MTIS 7102: Machine Learning and Neural Networks in Intelligence operations

APPLICATIONS CHALLENGE

The Nairobi Upper Hill Cyber Heist Unveiled

Background Story

In the heart of a bustling Nairobi Upper Hill, where skyscrapers reach for the heavens and neon lights dance through the night, lies the headquarters of CyberSec Solutions, a renowned cybersecurity firm known for its cutting-edge technologies and top-notch expertise.

Amidst the hum of computers and the click-clack of keyboards, a team of elite Intelligence analysts, including yourself, are at the forefront of a high-stakes cyber investigation. It all began with a seemingly innocuous email—a routine correspondence that harbored a sinister payload.

The email, disguised as a routine update from a trusted vendor, contained a hidden malware payload that silently infiltrated CyberSec Solutions' network defenses. Unbeknownst to the team, a sophisticated cybercriminal syndicate had launched a meticulously orchestrated attack aimed at stealing sensitive data and wreaking havoc on the firm's infrastructure which hosts the government Ecitizen Servers.

As the malware propagated through the network like a silent plague, alarms blared and warning lights flashed, Ecitizen servers down, signalling the onset of a digital siege. Panic ensued as the intelligence analysts scrambled to contain the breach and assess the extent of the damage.

In the wake of the cyber assault, it became clear that traditional methods alone would not suffice to thwart the attackers. The team needed to harness the power of cutting-edge technology—machine learning—to combat this elusive adversary.

Thus, the stage was set for your pivotal role in the investigation. Armed with your expertise in machine learning and data analysis, you embarked on a journey into the digital abyss, determined to unravel the intricacies of the cyber heist and bring the perpetrators to justice.

From dissecting malware strains with VirusTotal AI to conducting facial recognition analysis of CCTV video footage, each step of the investigation presented a new challenge that tested the limits of your skills and ingenuity. With each breakthrough, you edged closer to uncovering the truth behind the cyber assault and thwarting the nefarious plans of the cybercriminals.

As the clock ticked ominously and the pressure mounted, you remained undeterred, driven by a relentless pursuit of national security and a commitment to safeguarding the digital frontier. The fate of CyberSec Solutions—and the

countless individuals whose security hung in the balance—rested in your capable hands.

Now, armed with your expertise and a suite of cutting-edge machine learning tools, you stand ready to confront the forces of darkness and emerge victorious in the ultimate battle for national and cybersecurity supremacy.

Challenge 1

(Folder 1)

Natural Language Processing:

Investigation into CyberSec Solutions cyber-attack suspects computing devices recovered 4461 Files but have no time to go through each file checking for files containing context materials. Design NLP script that will parse the content of the files and extract files bearing keys words 'MPESA and HAWALA, KAFIR, mpesa, kafir, hawala 'words. The script should extract the identified files to a directory known as 'evidence and create a csv report of the files identified bearing the keyword'.

Challenge 2

(Folder 1)

File Format Classification:

To streamline the analysis process, create a classifying model to automatically detect and categorize files based on their respective formats.

Design a model to categorize all the files in the folder as either pdf, docx, jpeg, html etc. You are provided with file signatures (filetypes.xlsx) dataset of all possible identified files. You may use the dataset to help you in training the model.

Challenge 3

(Folder 2)

Malware Detection with Virus Total AI:

Following RAM analysis, investigators extracted **1456 files** from RAM Dump. Using Virus total machine learning API. Scan extracted files and generate the output of the findings in CSV format. Detailing all the files identified as malware. Use identified files to train malware detection ML model which may be used in future.

Challenge 4

(Folder 3)

Detecting File Manipulation:

In Folder 3, the focus is on detecting file manipulation by examining binary file headers and footers. Your aim is to develop ML model to detect potential file manipulation or tampering by flagging anomalies in file structures, enabling investigators to identify unauthorized modifications.

Challenge

Text classification plays a crucial role in reconstructing files with corrupted extensions found in **Folder 3.** After examining all the digital electronic media of suspects in the CyberSec Solutions case, it was discovered that the files in Folder 3 contain corrupted file extensions. The task is to design an ML model capable of reconstructing these files to their original format using provided headers and footer dataset (filetypes.xlsx) or applicable ML algorithim.

Challenge 5

(Folder 4)

Facial Recognition in Video Clips

In Folder 4, the focus shifts to Facial Recognition in Video Clips. The objective entails parsing video clips and extracting known faces. Leveraging state-of-the-art facial recognition algorithms, the task mandates the development of a Machine Learning (ML) model proficient in precisely identifying and extracting specific individuals from video footage.

Challenge:

The sponsors behind the CyberSec Solutions cyber-attack are suspected to be influential individuals featured in a YouTube video showcasing famous world faces (titled: Faces of the World - 10 YEARS OF TRAVEL.mp4). The sponsors' facial photographs are cataloged as sponsor1, sponsor2, sponsor3, and sponsor4. Design an ML model capable of extracting all video frames and accurately identifying and extracting each of these individuals' faces from the video.

Challenge 6

(Folder 5)
Predictive Insights through Machine Learning

In today's data-driven landscape, the power of Machine Learning (ML) to unearth predictive insights has become indispensable. By harnessing advanced algorithms and vast datasets, ML empowers us to extract valuable patterns and make informed decisions across diverse domains.

Challenge:

Folder 5 delves into Prediction tasks. The scenario involves being furnished with the computer browsing history of the primary suspect in the CyberSec Solutions case. The challenge is to construct a classification model aimed at predicting the most frequently visited category by the suspect hence generate other likely insights on the nature of suspect browsing habits. This entails tackling a multi-class classification problem, wherein each website is allocated to one of the predefined categories such as Shopping, Entertainment, social media, etc.

Challenge 7

(Folder 6)
Classification: Network Analysis:
Folder 6 delves into Classification tasks within the realm of Network Analysis. Using the provided dataset containing IP address ranges and their associated regions, design and implement a classification model to predict the region based on an input IP address.

Leverage a dataset containing CyberSec Solutions hosts vulnerabilities and their affiliated regions to develop a machine learning model capable of accurately predicting the region associated with a given IP address hence identify the vulnerability of specified IP.

Challenge 8

(Folder 7)
Steganography Detection:

The file in folder 7 is suspected to contain hidden materials related to the cyber criminals.

You are required to develop a specialized ML model capable of detecting hidden messages or data concealed within this digital image (future.jpg)

YOU WON THE CHALLENGE BADGE

Through your expertise in machine learning and data analysis, you have successfully navigated the intricate landscape of national security threats. Your technical reports and findings provided invaluable insights to the analysis department, empowering them to enhance future investigations and safeguard against emerging threats in the ever-evolving digital landscape. As an AI Technical Intelligence Analyst, you have remained at the forefront of innovation, continuously pushing the boundaries of what's possible in the realm of national security.

You have written a series of ML scripts to automate various analysis tasks, including malware detection, file classification, image extraction, facial recognition, anomaly detection, keyword search in documents, timeline analysis, metadata extraction, browser artifact analysis, and feature extraction from digital assets. You are our armoury.