What Machine Learning Is All About

Written for KAMIND, IT, Inc.

By Ravi Das

Introduction

Daily, we keep hearing about the number of Cyberattacks that are taking place. In fact, every day, major news headlines report a business or a corporation being hit with thousands, or even millions of Personal Identifiable Information (PII) records being maliciously hijacked.

Although Cyberattacks very often occur in mere minutes, if not seconds, the aftermath that it has brought on and the rebuilding that an organization must do can take a very long time.

In fact, this can be so detrimental that some companies never recover from it. The bottom line is that IT Security teams are simply too overburdened and overtaxed to keep up with all of this, let alone having to address future Cyberthreats.

Because of the enormous toll that this can take, many legitimate alerts and warnings often go unnoticed, which is known as Alert Fatigue (as reviewed in a previous blog). Many CIOs and CISOs are scrambling to find new ways to keep up with all of this.

One of the possible solutions to this is through the use of automated tools. One prime example of this is known as "Machine Learning". It can be used to not only relieve the hands of the IT Security staff when it comes to doing the daily, repetitious tasks; but it can also be used to help to model the Cyberthreat landscape into the future in just a matter of minutes. This is a task that would take hours and even days if it were to be attempted by a human being.

The goal of this blog is to provide a brief overview into Machine Learning. It is by no means a comprehensive one, as an entire book can be written on this subject matter.

A Definition of Machine Learning

Machine Learning, also known as "ML", can be specifically defined as follows:

"Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly."

(SOURCE: 1)

To break this down, ML is a subset of a much larger field known as "Artificial Intelligence". One of the primary goals of Machine Learning is to allow a Cybersecurity system to learn from a prearranged set of information and data, without any human assistance. From what it learns, this system is then able to project what the future holds in terms of threat vectors. Another objective of Machine Learning is to help filter out what alerts and warnings are for real, and which are not.

But keep in mind, in a very simplistic view, Machine Learning takes on more a "Garbage In, Garbage Out" methodology. In other words, the system is only as good as the information and data that is fed into it. Thus, in order to keep this system optimized on a real time basis, it must keep receiving and digesting various sorts of data sets 24 X 7 X 365. A cardinal rule of thumb in this instance is that the more intelligence feeds are used, the better, as this will lead the Cybersecurity system to discover a plethora of unhidden trends. As a result, this will make that much more robust in helping to combat Cyberattacks.

How Machine Learning Works

Machine Learning works in two different formats, depending upon the security requirements of the business or corporation:

