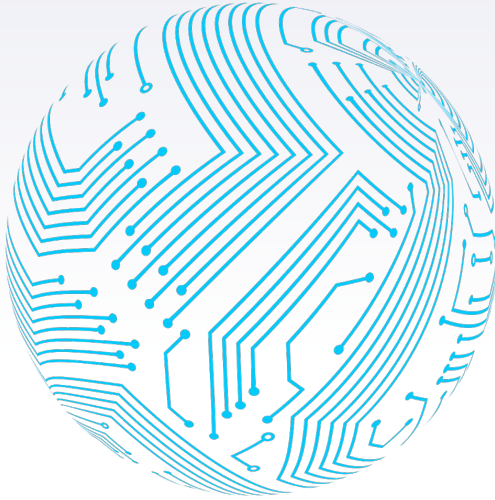


Inspired by:



 DATA SCIENCE
INSTITUTE

The World Data Science Institute is a Financial Data Science Research & Development Company!





Data Science And CyberSecurity

And how we got married

Author: Dairenkon Majime

What is cybersecurity?

A TOOL

Cybersecurity can be viewed as a set of tools to protect, with reliable ways, computers, networks, programs or data from companies, governments or even individual persons.

A PRACTICE

As well a tool, cybersecurity has many ways to protect data, and it is equipped with several processes to accomplish this goal.

What is data science?

A TOOL

Data science can be viewed as a set of tools to analyze and use data for decision-making processes. Can include machine learning algorithms to predict new events under past events, or infer things like behaviors using statistical methods for that.

A PRACTICE

As well as a tool, data science requires scientific methods to criticize results and make worthwhile outcomes.

A NECESSITY

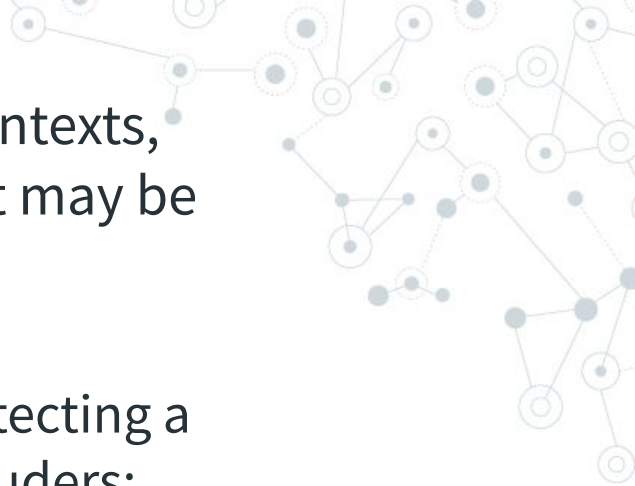
A few roles work together for data science to be what it is. Roles like mathematics, statistics in special, scientific computation, and business knowledge are required to make a good data science job.

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines, with some nodes highlighted in blue.

1.

Cybersecurity

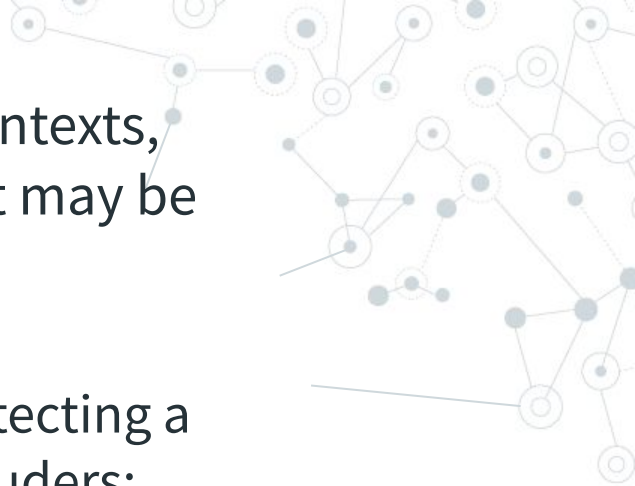
Let's delve into this!



