# FINDINGS

Rune mannaerts

# TABLE OF CONTENTS

# INTRODUCTION

In today's digital age, security threats are becoming increasingly complex and sophisticated. Organizations must have the ability to detect and respond to these threats in real-time to protect their assets, customers, and reputation. However, with the vast amounts of data generated by modern networks, it can be challenging for security teams to quickly identify and respond to security incidents.

This is where the XSOAR platform and machine learning (ML) model come in. XSOAR is a security orchestration, automation, and response (SOAR) platform that allows security teams to automate tasks and workflows, reducing the time and effort required to manage security incidents. The XSOAR ML model is a supervised machine learning algorithm that can be used to classify security events as either benign or malicious. The model is trained on a dataset of labeled security events, where each event is labeled as either benign or malicious. Once the model is trained, it can be used to classify new security events based on their features.

Python scikit is a popular machine learning library that can be used to build and train ML models. Scikit includes modules for data preprocessing, feature selection, model selection, and evaluation, as well as several popular algorithms, such as decision trees, random forests, and support vector machines. By using scikit, data scientists and ML practitioners can build and train ML models to solve a wide range of problems.

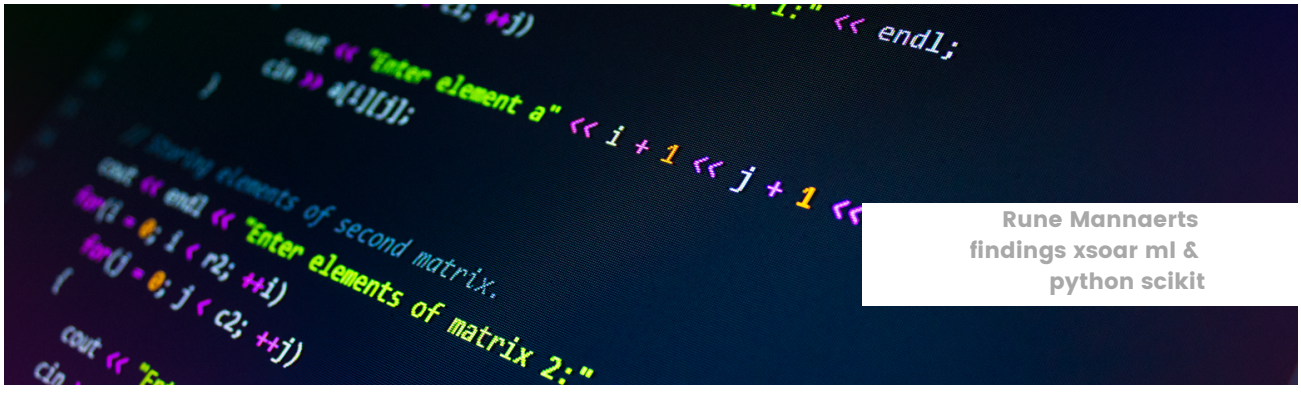
In this document, we will provide an overview of the XSOAR ML model and Python scikit, including how they work, what they are used for, and any benefits or drawbacks they may have. We will also provide examples of how these tools can be used in practice to detect and respond to security threats and solve other real-world problems. By the end of this document, readers should have a solid understanding of these tools and how they can be applied to solve important problems.

# FINDINGS: PYTHON SCIKIT

Based on my experience with Python scikit, I can confidently attest that this library is user-friendly and consistent, providing developers and data scientists with a wide range of powerful features and tools for building robust machine learning models with relative ease. One of its most impressive aspects is its ability to streamline the entire machine learning workflow, offering modules for data preprocessing, feature selection, model selection, and evaluation, which makes it a comprehensive solution for building and training machine learning models. This all-in-one approach saves significant time and effort, allowing developers and data scientists to focus on other aspects of their work.

Moreover, Python scikit's standardized API is consistent across all of its modules and functions, making it easier for users to learn and apply their knowledge to other parts of the library. This consistency makes it easier for developers and data scientists to create, maintain, and scale machine learning applications. However, it is important to keep in mind that the quality of the data used to train the machine learning model is crucial for its effectiveness. A poorly curated dataset can lead to inaccurate results and limit the effectiveness of the model. Therefore, it is essential for developers and data scientists to carefully curate their datasets and ensure that they are representative of the problem they are trying to solve.

