Real-world problems

Unit objectives

After completing this unit, you should be able to:

- Understand the network intrusion detection
- Gain knowledge on anomaly detection in big data
- Understand anomaly detection for autonomous robots

Network intrusion detection

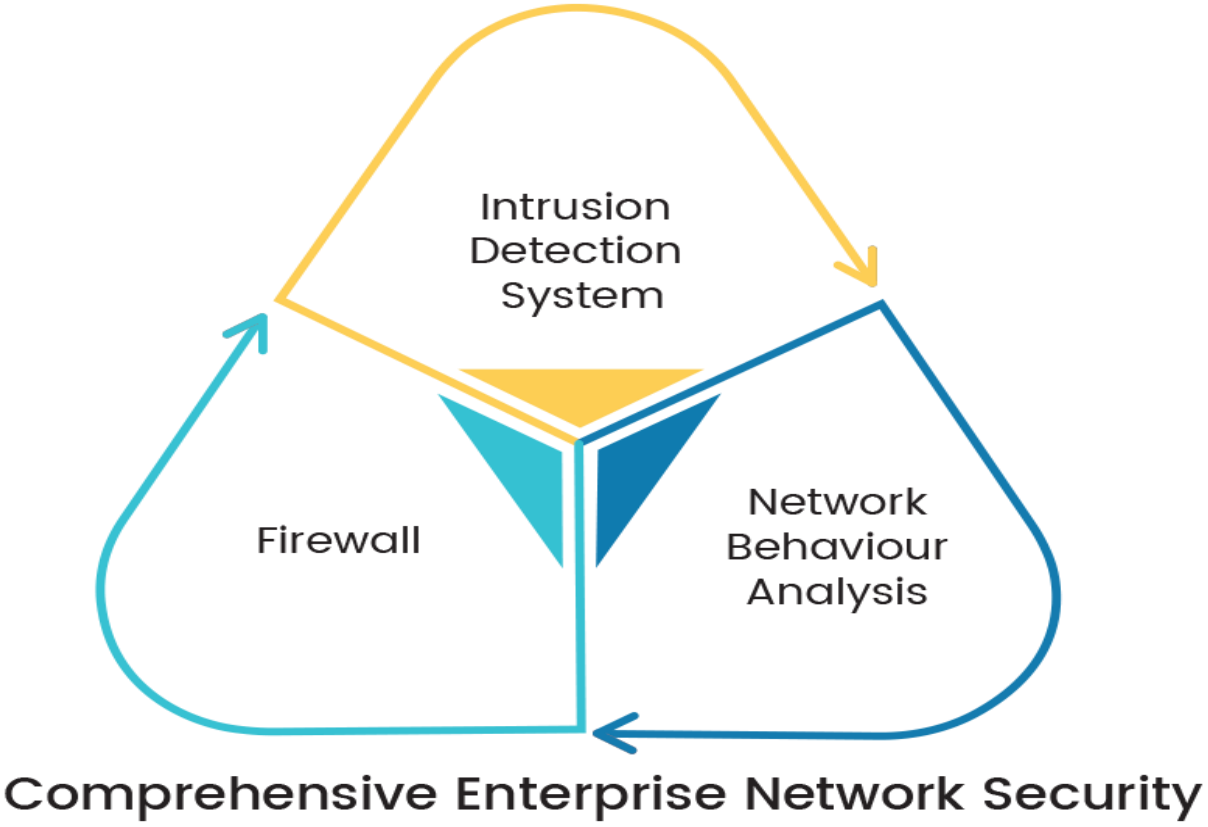


Figure: Network intrusion detection

Source: https://images.app.goo.gl/ZWkhczDWBdatVbmJ8

Understanding of IDS core operation

- A computer or software program carry out these useful roles is an intrusion detection system (IDS):
  - Evaluates a full cyber threat network infrastructure.
  - Senses a cyber threat immediately as it happens.
  - Implements an anti-attack preventive measure (intrusion protection systems) easily.
  - Submit files to a monitoring department or operator.

How an IDS works?

