

Configuration Manual:

Introduction:

This thesis aims at developing a software that detects an intrusion happening in the system using Machine Learning algorithms. The configuration manual will consist of all hardware and software needed to develop this project. The manual will explain the entire functionalities needed for installing the necessary configurations required to run the code.

Hardware Configurations:

- 4 GB RAM
- 10 GB HDD
- Intel 1.66 GHz Processor Pentium 4

Software Configurations:

- Jupyter Notebook
- Windows 7
- Python 3.6

Libraries used were:

Pandas (Fumo, 2017):

It is a machine learning library used in python. Pandas is an open source library. Pandas is usually used for data analytics and dataset loading. Pandas is being used in a variety of field for machine learning like finance, economics etc. It is highly user friendly and can represent dataset into tabular format for better understanding.

Numpy (Anon., n.d.):

Numpy is a machine learning library used in python. Numpy is used for dealing arrays in python. All calculation making use of 1-d or 2-d arrays are done with numpy. Numpy also consist of functions that can operate with linear algebra and fourier transform.

Sklearn (JAIN, 2015): One of the most important library for machine learning in python is Sklearn. Sklearn consist of lots of tools that can perform statistical modelling and perform classification, regression, clustering and dimensionality reduction.

Matplotlib (Anon., 2018): Matplotlib is a data visualization library. It is a python open source library used for plotting graphs from the results obtained from the model. These plots can help to better understand the scenario of the results. Different aspects of the results can be graphically formatted for better understanding.

Dataset:

The dataset used for this thesis is obtained from <https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/>. The dataset is capable of predicting nine different types of attacks. The dataset was created at Cyber Range Lab of the Australian Centre for Cyber Security (ACCS) using IXIA PerfectStorm tool for generating a hybrid of real modern normal activities and synthetic contemporary attack behaviours.

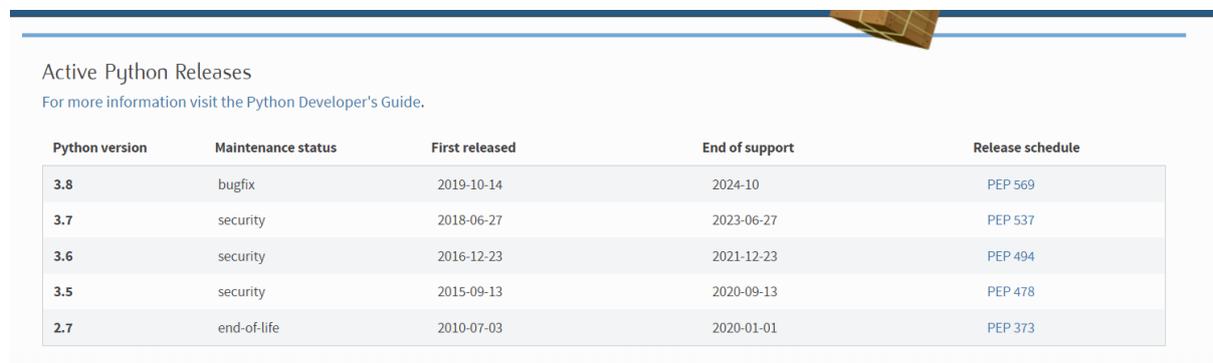
Functionalities:

This section will provide a complete guide of all the steps required to install necessary tools and software needed for the development.

Python:

It is an interpreted high level object oriented programming language. Python has open source machine learning libraries and tools that makes use of machine learning more efficient in python (Anon., 2020).

Installation of Python version 2.7 can be done from official website of python '<https://www.python.org/downloads/>'.



The screenshot shows the 'Active Python Releases' section of the Python website. It includes a link to the Python Developer's Guide and a table with the following data:

Python version	Maintenance status	First released	End of support	Release schedule
3.8	bugfix	2019-10-14	2024-10	PEP 569
3.7	security	2018-06-27	2023-06-27	PEP 537
3.6	security	2016-12-23	2021-12-23	PEP 494
3.5	security	2015-09-13	2020-09-13	PEP 478
2.7	end-of-life	2010-07-03	2020-01-01	PEP 373

All the imports used for developing this project:

```
4 import numpy as np
5 import pandas as pd
6 from sklearn import preprocessing
7 import xgboost as xgb
8 from sklearn import svm
9 from sklearn.metrics import classification_report, accuracy_score, roc_auc_score, average_precision_score, confusion_matrix
10 import sklearn
11 from sklearn.model_selection import train_test_split
12 import matplotlib.pyplot as plt
13 import seaborn as sns
14 from yellowbrick.classifier import ClassificationReport
15 from sklearn.ensemble import AdaBoostClassifier, RandomForestClassifier
16 from sklearn.ensemble import VotingClassifier
17 from mlxtend.classifier import StackingClassifier
18 from sklearn.neighbors import KNeighborsClassifier
19 from sklearn.linear_model import LogisticRegression
20 from sklearn.svm import SVC
21 from sklearn.model_selection import GridSearchCV
22 from mlxtend.feature_selection import SequentialFeatureSelector as SFS
23 from sklearn.externals import joblib
24 from sklearn.ensemble import ExtraTreesClassifier
25 from sklearn.datasets import load_iris
26 from sklearn.feature_selection import SelectFromModel
27
```

Installation of Jupyter Notebook:

Jupyter Notebook is a web-based application. It is a web application that supports Python-based code. It is very flexible: Which means it has the ability to configure, arrange the UI for supporting a very wide range of workflows in machine learning and data science field.

Jupyter Notebook can be installed from '<https://jupyter.org/install>'.

Installation

JupyterLab can be installed using `conda` or `pip`. For more detailed instructions, consult the [installation guide](#).

conda

If you use `conda`, you can install it with:

```
conda install -c conda-forge jupyterlab
```

pip

If you use `pip`, you can install it with:

```
pip install jupyterlab
```

If installing using `pip install --user`, you must add the user-level `bin` directory to your `PATH` environment variable in order to launch `jupyter lab`.

References

Anon., 2018. *Python | Introduction to Matplotlib*. [Online]

Available at: <https://www.geeksforgeeks.org/python-introduction-matplotlib/>

Anon., 2020. Python. Issue <https://www.python.org/doc/essays/blurb/>.

Anon., n.d. *NumPy Introduction*. [Online]

Available at: https://www.w3schools.com/python/numpy_intro.asp

Fumo, D., 2017. *Pandas Library in a Nutshell — Intro To Machine Learning*. [Online]

Available at: <https://medium.com/simple-ai/pandas-library-in-a-nutshell-intro-to-machine-learning-3-acbd39ec5c9c>

JAIN, K., 2015. *Analytics Vidya*. [Online]

Available at: <https://www.analyticsvidhya.com/blog/2015/01/scikit-learn-python-machine-learning-tool/>