In my experience, the quality of the dataset has a significant impact on the effectiveness of the model. A well-curated dataset can result in a highly accurate model, while a poorly curated dataset can result in an ineffective model, regardless of the library or algorithm used. Therefore, while Python scikit is a powerful and reliable library, its true potential can only be realized when working with high-quality, well-curated datasets. Overall, Python scikit is an excellent library for building and training machine learning models. Its user-friendly interface and consistent API make it easy to use, and its comprehensive set of modules and functions streamline the entire machine learning workflow. However, it is important to keep in mind that the effectiveness of any machine learning model ultimately depends on the quality of the data used to train it.

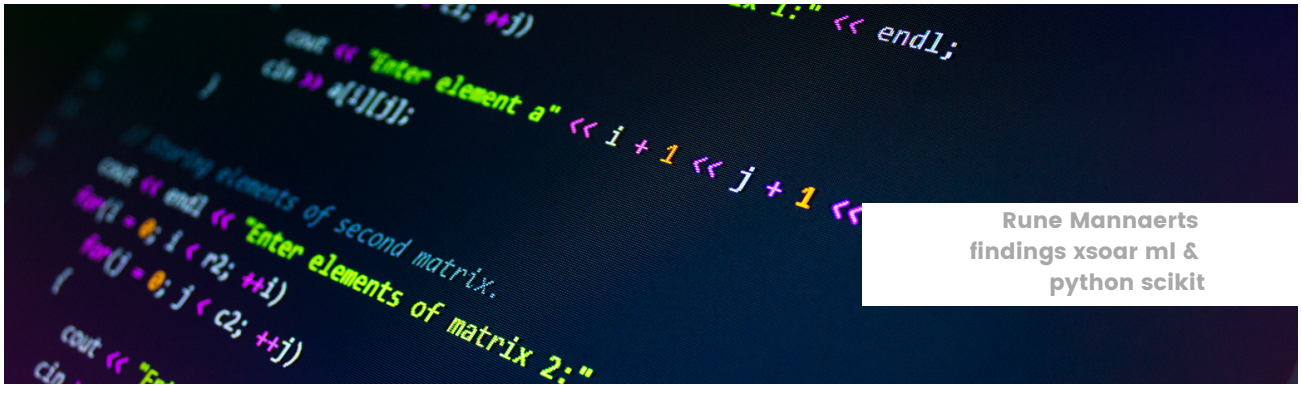


# FINDINGS: PYTHON SCIKIT

youtube video evidence:

ai script: https://youtu.be/DNuRqBbBvKw

dataset searcher: https://youtu.be/APicbdwiJN8



# FINDINGS: XSOAR ML

XSOAR machine learning models are known for their stability and effectiveness, and they have the potential to be very useful in various environments. However, one of the main challenges that I have encountered while working with these models is the issue of having limited data available to train them.

This is particularly true in light of a recent change in classification methods that took place earlier this year. Even if I were to use the previous classification method, there is not enough data available to base a model on, as many of the resolved cases were simply classified as either malicious or non-malicious. This lack of differentiation between the two categories makes it difficult to build effective machine learning models.

As a result, my evaluation of XSOAR machine learning models is primarily based on what I have seen online, rather than on my own personal experience. While I have been able to gather some information and insights from other sources, it is not the same as being able to build and train my own models with sufficient data.

It is worth noting that the lack of available data is not unique to XSOAR machine learning models. This is a common challenge in the field of machine learning, and it highlights the importance of data quality and quantity for effective model training. In the absence of sufficient data, it is difficult to draw meaningful conclusions or make accurate predictions.

Overall, XSOAR machine learning models are an exciting and promising tool for various applications, but the effectiveness of these models is heavily dependent on the quality and quantity of available data. While I have not had the opportunity to work with these models extensively myself, I have researched and learned about their capabilities and limitations, and I believe that they have the potential to be very useful in the right circumstances.

# FINDINGS: XSOAR ML

## New ML Model - Phishing machine learning classifier

Train your DBot ML to analyze emails, by pairing attributes of an incident with a "verdict".
A "verdict" is an arbitrary category, e.g. "Spam", "Malicious", "False positive".
Attributes are values of a field you select, e.g the values of "close reason".