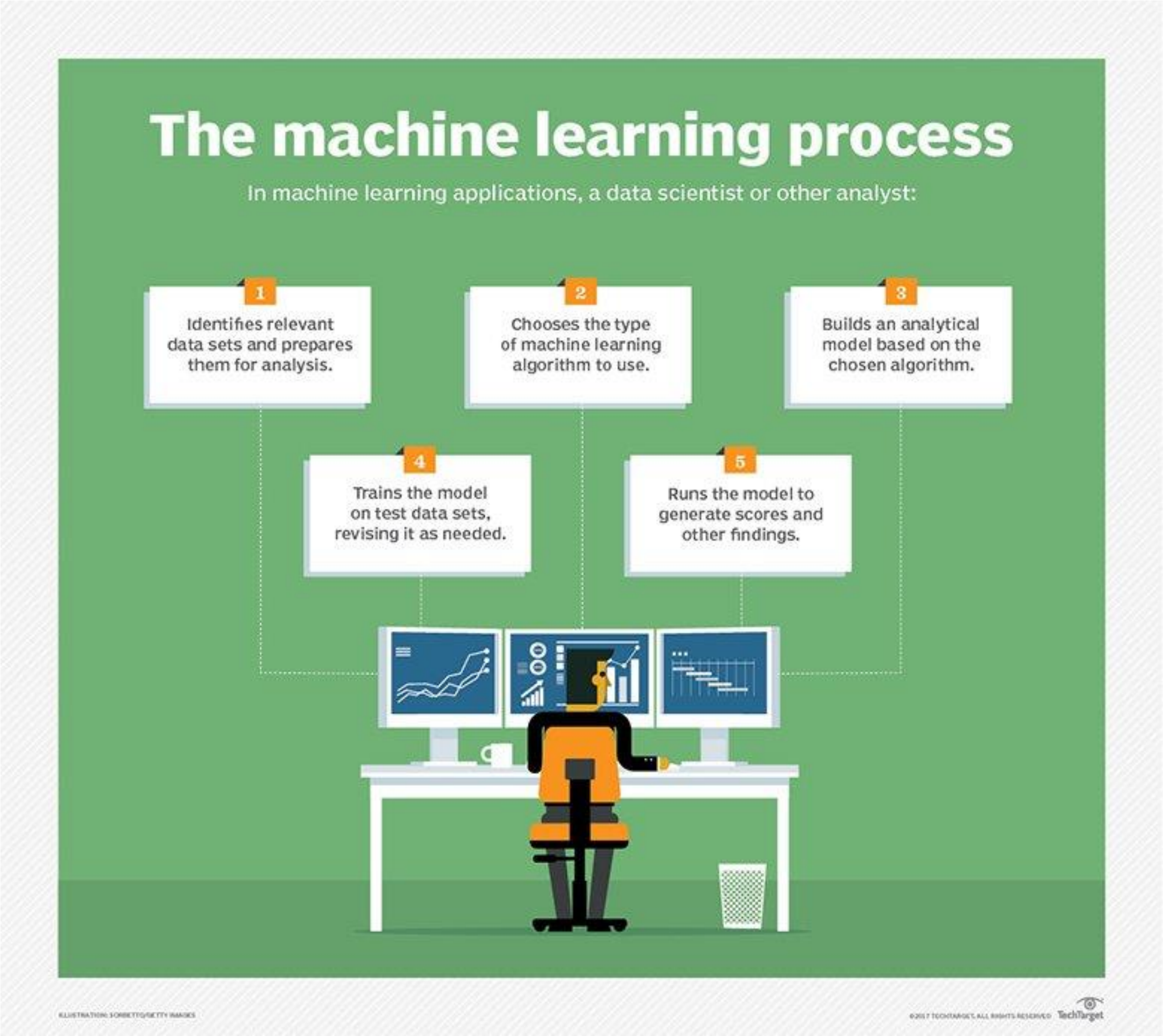
1) <u>A Supervised Approach</u>:

   In this scenario, the Cybersecurity system that used Machine Learning requires human intervention to varying degrees. For example, a highly skilled data analyst is often required in order to select the inputs and feed them. Also, the desired output must be programmed into it as well, so that it can provide the appropriate answers that are desired.

2) <u>An Unsupervised Approach</u>:

   This is where the Cybersecurity system can simply be fed the many intelligence feeds that it requires (and as mentioned, the more of this the better). From there, it will learn the various trends in the datasets, and even discover hidden or hard to notice trends and statistical correlations as well. These kinds of systems do not need to be programmed for what the desired outcome will be. In order to accomplish these sorts of tasks, very sophisticated algorithms known as “Neural Networks” and “Deep Learning” are very often used.

Both approaches are illustrated in the diagram below:



The machine learning process

In machine learning applications, a data scientist or other analyst:

1 Identifies relevant data sets and prepares them for analysis.

2 Chooses the type of machine learning algorithm to use.

3 Builds an analytical model based on the chosen algorithm.

4 Trains the model on test data sets, revising it as needed.

5 Runs the model to generate scores and other findings.

ILLUSTRATION: SORBETTO/GETTY IMAGES

©2017 TECHTARGET. ALL RIGHTS RESERVED TechTarget

(SOURCE: 2).

The Types of Machine Learning Algorithms

Machine Learning makes use of a wide variety of Learning Models (also referred to as “Learning Algorithms”), and some of the widely used ones are as follows:

- <u>Reinforcement Learning</u>:

  This is a kind of algorithm in which the ultimate outcome is to determine which course of action to take. As it relates to Cybersecurity, this could mean whether a malicious file should be completely purged or transferred to an isolated environment so that it can be studied further.

- <u>Feature Learning</u>:
  This is where several algorithms are used in order to examine the datasets from different angles or perspectives when it is fed into the Cybersecurity system.

