature 2
Feature 3
Feature 2
Feature 3
Feature 4
Feature 1
Feature 3
Feature 4


Feature 1
Feature 2
Feature 3
Feature 1
Feature 2
Feature 1
Feature 2
Feature 3
Feature 4
Feature 2
Feature 3
Feature 4
Feature 1
Feature 3
Feature 1
Feature 3
Feature 1
Feature 3
Feature 4
Feature 1
Feature 4

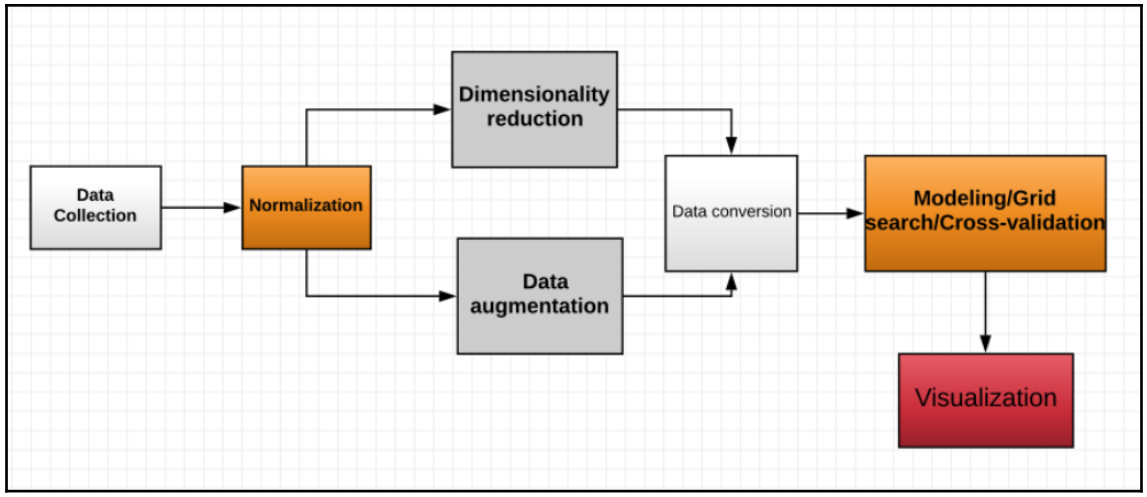
```
root@kali: /home/ghost/Chapter10
File  Edit  View  Search  Terminal  Help
>>> from pandas import read_csv
>>> from sklearn.feature_selection import RFE
>>> from sklearn.linear_model import LogisticRegression
>>> # load data
... url = "https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-diabetes.data.cs
v"
>>> names = ['preg', 'plas', 'pres', 'skin', 'test', 'mass', 'pedi', 'age', 'class']
>>> dataframe = read_csv(url, names=names)
>>> array = dataframe.values
>>> X = array[:,0:8]
>>> Y = array[:,8]
>>> # feature extraction
... model = LogisticRegression()
>>> rfe = RFE(model, 3)
>>> fit = rfe.fit(X, Y)
>>> print("Num Features: %d") % fit.n_features_
Num Features: 3
>>> print("Selected Features: %s") % fit.support_
Selected Features: [ True False False False False  True  True False]
>>> print("Feature Ranking: %s") % fit.ranking_
Feature Ranking: [1 2 3 5 6 1 1 4]
>>>
```




ghost@kali: ~/Chapter10
File Edit View Search Terminal Help
ghost@kali:~/Chapter10$ python
Python 2.7.12+ (default, Aug 4 2016, 20:04:34)
[GCC 6.1.1 20160724] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> from sklearn.svm import LinearSVC
>>> from sklearn.datasets import load_iris
>>> from sklearn.feature_selection import SelectFromModel
>>> iris = load_iris()
>>> X, y = iris.data, iris.target
>>> X.shape
(150, 4)
>>> lsvc = LinearSVC(C=0.01, penalty="l1", dual=False).fit(X, y)
>>> model = SelectFromModel(lsvc, prefit=True)
>>> X_new = model.transform(X)
>>> X_new.shape
(150, 3)
>>>


ghost@kali: ~/Chapter10
File Edit View Search Terminal Help
ghost@kali:~/Chapter10$ python
Python 2.7.12+ (default, Aug 4 2016, 20:04:34)
[GCC 6.1.1 20160724] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> from sklearn.ensemble import ExtraTreesClassifier
>>> from sklearn.datasets import load_iris
>>> from sklearn.feature_selection import SelectFromModel
>>> iris = load_iris()
>>> X, y = iris.data, iris.target
>>> X.shape
(150, 4)
>>> clf = ExtraTreesClassifier()
>>> clf = clf.fit(X, y)
>>> clf.feature_importances_
array([0.17484223, 0.07394988, 0.47573886, 0.27546903])
>>> model = SelectFromModel(clf, prefit=True)
>>> X_new = model.transform(X)
>>> X_new.shape
(150, 2)
>>>


jupyter
Install
About Us
Community
Documentation
NBViewer
Widgets
Blog
jupyter
Project Jupyter exists to develop open-source software, open-standards, and services for interactive computing across dozens of programming languages.


Data Collection
Normalization
Dimensionality reduction
Data augmentation
Data conversion
Modeling/Grid search/Cross-validation
Visualization