**1. Details and Scope**

Model name*: phishing

Description: Type model description

Incident type*: Cegeka - Microsoft Defender For Endpo...

Date range: Year to date

**2. Select field**

Field values should dictate verdicts

Incident field*: Triage Result

| Field Values |
| --- |
| |

**3. Pair values and verdicts**

Drag one value or more from the left column (Field Values) to the appropriate verdict column on the right.
You can rename or add verdicts to fit your needs.

| Verdict: Malicious | Verdict: Non-Malicious |
| --- | --- |
| suspicious (8) | benign positive (3) |
| true positive (6) | false positive (3) |
| Total: 14* | Total: 6* |

Edit verdicts

* Not enough incidents for a trusted verdict. Modify the date range to return more results.


Rune Mannaerts
findings xsoar ml &
python scikit

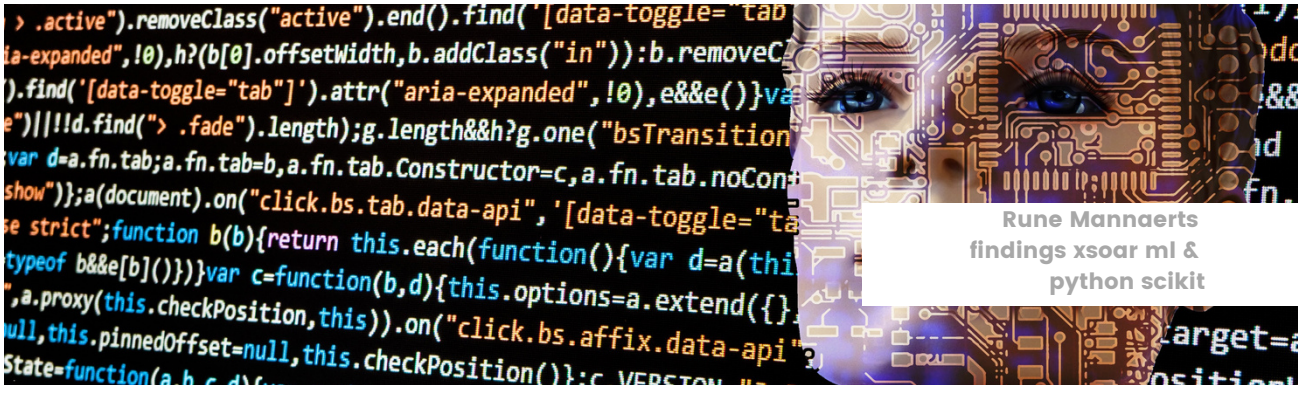
# FINDINGS: OPENAI CHATGPT

As part of my study, I analyzed OpenAI's ChatGPT, a robust language model that has been developed using deep learning techniques. One of the most remarkable features of ChatGPT is its ability to comprehend natural language and provide coherent and contextually relevant responses to a broad range of prompts. This is a significant advantage of the model, as it can be trained to perform specific tasks, making it a versatile tool for various applications.

During my investigation, I developed an integration for ChatGPT based on an existing model. Although the integration was successful, I found that there were certain limitations to the language model and deep learning network. Specifically, I noticed that the translation of prompts to useful data was not always consistent for the same prompt or human perception. This implies that the responses from ChatGPT were somewhat predictable, which can be problematic when trying to obtain accurate and reliable information.

Moreover, the language model can sometimes produce responses that are unexpected or not relevant to the prompt. While this can be useful in some cases, it can also lead to confusion and errors when relying solely on ChatGPT's responses. Therefore, I recommend that ChatGPT should not be solely relied upon, but rather used as a supplementary tool in combination with other tools to strengthen and integrate them together for a more targeted ML model.

In conclusion, I found that ChatGPT is a powerful language model that can be fine-tuned and trained for specific tasks, making it a versatile tool for a variety of applications. However, it has some limitations, including inconsistent translations of prompts to useful data and the potential for unexpected or irrelevant responses. Therefore, I recommend using ChatGPT as a supplementary tool in combination with other tools to strengthen and integrate them together for a more targeted ML model, rather than solely relying on it.