- Anomaly Detection:

  This is a more specialized kind of algorithm which detects statistical outliers, or other types of anomalous data pieces that lie in the overall dataset. From here, they can be put together in order to discover hard to notice trends. This is very useful for Cybersecurity, as it can be used to help detect potential Cyberthreats that are covert in nature.

- Association Learning:

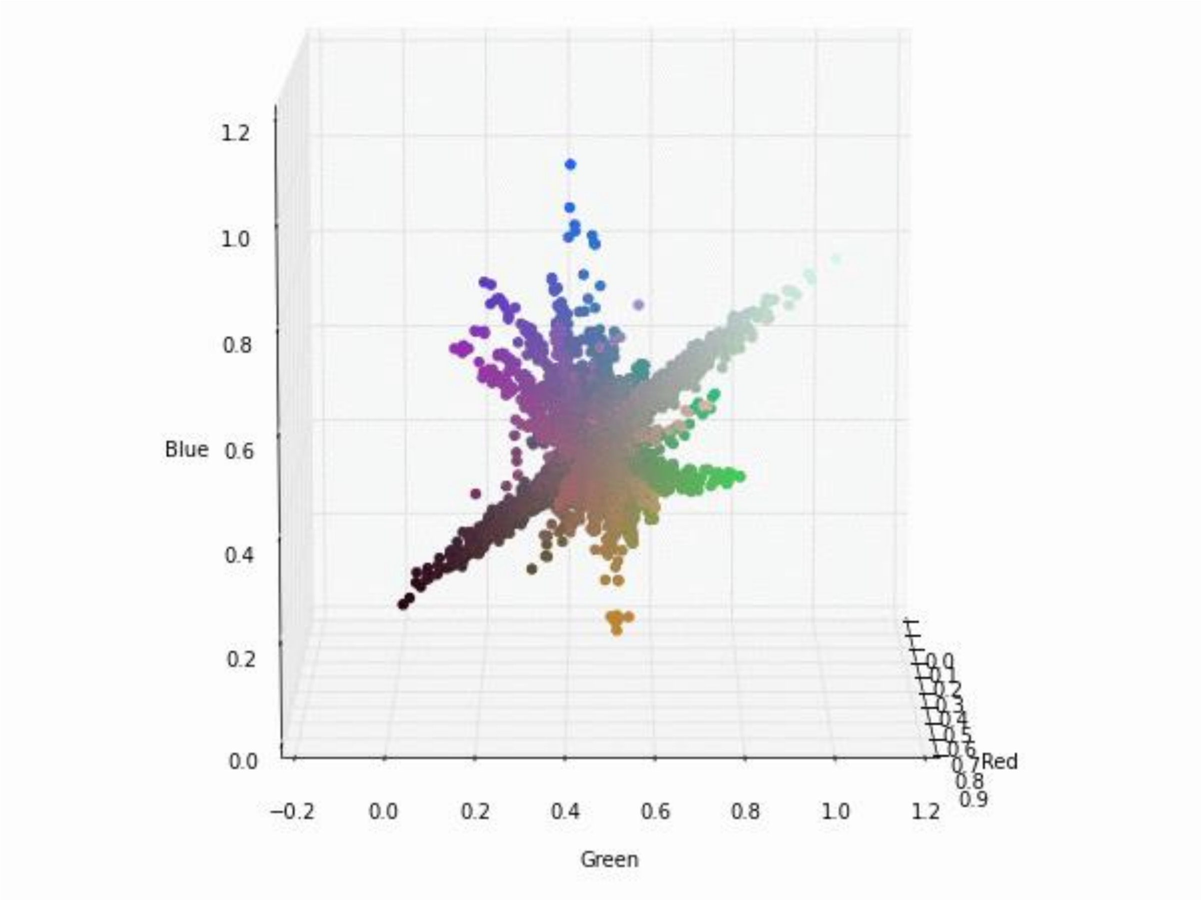
  In this instance, various rules are preestablished so that the Cybersecurity system can discover any statistical correlations from the datasets that are being fed into it. From here, various hypothesis can then be formulated about a potential threat vector, and it can be very useful when conducting Threat Hunting exercises.