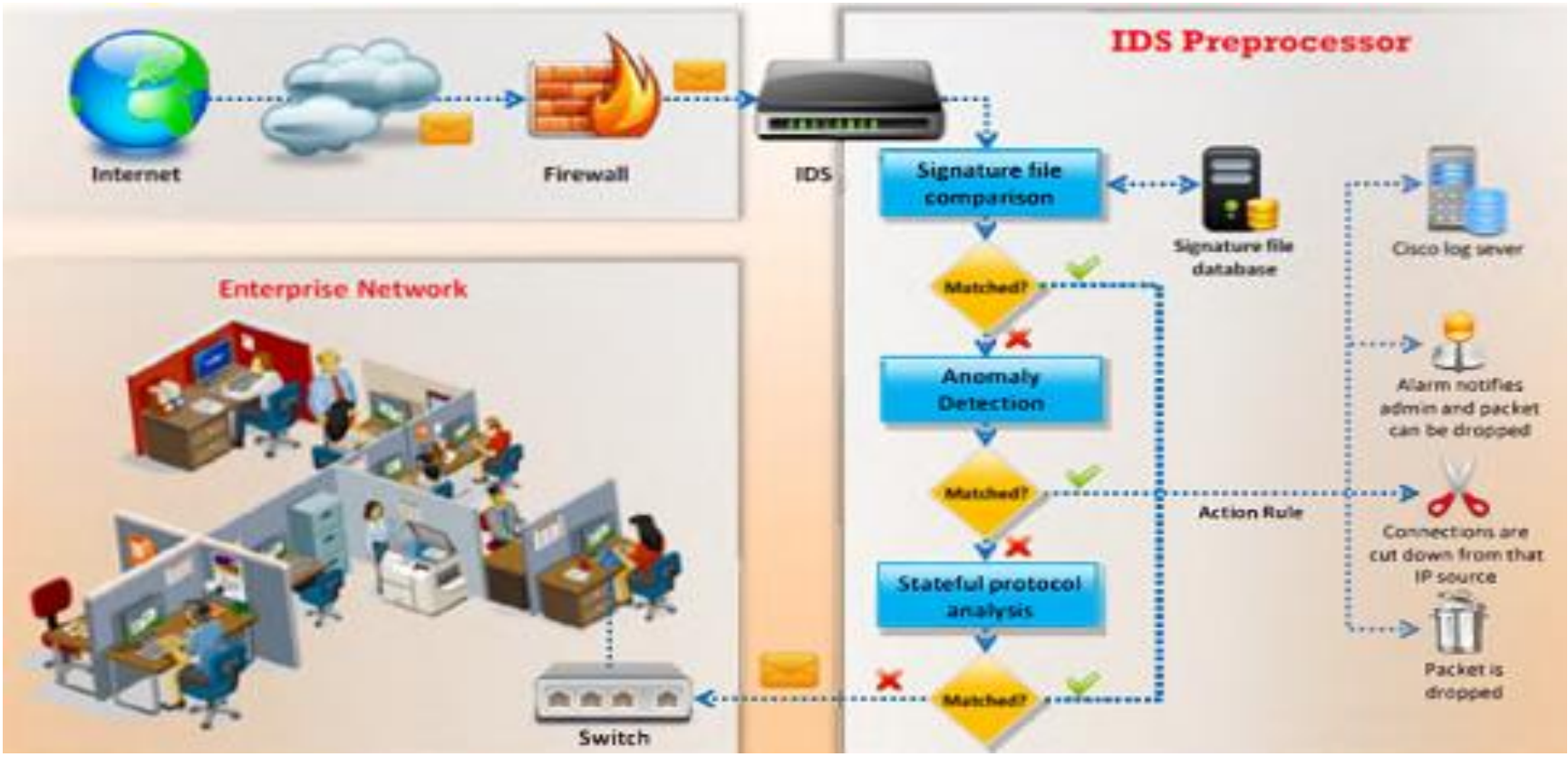


Figure: IDS working model

Source: https://images.app.goo.gl/FaBeWcdTYnL18rBD8

Types of intrusion detection systems

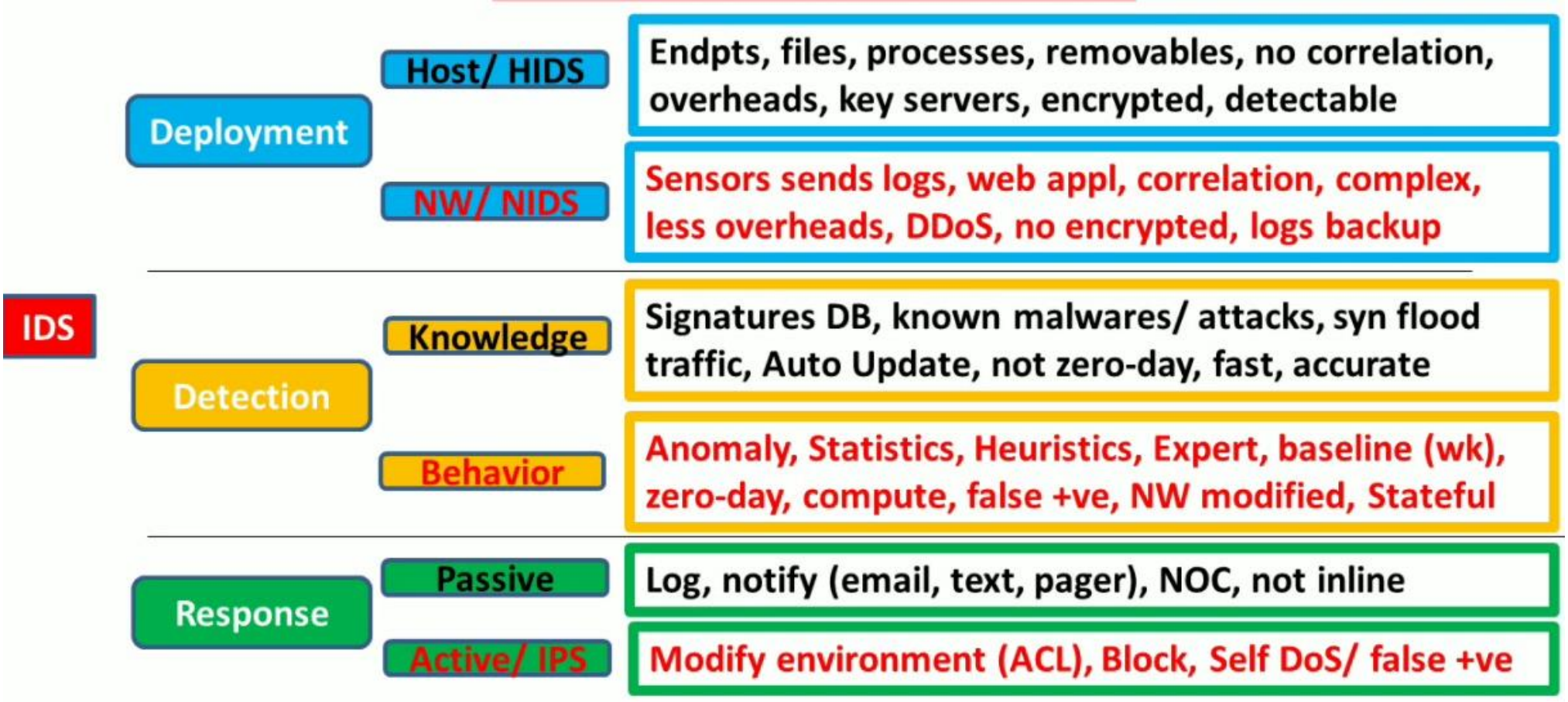


Figure :IDS categories

Source: https://images.app.goo.gl/kg6Hf9kXFMeCxxFfA

Self evaluation: Exercise 20

- To continue with the training, after learning the various steps involved in pattern recognition and anomaly detection, it is instructed to utilize the concepts to perform the following activity.

- You are instructed to write the following activities using python code.

- Exercise 20: OpenCV (Object Detection with CAM).

Fundamental concerns of intrusion detection systems

- IDS is not scalable.
- False positives & negatives.
- Experienced administrators required.
- Encrypted packets.
- Protocol-based attacks.
- Ongoing updates.

Intrusion detection vs. intrusion prevention

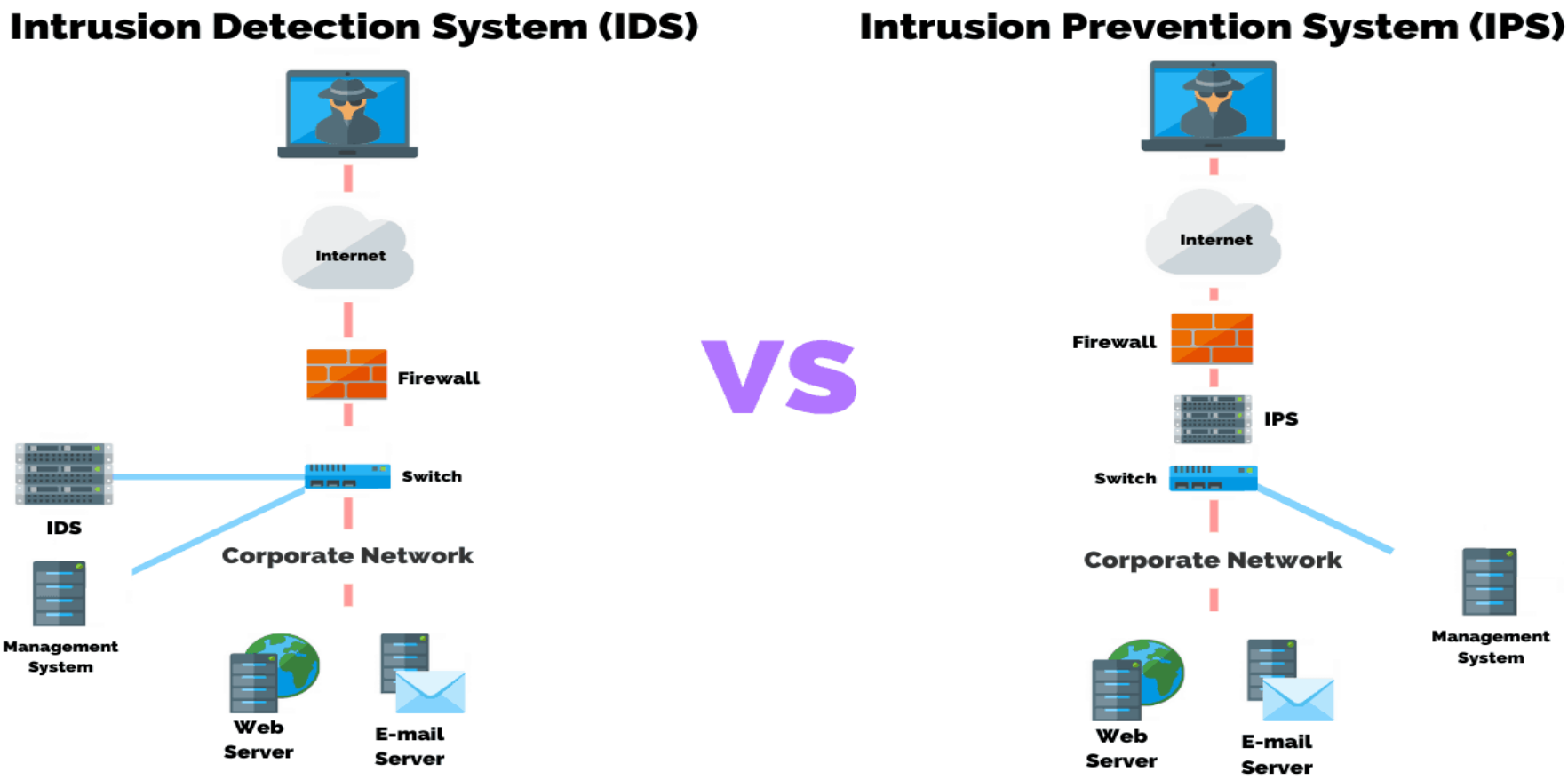


Figure: Intrusion Detection vs. Intrusion Prevention

Source: https://images.app.goo.gl/fXqCTJYByNixx6vT6

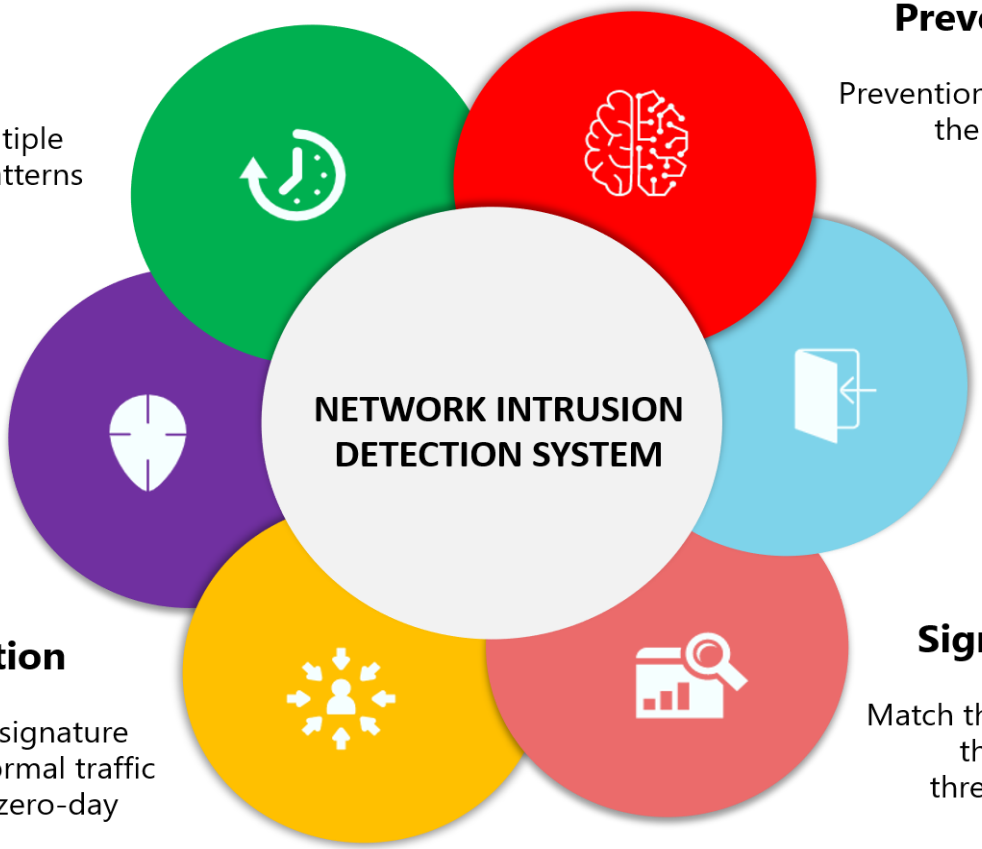
The future of IDS

Network Traffic Processing

Convert the traffic into multiple network traffic parameter patterns (Signatures)