The phrase "cybersecurity" is used in a range of contexts, ranging from business to mobile computing, and it may be broken down into various areas. Include:

- network security, which is concerned with protecting a computer network from cyber attackers or intruders;

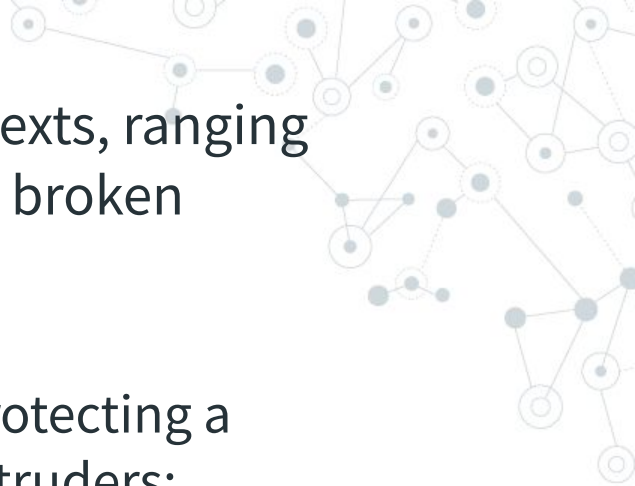




The phrase "cybersecurity" is used in a range of contexts, ranging from business to mobile computing, and it may be broken down into various areas. Include:

- network security, which is concerned with protecting a computer network from cyber attackers or intruders;
- application security, which is concerned with keeping software and devices free of dangers or cyber-threats; and

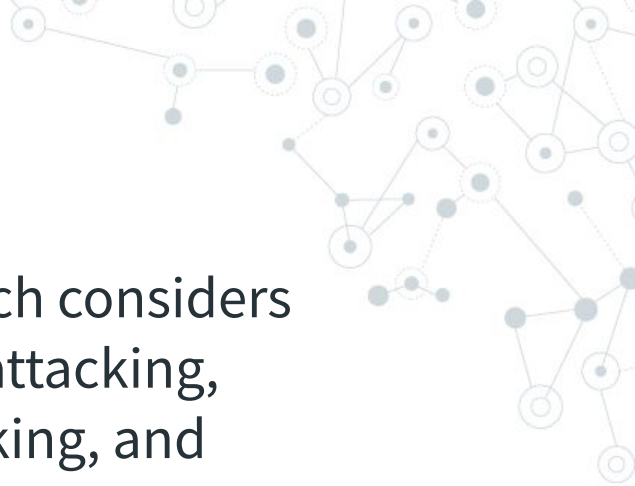





The term "cybersecurity" is used in a range of contexts, ranging from business to mobile computing, and it may be broken down into various areas. Include:

- network security, which is concerned with protecting a computer network from cyber attackers or intruders;
- application security, which is concerned with keeping software and devices free of dangers or cyber-threats; and
- data security, which is concerned with keeping data secure.





The risks typically associated with any attack, which considers three security factors, such as threats, i.e., who is attacking, vulnerabilities, i.e., the weaknesses they are attacking, and impacts, i.e., what the attack does. A security incident is an act that threatens the confidentiality, integrity, or availability of information assets and systems.






Several types of cybersecurity incidents that may result in security risks on an organization's systems and networks or an individual.

Unauthorized access

that describes the act of accessing information to network, systems or data without authorization that results in a violation of a security policy





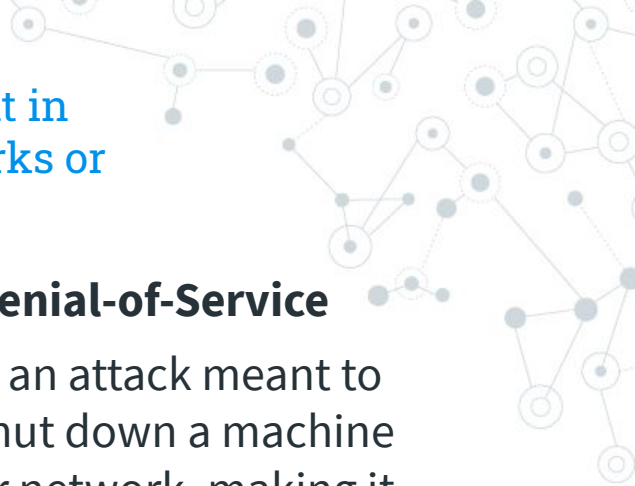
Several types of cybersecurity incidents that may result in security risks on an organization's systems and networks or an individual.

Unauthorized access

that describes the act of accessing information to network, systems or data without authorization that results in a violation of a security policy

Malware

known as malicious software, is any program or software that intentionally designed to cause damage to a computer, client, server, or computer network, e.g., botnets.



Several types of cybersecurity incidents that may result in security risks on an organization's systems and networks or an individual.