The Machine Learning Models

It is the summation of the different algorithms that make up a Machine Learning Model. Some of these include the following:

- Artificial Neural Networks:

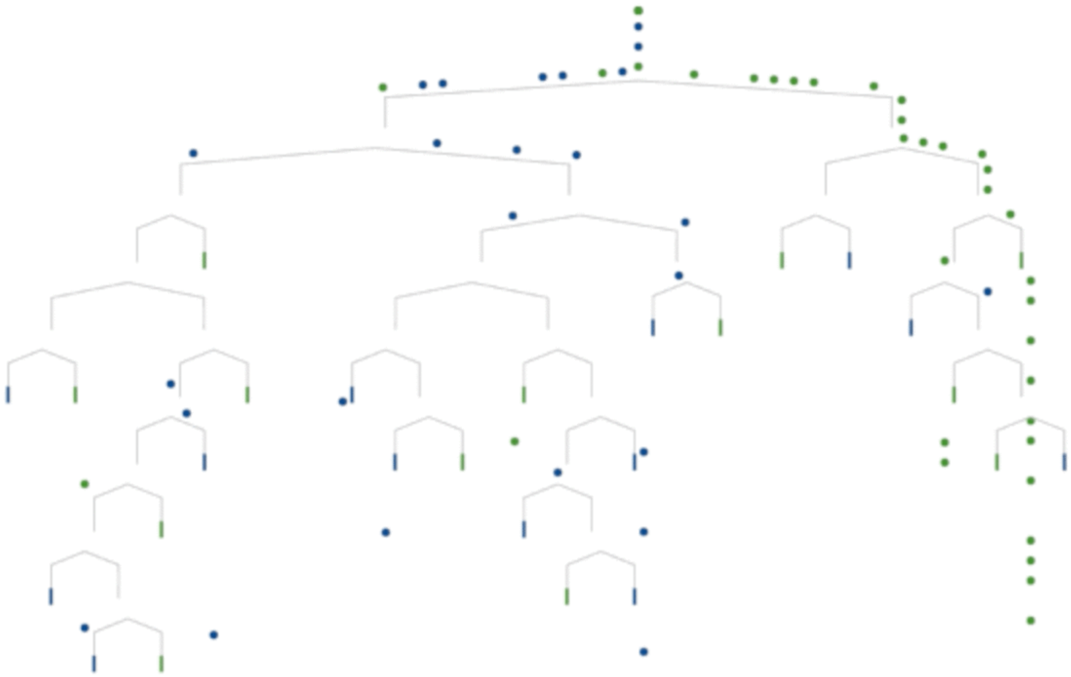
  This kind of model tries to mimic the thought and reasoning processes that are found in the human brain. As it relates to Cybersecurity, the system will learn from various examples that are presented to it, and from there, attempt to make projections as to what the future Threat Landscape will look like. The key advantage here is that it can do all of this in real time and provide needed answers to the IT Security in just a matter of a few seconds. An illustration of this can be seen below:



(SOURCE: 3).

- <u>Decision Trees</u>:

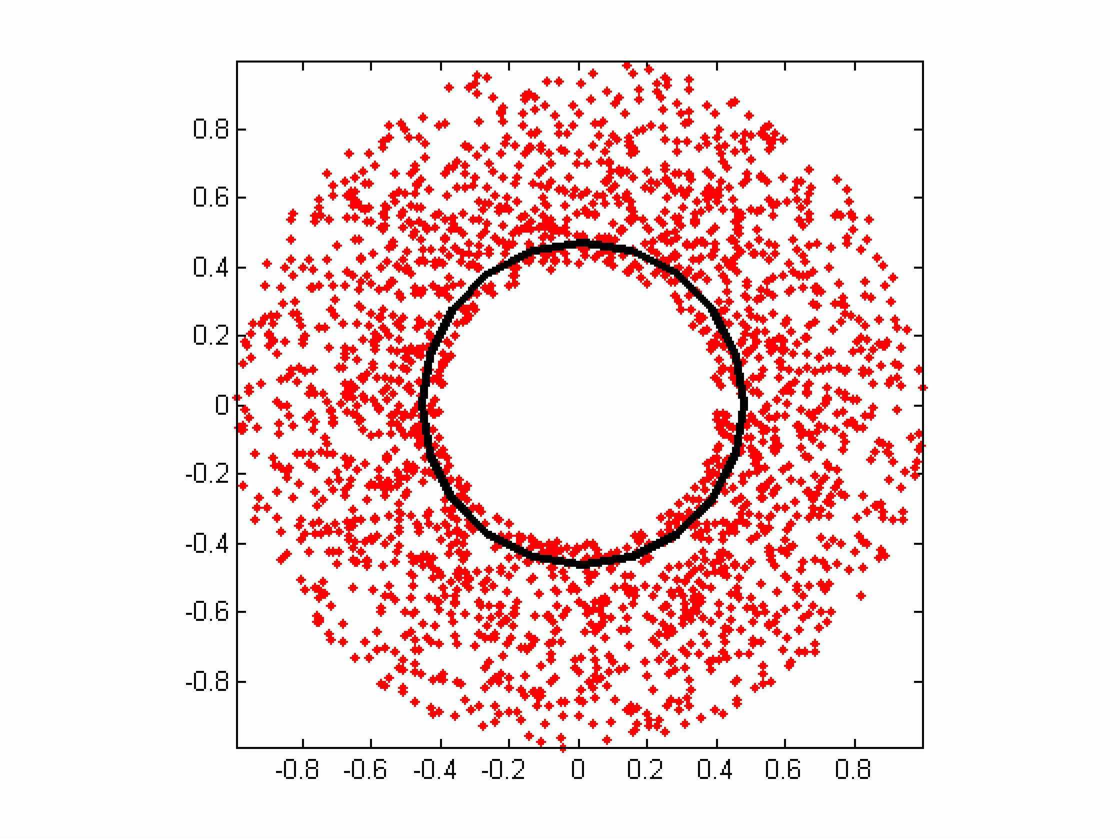
  This is a type of model that relies more upon the principles of statistics. For example, observations can be made about a potential Cyberthreat, and then conclusions can be drawn from it, as to where it could it hit in the IT Infrastructure of an organization. This kind of model is illustrated below:



(SOURCE: 3).

- <u>The Gaussian Mixture Model</u>:

  This type of model is a much more sophisticated by one nature, which makes use of very sophisticated Data Clustering Algorithms and Pattern Recognition. This kind of Machine Learning model is used in another subset of Artificial Intelligence known as “Computer Vision”. This is illustrated below:



(SOURCE: 3).

Applications of Machine Learning

Apart from its potential, heavy usage in Cybersecurity, Machine Learning is used in different sectors as well. Examples of this are as follows:

- Insurance Claims Analysis: Various types of algorithms are being used to help price insurance policy premiums, as well as detect fraudulent behavior in a real time basis.

- Medical Diagnosis: In this scenario, Machine Learning is being used to efficiently store all kinds of medical information and data in a safe and secure manner. The algorithms that have evolved in this field have become so sophisticated that they are even being used to predict the medical diagnoses of patients.

- Search Engines: We all use Google every day to search for information on the Internet, and the results are displayed in just a second or two. How is this done? It has been made possible using Machine Learning. End user behavior is also carefully analyzed so that search results can be greatly improved over time.

- The Financial Markets: Machine Learning is being used here to help predict how the DOW, NASDAQ, S&P 500, etc. will react to certain events that take place on a global basis. Also, traders and managers are using the algorithms of Machine Learning to spot unhidden trends and make trades instantaneously based upon that.

Conclusions

Overall, this blog examined what Machine Learning is, how it works, and the various algorithms and models that use it. It holds a great promise in Cybersecurity, especially when it comes to automating repetitious tasks, and predicting what future threat vectors could look like. But there are limitations to it as well, and this will be examined further in a future blog.

Sources

1) https://www.expertsystem.com/machine-learning-definition/
2) https://searchenterpriseai.techtarget.com/definition/machine-learning-ML
3) https://emerj.com/ai-glossary-terms/what-is-machine-learning/