Prevention System

Prevention System acts based on the threat detected



Threat Reporting

Report/Log the threat

Threat Classification

Classify the threat based on Signature matching

Anomaly Detection

Match the input network signature against already available normal traffic signatures to detect the zero-day attacks

Signature Matching

Match the input signature against the already available threat signature patterns

Figure: Securing IoT with Intrusion Detection Systems

Source: https://images.app.goo.gl/yMGXBCuSCMGAHZuM9

Anomaly detection in big data

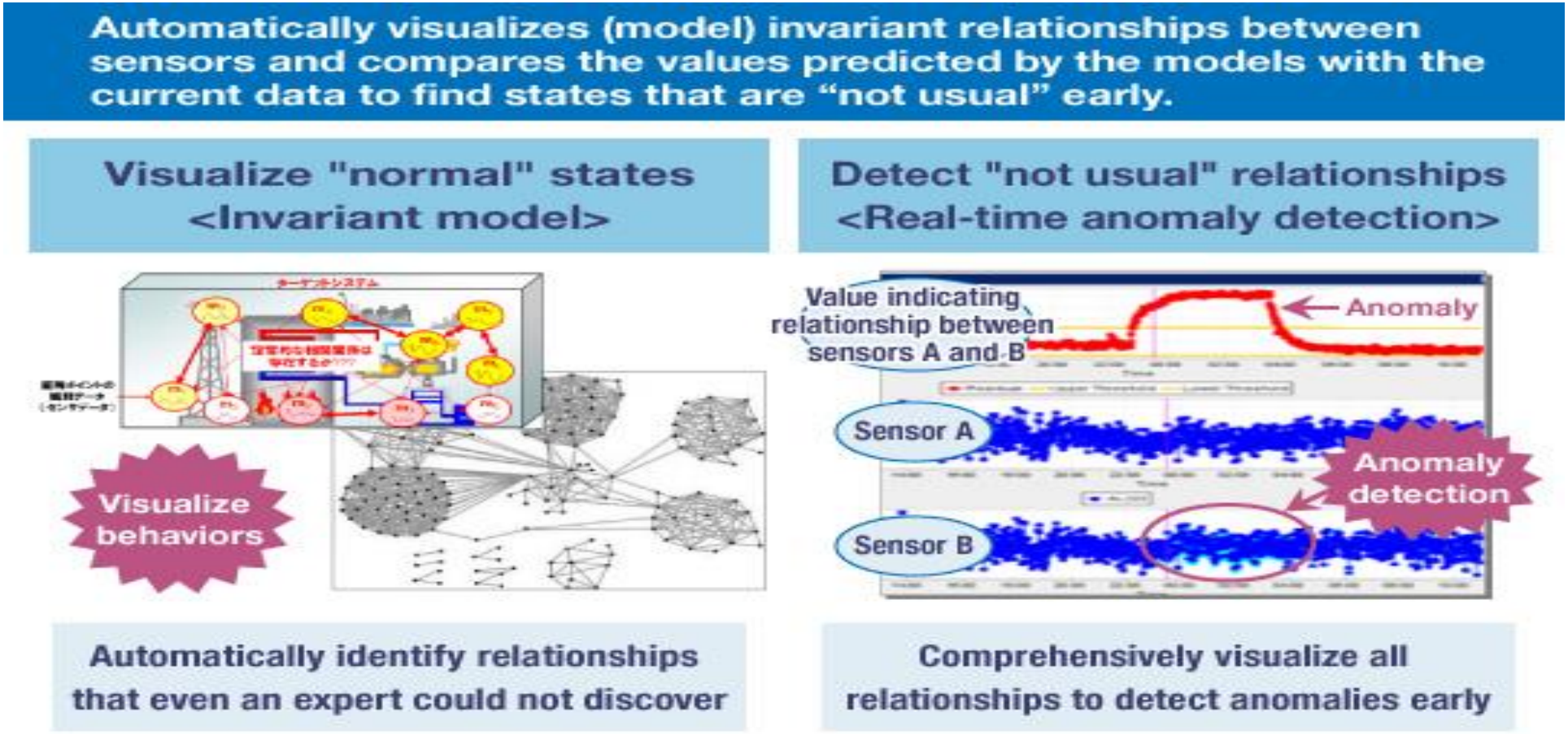


Figure: Bigdata visualization anomalies

Source: https://images.app.goo.gl/9SS6RTaHGHJXkLzk7

Key attributes of advanced anomaly detection

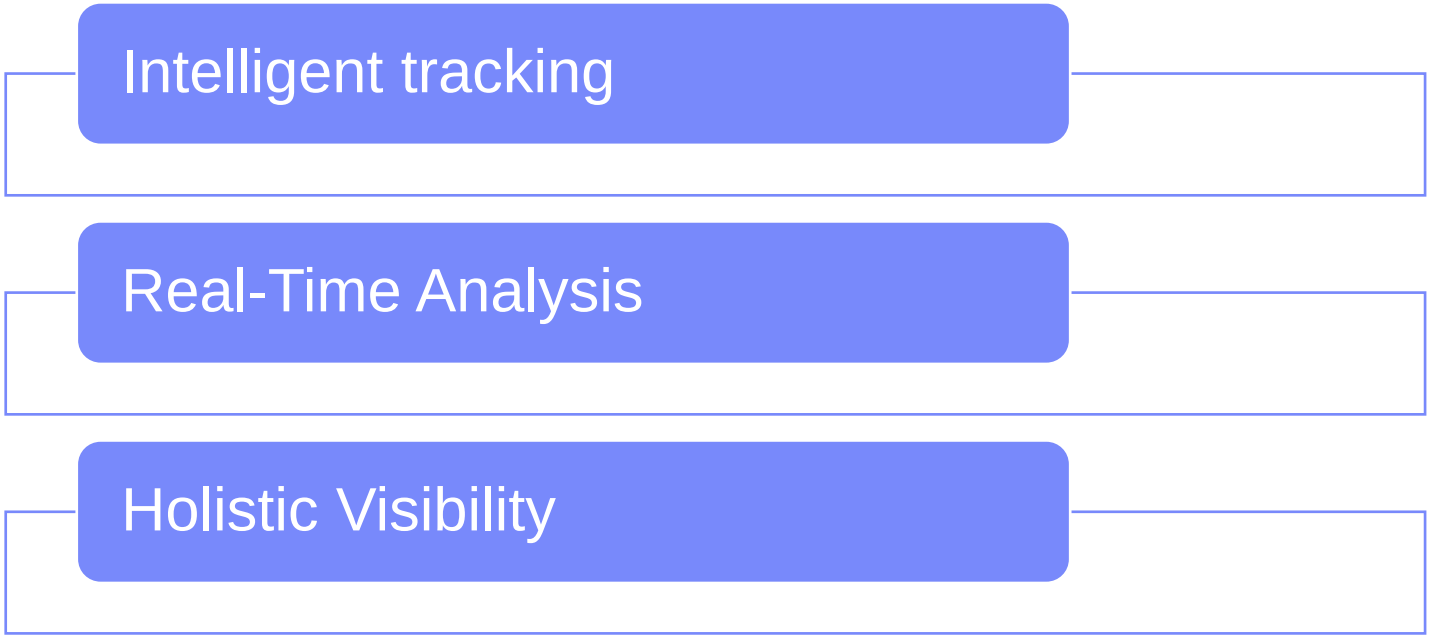


Figure: Key attributes of advanced anomaly detection

Self evaluation: Exercise 21

- To continue with the training, after learning the various steps involved in pattern recognition and anomaly detection, it is instructed to utilize the concepts to perform the following activity.

- You are instructed to write the following activities using python code.

- Exercise 21: OpenCV (Object Detection with Video).