Unauthorized access

that describes the act of accessing information to network, systems or data without authorization that results in a violation of a security policy.

Malware

known as malicious software, is any program or software that intentionally designed to cause damage to a computer, client, server, or computer network, e.g., botnets.

Denial-of-Service

is an attack meant to shut down a machine or network, making it inaccessible to its intended users by flooding the target with traffic that triggers a crash.



But, what are the defense strategies in cybersecurity?

Defense strategies are needed to protect data or information, information systems, and networks from cyber-attacks or intrusions.



But, what are the defense strategies in cybersecurity?

More granularly, they are responsible for preventing data breaches or security incidents and monitoring and reacting to intrusions, which can be defined as any kind of unauthorized activity that causes damage to an information system.



But, what are the defense strategies in cybersecurity?

Signature-based IDS

A signature can be a predefined string, pattern, or rule that corresponds to a known attack. A particular pattern is identified as the detection of corresponding attacks in a signature-based IDS.

Anomaly-based IDS

In an anomaly-based intrusion detection system, the behavior of the network is first examined to find dynamic patterns, to automatically create a data-driven model, to profile the normal behavior, and thus it detects deviations in the case of any anomalies



1.

Data Science

Let's delve into this!

Let's review some concepts

Artificial Intelligence

A Science of making things smart or, in other words, human tasks performed by machines (e.g., Visual Recognition, NLP, etc.)..

Data Science

The interdisciplinary field of data collection, preprocessing, inferring, or making decisions by analyzing the data.

Machine Learning

An Approach(just one of many approaches) to AI that uses a system that is capable of learning from experience.

Let's review some concepts

Artificial Intelligence

A Science of making things smart or, in other words, human tasks performed by machines (e.g., Visual Recognition, NLP, etc.)..

The main point is that AI is not exactly machine learning or smart things. It can be a classic program installed in your robot cleaner like edge detection.

Data Science

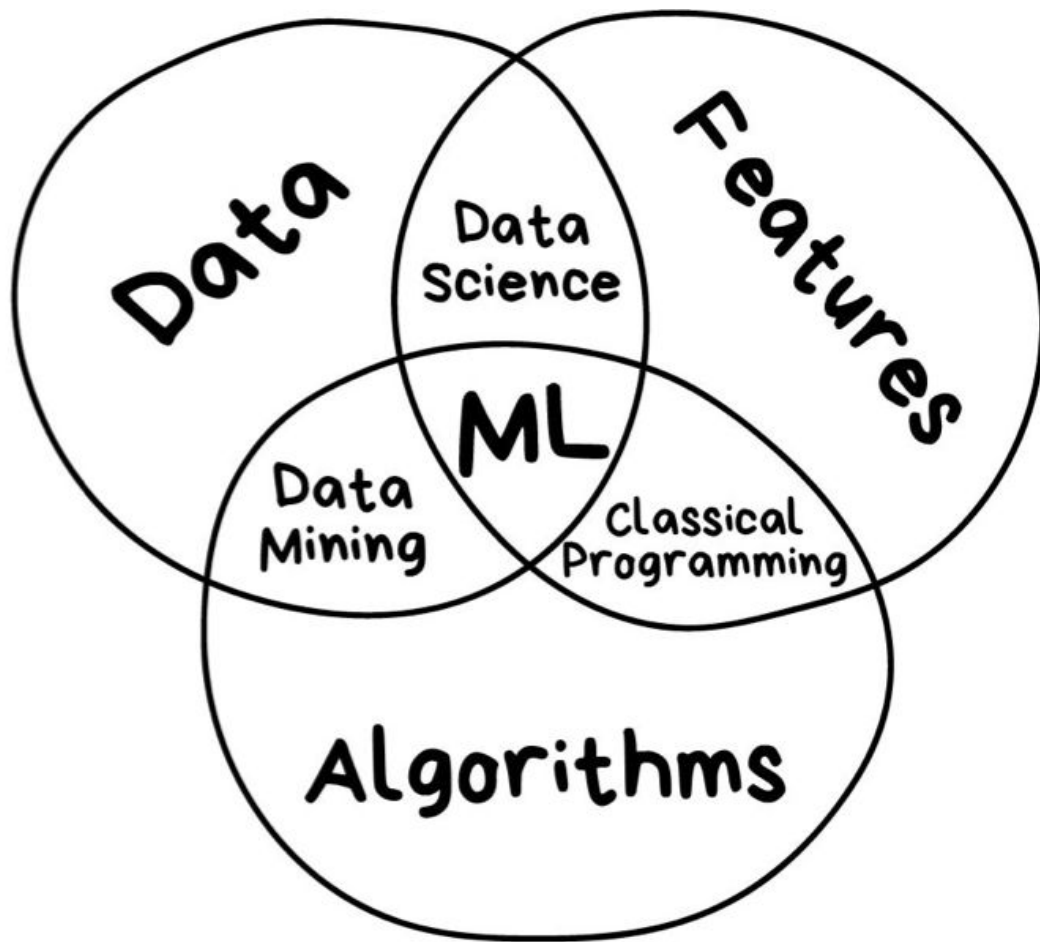
The interdisciplinary field of data collection, preprocessing, inferring, or making decisions by analyzing the data.