# FINDINGS: OPENAI CHATGPT

The present collection of images serves as illustrative examples derived from a script I developed. The intention behind showcasing these images is to highlight the inherent inconsistencies observed within the utilized API. To ensure the security of sensitive information, certain data points have been deliberately redacted.

**CVE-2020-15783 Detail**

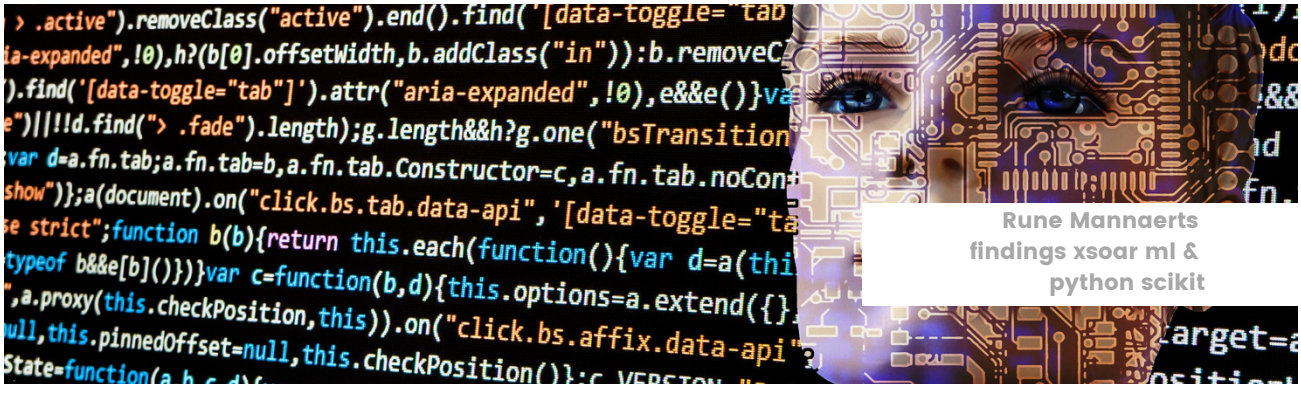
**Description**

A vulnerability has been identified in SIMATIC S7-300 CPU family (incl. related ET200 CPUs and SIPLUS variants) (All versions), SIMATIC TDC CPU555 (All versions), SINUMERIK 840D sl (All versions). Sending multiple specially crafted packets to the affected devices could cause a Denial-of-Service on port 102. A cold restart is required to recover the service.

text

CVE-2020-15783: Microsoft Azure for Windows Elevation of Privilege Vulnerability
CVE-2020-15782: Microsoft Azure for Windows Remote Code Execution Vulnerability
CVE-2020-15781: Microsoft Azure for Windows Denial of Service Vulnerability
CVE-2020-15780: Microsoft Azure for Windows Information Disclosure Vulnerability

the following link is used to determine the above descriped cve has nothing to do with an elevation of privilege vulnerability on the microsoft azure platform:
"https://cert-portal.siemens.com/productcert/pdf/ssa-492828.pdf"



# CONCLUSION

in summary, my analysis of XSOAR machine learning models, Python scikit, and OpenAI ChatGPT has yielded valuable insights into their strengths and limitations.

XSOAR machine learning models have demonstrated stability and effectiveness, but their potential usefulness is heavily dependent on the quality and quantity of available data. Without sufficient data, it is difficult to draw meaningful conclusions or make accurate predictions.

Python scikit is a powerful and user-friendly library that offers comprehensive modules and functions for building and training machine learning models. However, the effectiveness of the model ultimately depends on the quality of the data used to train it. A well-curated dataset is essential to achieve high accuracy.

OpenAI ChatGPT is a remarkable language model that can be fine-tuned and trained for specific tasks, making it a versatile tool for a variety of applications. However, it has certain limitations, including inconsistent translations of prompts to useful data and the potential for unexpected or irrelevant responses.

Taken together, these findings underscore the importance of high-quality and representative data for effective machine learning. Furthermore, while these tools offer powerful solutions for various applications, they are most effective when used in combination with other tools and methodologies to create a more targeted and robust machine learning model.