The real-world impact of anomaly detection

Figure: Robot dog reminds park goers about social distancing | Coronavirus

Source: https://images.app.goo.gl/RWug9Bhv44zNNJWz9

Anomaly detection on 5G: Possibilities and opportunities

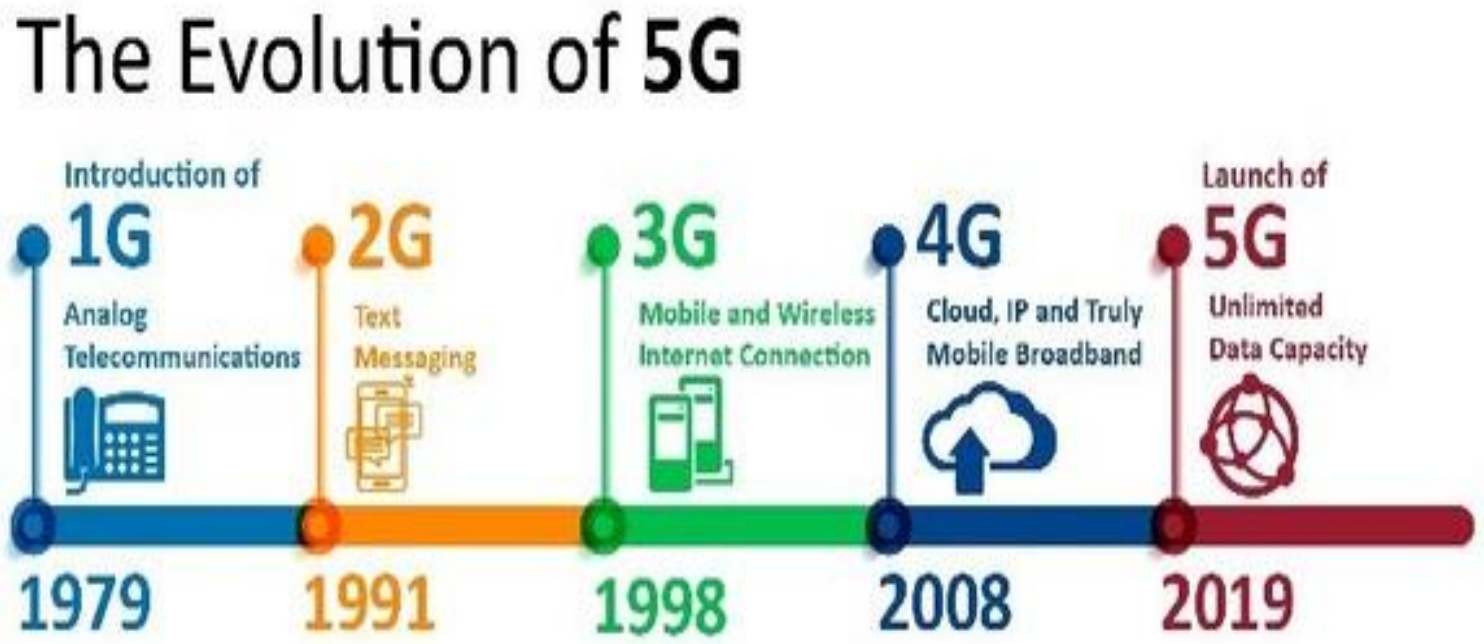


Figure: The Evolution of 5G

Source: https://images.app.goo.gl/43eqdfsmck7Cbhoa7

Real time anomaly detection in docker, Hadoop cluster

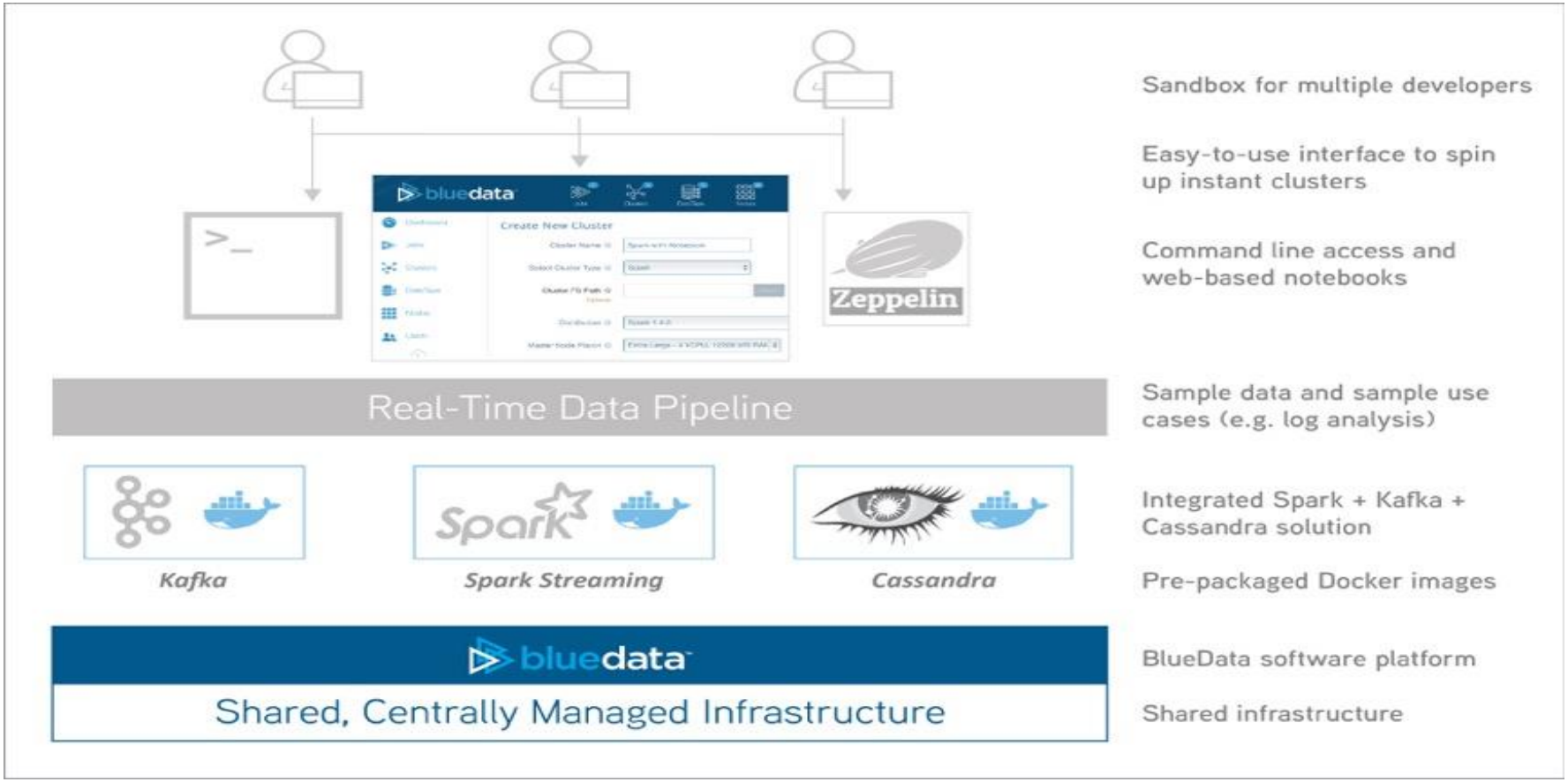


Figure: Real Time Anomaly Detection in Docker, Hadoop cluster

Source: https://images.app.goo.gl/u9hxmmh17NomAnJV8

Anomaly detection in IoT

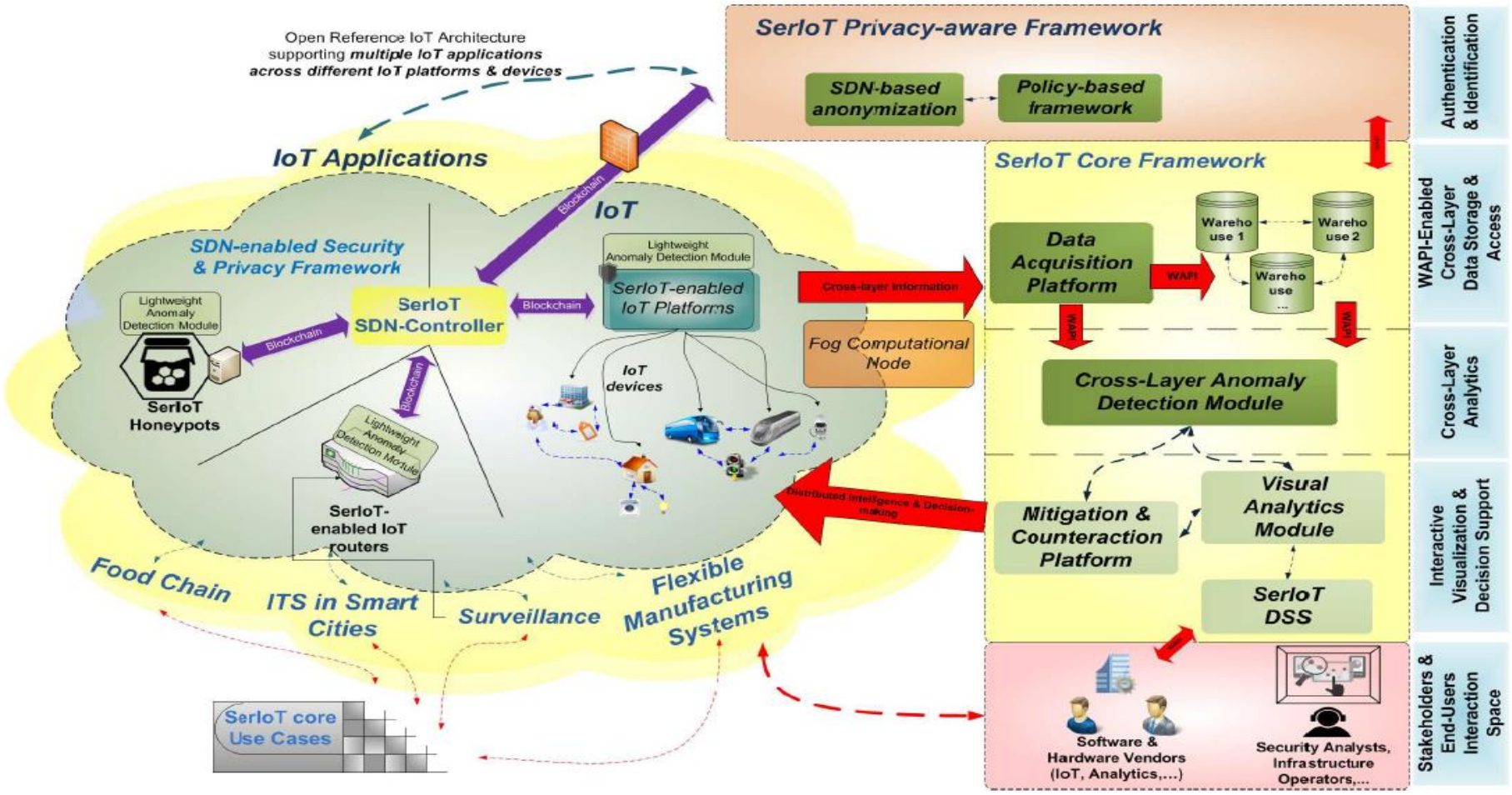


