Machine Learning , Analytics & Cyber Security the Next Level Threat Analytics

Presented BY: Manjunath N V

yoda@security-exploits.com

LinkedIn: https://www.linkedin.com/in/manjunath-kumar-1a222a18

Who am I – Manjunath N V

- Education
  - B.E. in Computer Science (Bangalore University)
  - Post Graduation in S.O.C (university of Edinburgh, Glasgow)
  - Post Graduation in Digital Network Communication(London Metropolitan university)
- Professional Experience
  - 15+ Years of Consulting and Training experience in Programming, Networking, testing & Security.
  - Hold 20+ Active Professional Certifications in SECURITY
  - Have Trained 3000+ People in Last 8 Years in IT Security
- WARNING – **MAD** about Security (can talk Hours on the Subject)

Topics Covered

- Theoretical Nature of
  - Definitions
  - Importance of these Technologies
  - Where to Find More resources
- Hands On Materials
  - Lab Setup
  - Basic Demonstration
- Guidance to Projects
  - Market Demand for Technologies
  - Project IDEAS

Motivation

- JOBS

  - Live online Openings

Final OUTPUT (DEMO)

- A working Docker IMAGE
  - With Python Library installed
  - Saved as LOCAL DOCKER IMAGE
  - Seems Very simple but need to understand MANY Concepts such as
    - Virtualisation
    - Containers
    - Devops

(WILL SHARE LOT OF self study LAB books for the MOTIVATED)

Definitions

- **Machine Learning (ML)**

Google’s definition - Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data.

Definitions (Contd.)

- **Data Analytics**

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making.

Definitions (Contd.)

- **Cyber Security**

Cybersecurity is the body of technologies, processes and practices designed to protect networks, computers, programs and data from attack, damage or unauthorized access. In a computing context, **security** includes both **cybersecurity**and physical **security**.

Definitions (Finally.)

- **ML + Data Analytics + Cyber Security**

Machine learning has been quickly adopted in cybersecurity for its potential to automate the detection and prevention of attacks, particularly for next-generation antivirus (NGAV) products. ML models in NGAV have fundamental advantages compared to traditional AV, including the higher likelihood of identifying novel, zero-day attacks and targeted malware, an increased difficulty of evasion, and continued efficacy during prolonged offline periods

Implementation Technologies - 1

- Virtualisation
  - Type 1
    - ESXi,KVM
  - Type 2
    - Vmware Workstation,VirtualBOX
- Containers
  - DOCKER
  - LXC
  - KUBERNETES

Implementation Technologies - 1

Devops implementation

DEVOPS LIFECYCLE

DEVOPS – Technology Mapping

What is Docker? (the Silver Bullet)

- Docker is a ‘container technology’
  - Linux-specific
    - can’t run Mac OSX, Windows *in* docker containers
    - But *can* run docker containers *on* Mac OSX & Windows
  - Shrink-wrap your software, run it on any Linux platform

- *Not* a virtual machine
  - Similar to virtual machines, but more lightweight
    - Smaller, faster to start, easier to maintain and manage
    - Lighter on system resources => vastly more scalable
  - VM-thinking will lead to poor results, avoid it!

why use Docker?

- Portability:
  - No need to rebuild your application for a new platform!
    - Build a container once, run it anywhere
      - Cori/Edison/Genepool/…
      - AWS/GCP/…
    - Stable s/w versions across all platforms, no runtime glitches
  - Think of it as ‘modules-to-go’
    - Instead of ‘module load PQR’ you ‘docker pull PQR’
    - No waiting for modules to be built/deployed for you!

- Reproducibility:
  - Because your s/w is stable, your pipeline is reproducible
    - Run the exact same binaries again 10 years from now

What can you do with it?

- Computational workloads
  - Use applications without having to install them
  - Run your applications anywhere; clouds
  - **Reproducible pipelines** – today's focus

- Services
  - Web portals/gateways (**R/Shiny**, Apache, Jupyter…)
  - Persistent workflow manager interfaces (Fireworks…)
  - Continuous build systems (**Gitlab**…)
  - For prototyping or for production running (databases etc)
  - All those things you run in the background on the login nodes today!

Building a container: the Dockerfile

- A recipe for building a container
- Start with a base image, add software layer by layer
  - Choosing the base image has a big effect on how large your container will be: go small (‘alpine’ or ‘busybox’)!
- Add metadata describing the container
  - Always a good idea
- Set the command to run when starting the container, map network ports, set environment variables
  - Not strictly needed for batch applications, useful for services (web apps, databases…)

Some Examples

- SPAM detection
  - Distinguish between SPAM and legitimate email
  - % of emails correctly classified
  - Hand-labeled emails
- Detecting catalog duplicates
  - Distinguish between duplicate and non-duplicate catalog entries
  - False positive/negative rate based on business criteria
  - *H*and-labeled duplicates and non-duplicates
- Go learner
  - *P*laying Go
  - % of games won in tournament
  - Practice games against itself

Programming Language - Why python?

- So many tools
  - Preprocessing, analysis, statistics, machine learning, natural language processing, network analysis, visualization, scalability
- Community support
- “Easy” language to learn
- Both a scripting and production-ready language

External libraries

A very complete list can be found at PyPi the Python Package Index:
https://pypi.python.org/pypi

To install, use pip, which comes with Python:

`pip install` *`package`*

or download, unzip, and run the installer directly from the directory:

```
python setup.py install
```

If you have Python 2 and Python 3 installed, use pip3 (though not with Anaconda) or make sure the right version is first in your PATH.