It's new interdisciplinary field that synthesizes and builds on statistics, informatics, computing, communication, management, and sociology to study data and its environments, to transform data to insights and decisions by following a data-to-knowledge-to-wisdom thinking and methodology

Machine Learning

An Approach(just one of many approaches) to AI that uses a system that is capable of learning from experience.

In other words, ML is a system that can recognize patterns by using examples rather than by programming them. If your system learns constantly, makes decisions based on data rather than algorithms, and change its behavior, it's Machine Learning.






1.2

Machine Learning







Machine learning can be classified into three major categories concerning methodology

Supervised ML

The targeted labels or classes are already known for the data, and those labels and classes are used to learn for the computations, e.g. classification and regression.






Machine learning can be classified into three major categories concerning methodology

Supervised ML

The targeted labels or classes are already known for the data, and those labels and classes are used to learn for the computations, e.g. classification and regression.

Unsupervised ML

The targeted value is not already known. Unsupervised learning mainly focuses on finding out relationships between samples. It works by finding the patterns among data such as clustering



Machine learning can be classified into three major categories concerning methodology

Supervised ML

The targeted labels or classes are already known for the data, and those labels and classes are used to learn for the computations, e.g. classification and regression.

Unsupervised ML

The targeted value is not already known. Unsupervised learning mainly focuses on finding out relationships between samples. It works by finding the patterns among data such as clustering

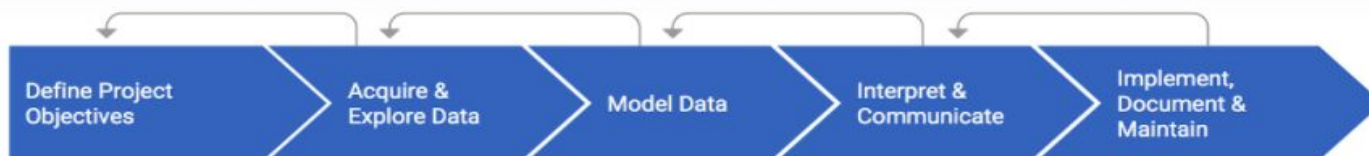
Semi-Supervised ML

Where there is a portion of data labelled or needing human experts during the acquisition of data. The human expert during the labelling process will surely help to solve the problem and improve the accuracy of the model.

Watch a short video explaining a little bit about ML!



The Machine Learning Life Cycle



1. Define Project Objectives

- ☐ Specify business problem
- ☐ Acquire subject matter expertise
- ☐ Define unit of analysis and prediction target
- ☐ Prioritize modeling criteria
- ☐ Consider risks and success criteria
- ☐ Decide whether to continue

2. Acquire & Explore Data

- ☐ Find appropriate data
- ☐ Merge data into single table
- ☐ Conduct exploratory data analysis
- ☐ Find and remove any target leakage
- ☐ Feature engineering

3. Model Data

- ☐ Variable selection
- ☐ Build candidate models
- ☐ Model validation and selection

4. Interpret & Communicate

- ☐ Interpret model
- ☐ Communicate model insights

5. Implement, Document & Maintain

- ☐ Set up batch or API prediction system
- ☐ Document modeling process for reproducibility
- ☐ Create model monitoring and maintenance plan

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric rings, suggesting a hierarchical or multi-layered structure. The lines are thin and gray, connecting the nodes in a non-linear fashion.

2.

AI and Cybersecurity

A decorative network diagram in the bottom-right corner, similar to the one in the top-left. It shows a cluster of nodes connected by lines, with some nodes being larger and more prominent than others, creating a sense of depth and complexity.