Figure: Anomaly Detection in IoT

Source: https://images.app.goo.gl/ayCxMUueCDq9XE2N8

Detection of deviations in deep learning time series results

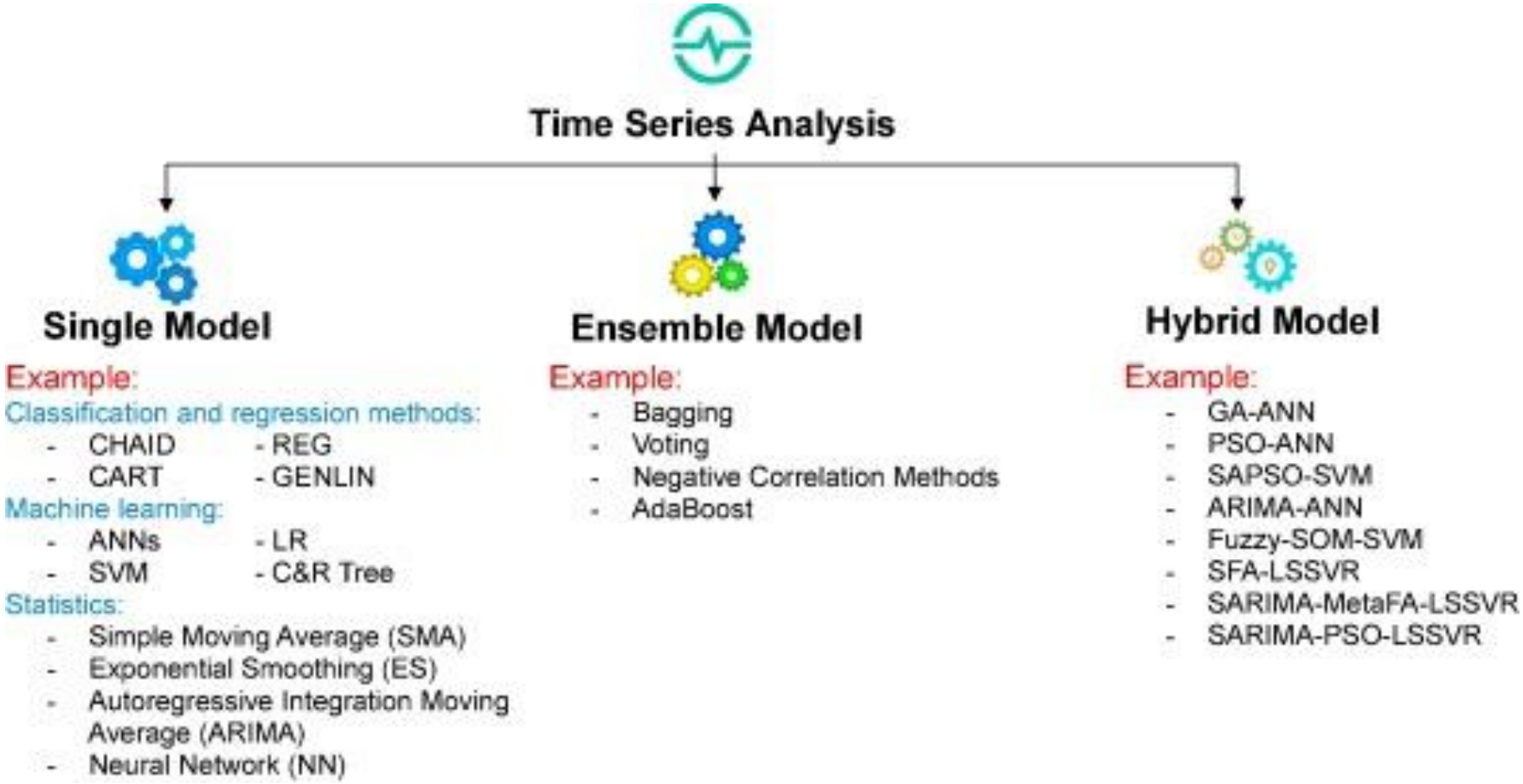


Figure: Detection of deviations in Deep Learning time series results

Source: https://images.app.goo.gl/2UUTc7Eq8w2pDDrV6

Self evaluation: Exercise 22

- To continue with the training, after learning the various steps involved in pattern recognition and anomaly detection, it is instructed to utilize the concepts to perform the following activity.

- You are instructed to write the following activities using python code.

- Exercise 22: OpenCV (Color Filtration).

Anomaly detection use cases

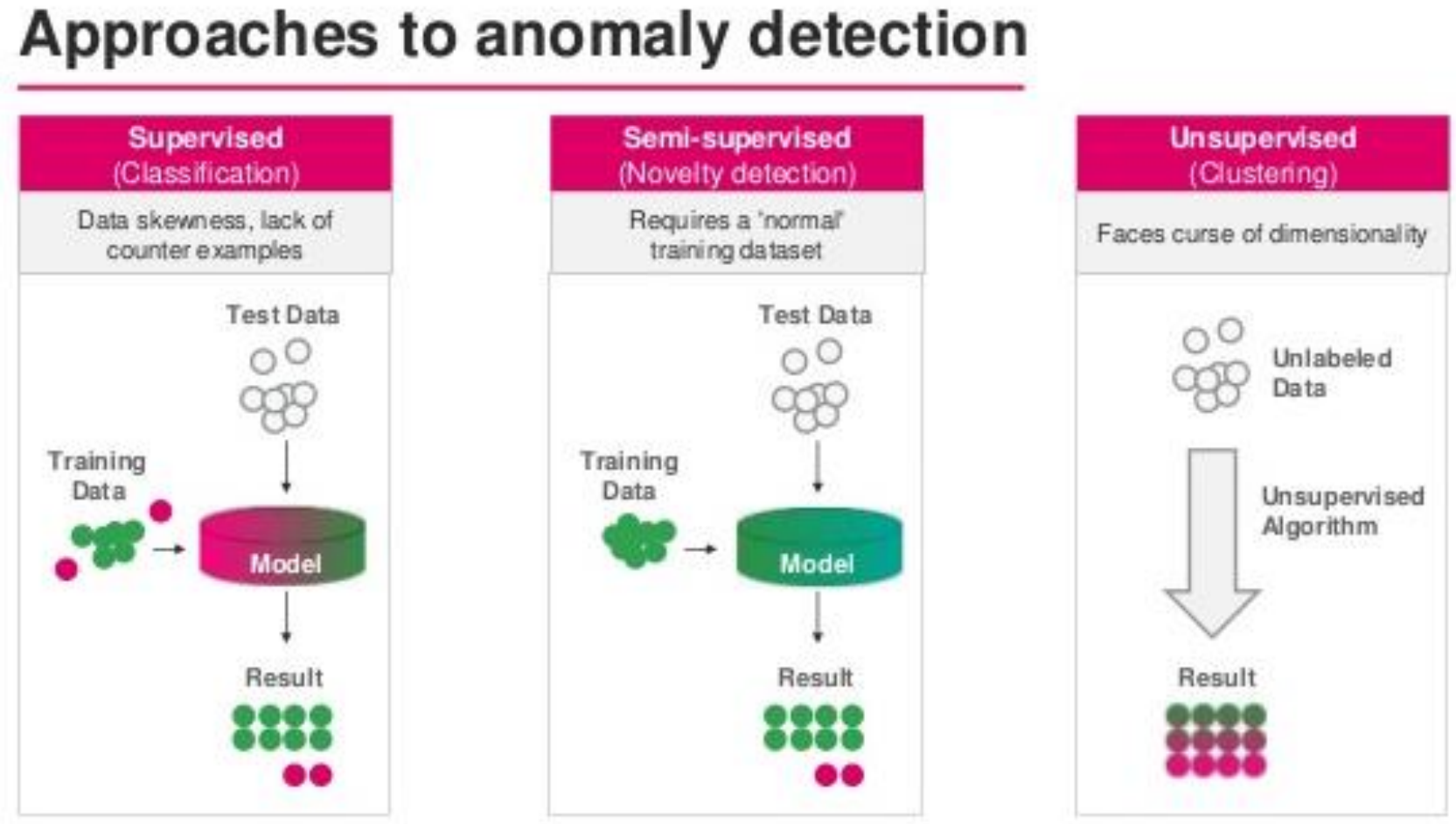


Figure: Approaches to anomaly detection

Source: https://images.app.goo.gl/KQnkiHCvv6Ryqa4V8

Anomaly detection with time series forecasting

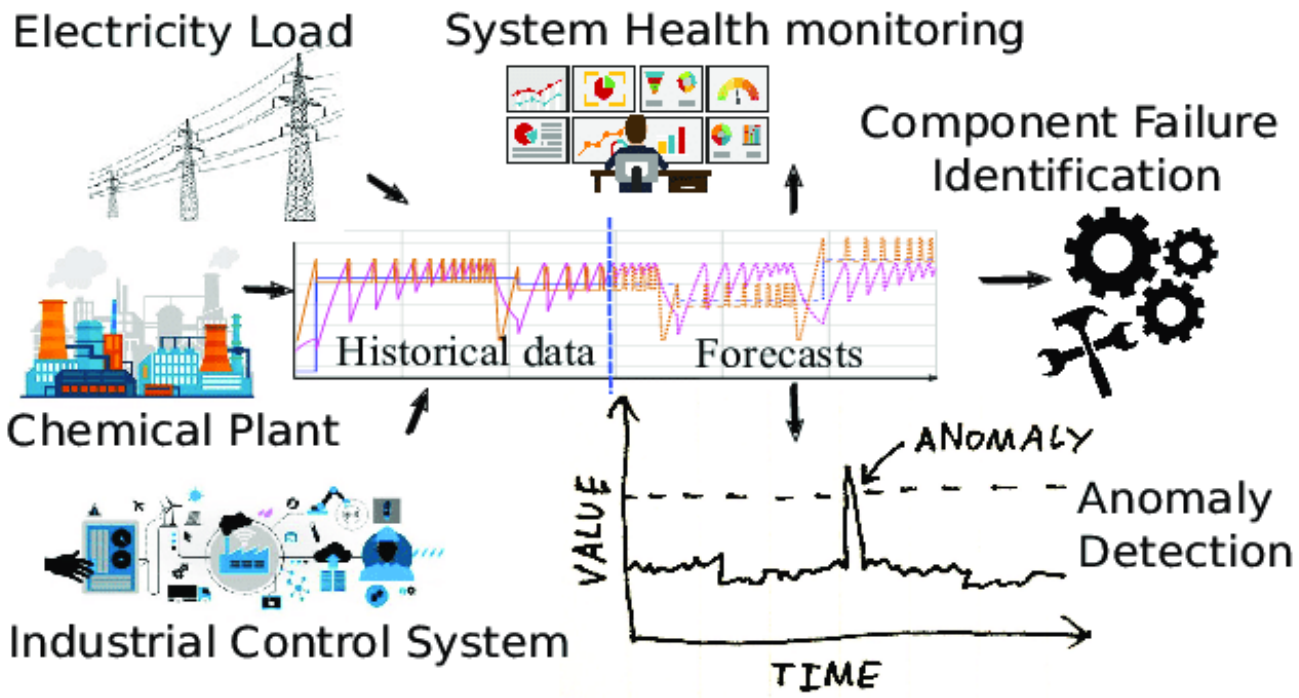


Figure: Anomaly Detection with Time Series Forecasting

Source: https://images.app.goo.gl/9NhyxHrSHdVr6nB78

Self evaluation: Exercise 23

- To continue with the training, after learning the various steps involved in pattern recognition and anomaly detection, it is instructed to utilize the concepts to perform the following activity.

- You are instructed to write the following activities using python code.

- Exercise 23: OpenCV (Object Detection with haar cascade).

What is time series analysis?

- Time period analysis is also time period research. In attempt to elucidate the framework and role provided by the time series, the study is carried out.

- A mathematical model may be conveniently established by knowing the time series process, such that additional forecasts, surveillance and management could be made.

Time series data models

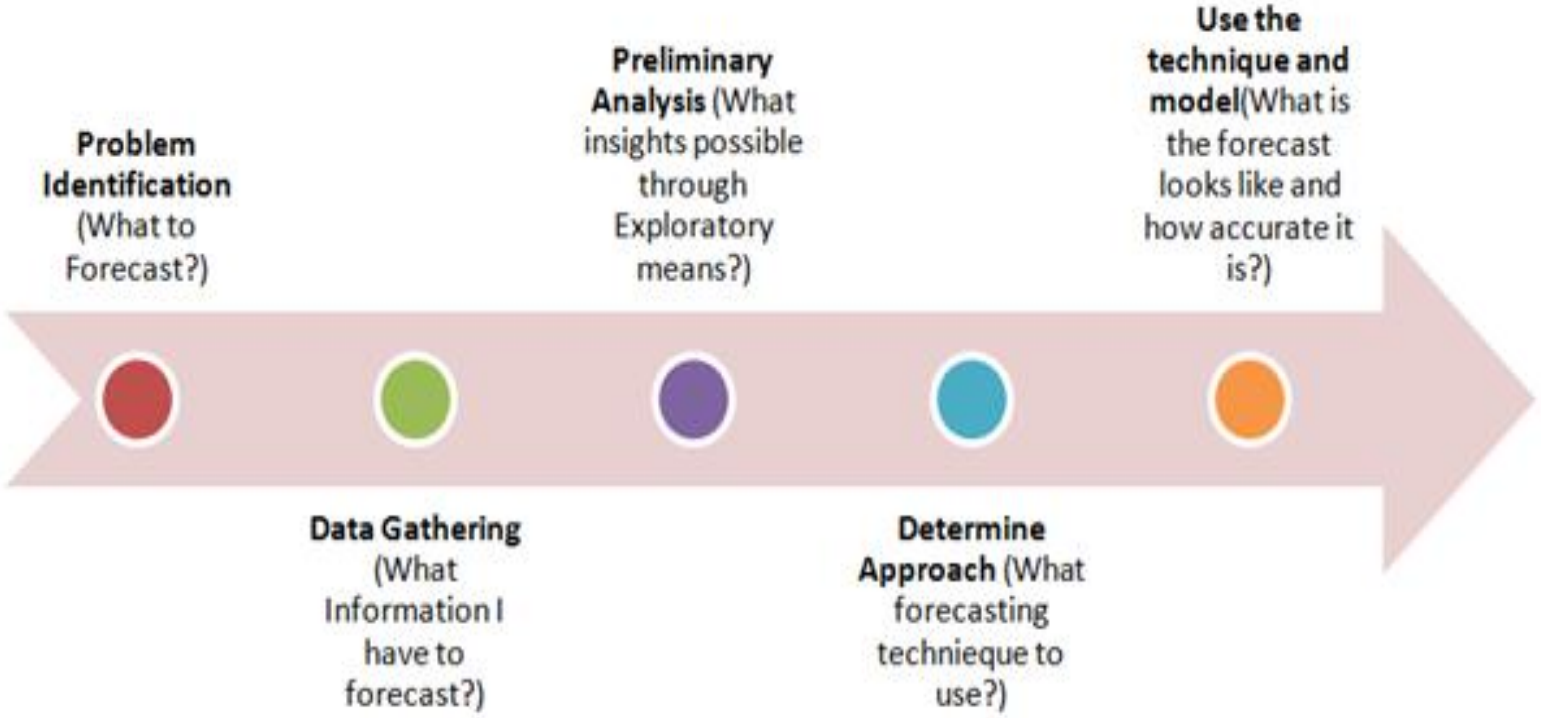


Figure: ARIMA forecasting process

Source: https://images.app.goo.gl/RZ4yJXcDvFZqv7R96

Self evaluation: Exercise 24

- To continue with the training, after learning the various steps involved in pattern recognition and anomaly detection, it is instructed to utilize the concepts to perform the following activity.

- You are instructed to write the following activities using python code.