Well, but how is data science related with cybersecurity?

Let's see some practical cases and discover together!

Due to the pandemic, more employees are working from home than ever before. To stay updated with the work and collaborate, employees and even college students are using text messages. Whether it is SMS or internet-based texting application like WhatsApp or Telegram hackers under the pretense of the umbrella-term "COVID-19" are phishing and scamming people.

Well, but how is data science related with cybersecurity?

Let's see some practical cases and discover together!

Due to the pandemic, more employees are working from home than ever before. To stay updated with the work and collaborate, employees and even college students are using text messages. Whether it is SMS or internet-based texting application like WhatsApp or Telegram hackers under the pretense of the umbrella-term "COVID-19" are phishing and scamming people.

In this Machine learning use case, the MTD system(Mobile Threat Defense System) is used. In this, ML models are trained to segregate the hackers from genuine informational Covid-19 messages. Like mobile, laptops, PC, etc., different endpoints are safeguarded. They are safeguarded by the Unified Endpoint Management program. UEM is highly effective for text-based applications and SMSs. Herein, the model is trained with many datasets to identify the threats amongst the authentic messages.

Well, but how is data science related with cybersecurity?

Let's see some practical cases and discover together!

Machine learning is already abundant when it is concerned with mobile devices. Whether it is iOS or Android, data privacy, security patches, anti-virus applications already use ML. Google is already using Machine Learning in security for mobile devices. ML is used to prevent cyber attacks in networks, devices, and vulnerability assessment tools themselves.

Well, but how is data science related with cybersecurity?

Let's see some practical cases and discover together!

Machine learning is already abundant when it is concerned with mobile devices. Whether it is iOS or Android, data privacy, security patches, anti-virus applications already use ML. Google is already using Machine Learning in security for mobile devices. ML is used to prevent cyber attacks in networks, devices, and vulnerability assessment tools themselves.

Wandera, a cybersecurity space leader, uses its ML algorithm. They detected 500 ransomware strains in the different companies' business mobile devices. Apple's Siri, Google Assistant, and Amazon's Alexa, are personal, AI-driven assistance. They have significant responsibilities of securing the voice-based commands using ML. Also, to identify the actual owner's voice against a hacker's control.

Well, but how is data science related with cybersecurity?

Let's see some practical cases and discover together!

There is no doubt that Machine learning and AI are better than humans when identifying any loopholes or making any errors. ML in Cybersecurity was introduced when data usage increased rapidly. For humans finding and analyzing any threats was considered as finding a needle in a haystack. MIT introduced a system called AI2. It is an adaptive machine learning security platform that helped analysts find those 'needles in the haystack.'

Well, but how is data science related with cybersecurity?

Let's see some practical cases and discover together!

There is no doubt that Machine learning and AI are better than humans when identifying any loopholes or making any errors. ML in Cybersecurity was introduced when data usage increased rapidly. For humans finding and analyzing any threats was considered as finding a needle in a haystack. MIT introduced a system called AI2. It is an adaptive machine learning security platform that helped analysts find those 'needles in the haystack.'

This system could filter out all the malicious activities out of millions of actions taken during one day. AI2 brought down the threat rate by 85%. Vulnerability assessment tools became common among analysts for detections of any attacks.

Well, but how is data science related with cybersecurity?

Let's see some practical cases and discover together!

Network security is of utmost importance for any business. Understanding the various topology of the network security architecture is a challenge. Even for many cybersecurity specialists. With the amount of data coming in and out of the network, it is no joking matter. Along with analyzing the data, maintaining the web, and identifying the connection behavior.

Well, but how is data science related with cybersecurity?

Let's see some practical cases and discover together!

Network security is of utmost importance for any business. Understanding the various topology of the network security architecture is a challenge. Even for many cybersecurity specialists. With the amount of data coming in and out of the network, it is no joking matter. Along with analyzing the data, maintaining the web, and identifying the connection behavior.

The enhanced ML-based network security system will track all outgoing and incoming calls/data. To detect any suspicious information patterns in the network. Many software can monitor networks by using anomaly detection software. It is used to alert human authorities in case of discrepancies in data like previous cyber threats.





Thank you so much!

Sources:

<https://www.hitechnectar.com/blogs/machine-learning-in-cybersecurity/>

<https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00318-5>

<https://towardsdatascience.com/fraud-detection-the-problem-solutions-and-tools-dd8977b435c9>

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9277523>