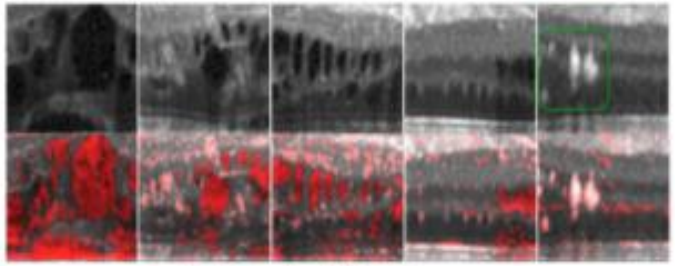
- Exercise 24: Graph theory.

How to find anomaly in time series data?

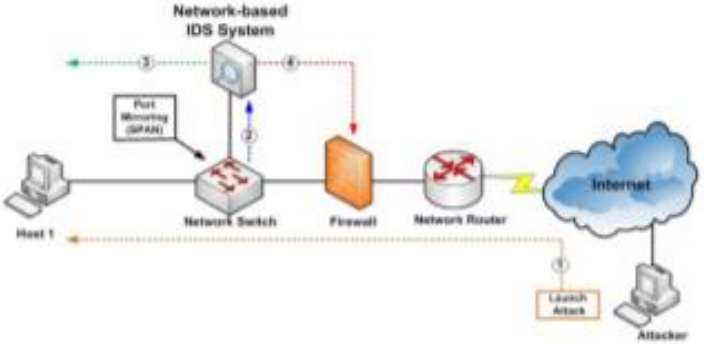
Figure: Anomaly in time series data

Anomaly detection using machine learning

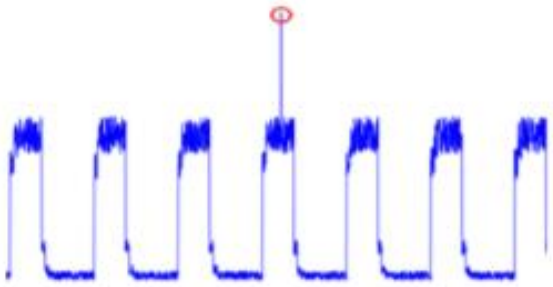
(a) Illegal Traffic Flow detection



(b) Detecting Retinal Damage



(c) Cyber-Network Intrusion detection



(d) Internet Of Things (IoT) Big-Data
Anomaly detection

Figure: Anomaly Detection using Machine Learning

Source: https://images.app.goo.gl/3qwPTMzDRUUU7Ett6

Anomaly detection using deep learning

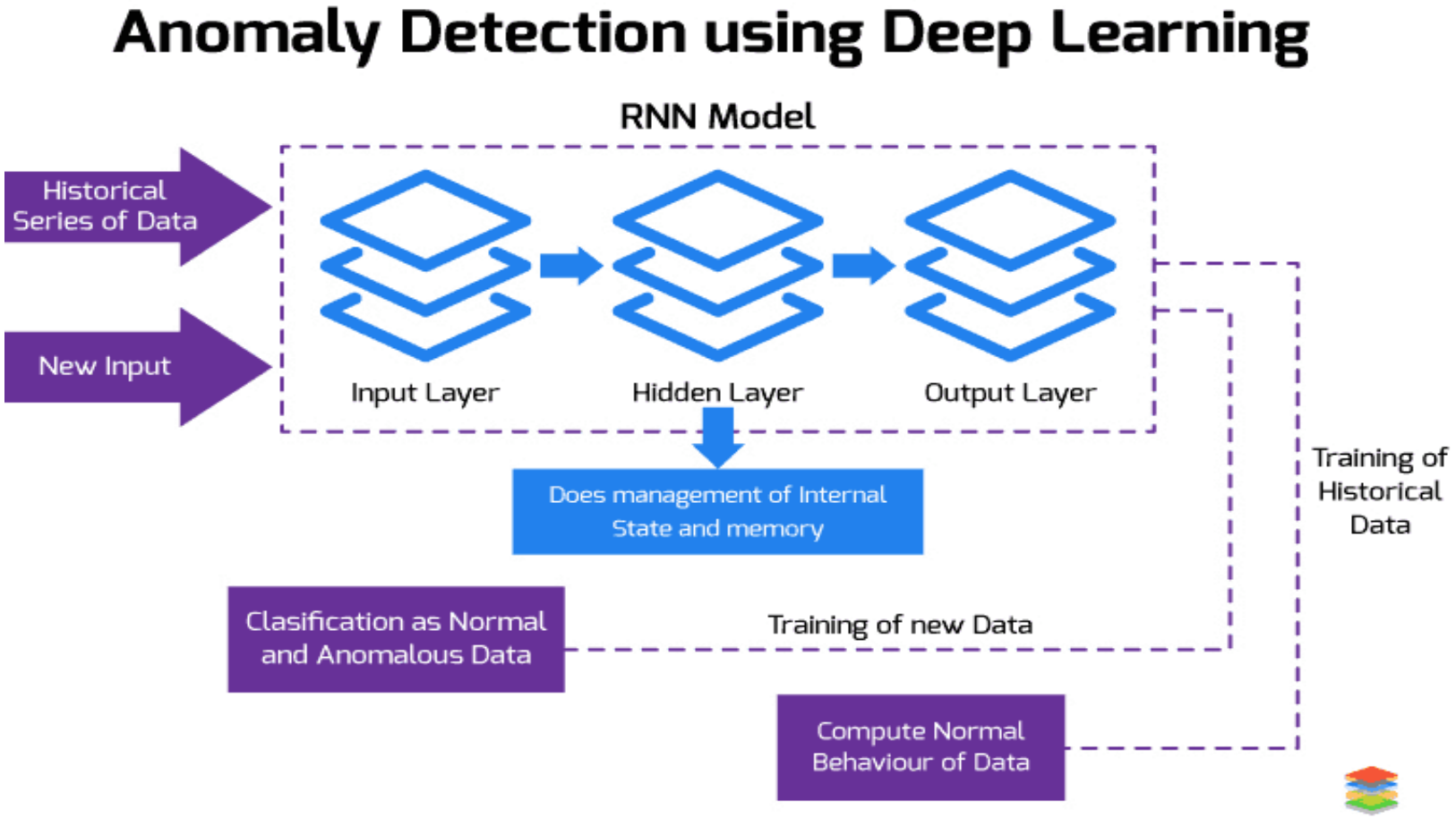


Figure: Anomaly Detection using Deep Learning

Source: https://images.app.goo.gl/NrZJ3oR11XrByVZA8

Anomaly detection for an e-commerce pricing system

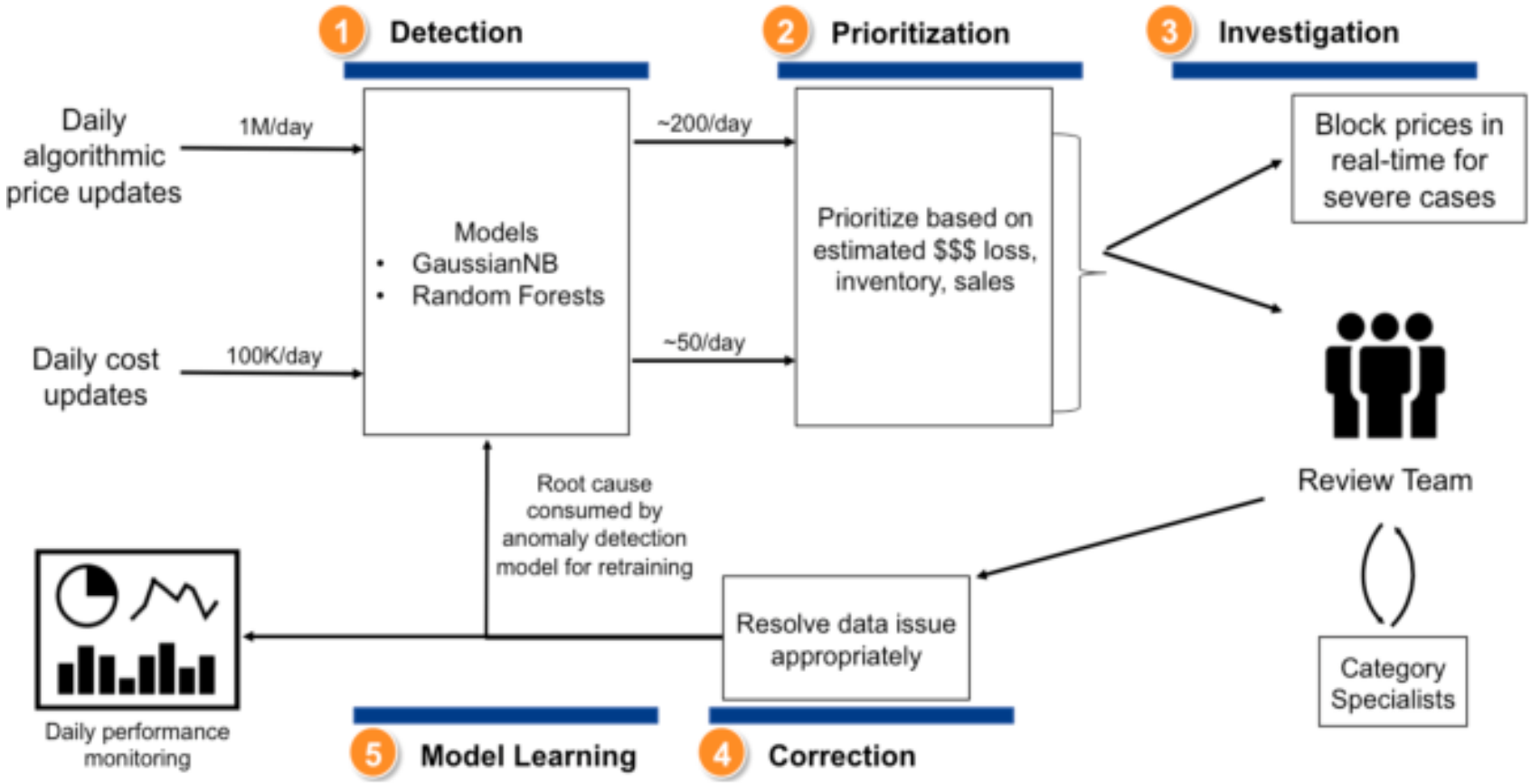


Figure: Anomaly Detection for an E-commerce Pricing System

Source: https://images.app.goo.gl/MtbnBSJFoQha7vkp9

IBM's Watson AIOps automates IT anomaly detection and remediation

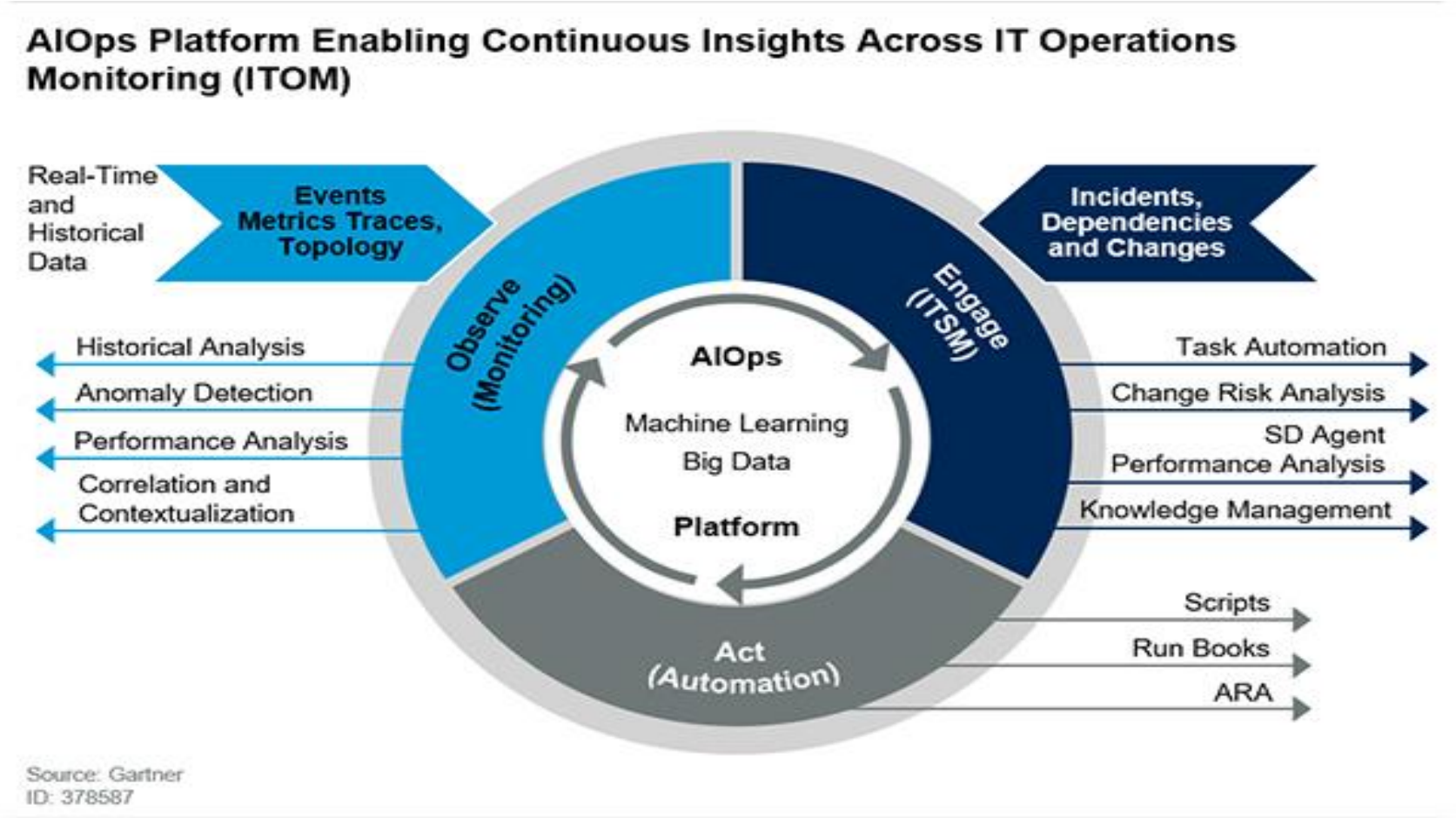


Figure: AIOps

Source: https://images.app.goo.gl/KC2fhvxbxnX6fFzS7

Self evaluation: Exercise 25

- To continue with the training, after learning the various steps involved in pattern recognition and anomaly detection, it is instructed to utilize the concepts to perform the following activity.

- You are instructed to write the following activities using python code.

- Exercise 25: GUI for pattern detection.

Checkpoint (1 of 2)

Multiple choice questions:

1. The numerical output of a sigmoid node in a neural network:
   a) Is unbounded, encompassing all real numbers
   b) Is unbounded, encompassing all integers
   c) Is bounded between 0 and 1
   d) Is bounded between -1 and 1

2. What would you do in PCA to get the same projection as SVD?
   a) Transform data to zero mean
   b) Transform data to zero median
   c) Not possible
   d) None of these

3. Regarding bias and variance, which of the following statements are true?
   a) Models which overfit have a high bias
   b) Models which overfit have a low bias
   c) Models which underfit have a high variance
   d) Models which underfit have a low variance

Checkpoint solutions (1 of 2)

Multiple choice questions:

1. The numerical output of a sigmoid node in a neural network:
   a) Is unbounded, encompassing all real numbers
   b) Is unbounded, encompassing all integers
   c) Is bounded between 0 and 1
   d) Is bounded between -1 and 1

2. What would you do in PCA to get the same projection as SVD?
   a) Transform data to zero mean
   b) Transform data to zero median
   c) Not possible
   d) None of these

3. Regarding bias and variance, which of the following statements are true?
   a) Models which overfit have a high bias
   b) Models which overfit have a low bias
   c) Models which underfit have a high variance
   d) Models which underfit have a low variance

Checkpoint (2 of 2)

Fill in the blanks:

1. In k nearest neighbors, k=1 increases the _____.
2. K-means _____ is guaranteed to converge to a local minimum.
3. _____(CDF) is a method to describe the distribution of random variables.
4. In kernelized SVMs, the _____ K has to be positive semi-definite.

True or False:

1. A cumulative distribution function (CDF) cannot be less than 0 or bigger than 1. True/False
2. K-Means Clustering is guaranteed to converge (i.e., terminate). True/False
3. Nearest neighbors is a parametric method. True/False

Checkpoint solutions (2 of 2)

Fill in the blanks:

1. In k nearest neighbors, k=1 increases the **complexities**.
2. K-means **clustering** is guaranteed to converge to a local minimum.
3. **Cumulative Distribution Function (CDF)** is a method to describe the distribution of random variables.
4. In kernelized SVMs, the **kernel matrix** K has to be positive semi-definite.

True or False:

1. Acumulative distribution function (CDF) cannot be less than 0 or bigger than 1. **True**
2. K-Means Clustering is guaranteed to converge (i.e., terminate). **True**
3. Nearest neighbors is a parametric method. **False**

Question bank

Two mark questions:

1. What is overfitting, and how can you avoid it?
2. What is ‘training set’ and ‘test set’ in a machine learning model? How much data will you allocate for your training, validation, and test sets?
3. How do you handle missing or corrupted data in a dataset?
4. What is semi-supervised machine learning?

Four mark questions:

1. How can you choose a classifier based on a training set data size?
2. Describe the confusion matrix with respect to machine learning algorithms.
3. What is a false positive and false negative and how are they significant?
4. What are the three stages of building a model in machine learning?

Eight mark questions:

1. What is deep learning in modern businesses?
2. What are the applications of supervised machine learning in modern businesses?

Unit summary

Having completed this unit, you should be able to:

- Understand the network intrusion detection.
- Gain knowledge on anomaly detection in big data.
- Understand anomaly detection for autonomous robots.