

A close-up photograph of a person's hands holding a white smartphone. The person is wearing a white button-down shirt. The background is a light, neutral color with some circular patterns. A bright blue horizontal bar is overlaid on the right side of the image, containing white text.

# Anomaly detection in Telecommunications

Valentina Đorđević  
Data Scientist @Things Solver  
[valentina@thingsolver.com](mailto:valentina@thingsolver.com)

# Content

01	Defining the term “anomaly”	<ul style="list-style-type: none"><li>• What are anomalies</li><li>• Types of anomalies</li><li>• Detecting anomalies in the data</li></ul>
02	Problem definition	<ul style="list-style-type: none"><li>• Main concepts &amp; data explanation</li><li>• User requirements definition</li><li>• Determine project objectives</li></ul>
03	Isolation Forest algorithm	<ul style="list-style-type: none"><li>• Explaining the concept of instance “isolation”</li><li>• Phases and steps of IF algorithm</li><li>• Determining the anomaly score</li></ul>
04	Autoencoder neural network	<ul style="list-style-type: none"><li>• Explaining the concepts of autoencoder neural network<ul style="list-style-type: none"><li>◦ Encoder part</li><li>◦ Decoder part</li></ul></li><li>• Determining the anomaly score</li></ul>
05	Comparative analysis of the applied techniques	<ul style="list-style-type: none"><li>• Evaluation of applied models</li><li>• Pros ‘n’ cons</li></ul>



Anomaly detection

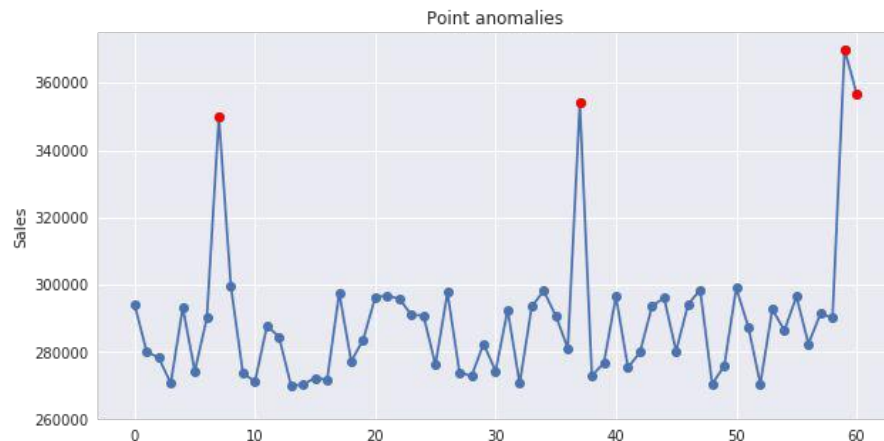
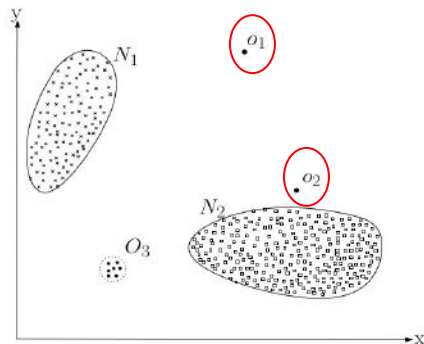
# Anomalies in the data.

- ❑ Anomaly represents the type of behaviour in the data that differs significantly from some expected behaviour.
  
- ❑ Anomaly != Outlier
  
- ❑ Types of anomalies:
  1. Point anomalies
  
  2. Contextual anomalies
  
  3. Collective anomalies

# Point anomaly.

**Point anomaly** is an instance that could be considered as anomalous among other instances in the dataset.

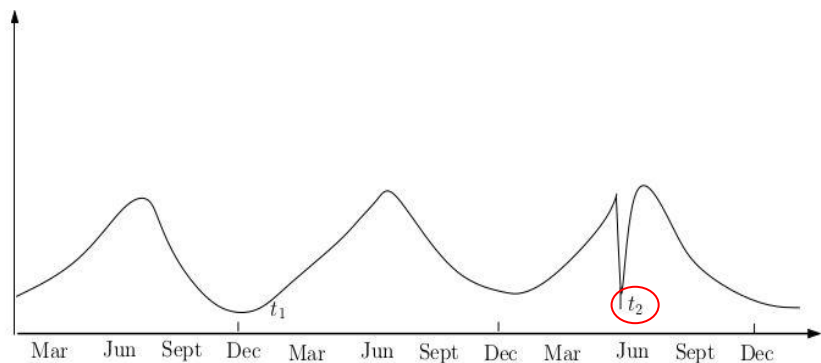
Point anomalies often represent some extremum, irregularity or deviation that happens randomly and have no particular meaning.



# Contextual anomaly

**Contextual anomaly** is an instance that could be considered as anomalous in some specific context.

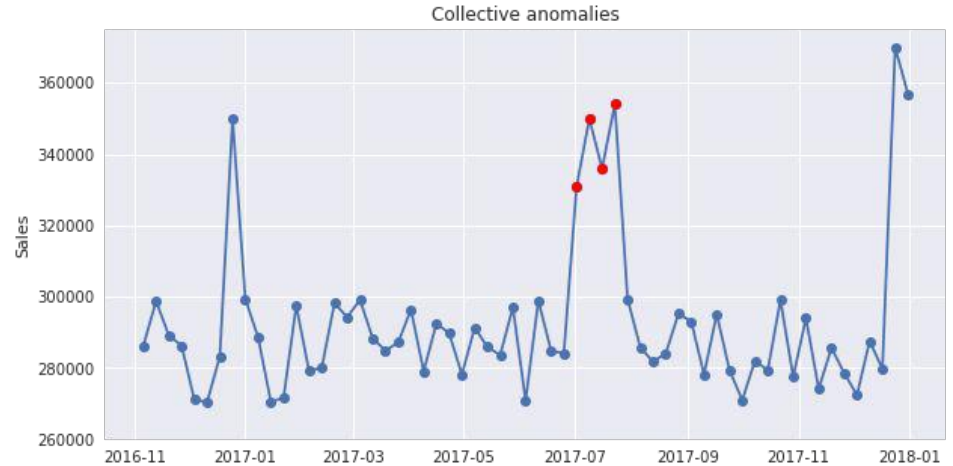
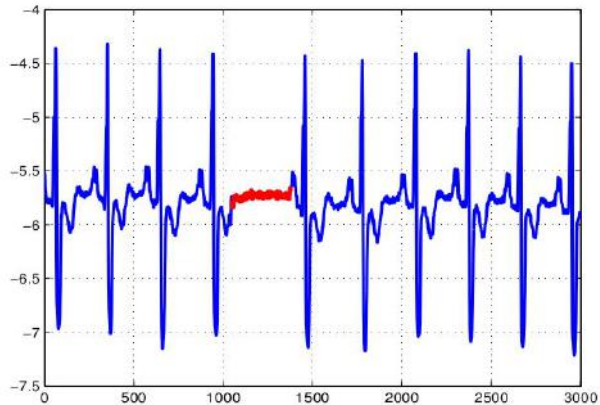
The contextual anomaly is determined by combining contextual and behavioural features, like space and/or time with some quantitative measurement (total money spent, average temperature, average end user throughput,...)



# Collective anomaly.

**Collective anomaly** is often represented as a group of correlated, interconnected or sequential instances.

While each particular instance of this group doesn't have to be anomalous itself, their collective occurrence is anomalous.



# Anomaly detection techniques

Supervised anomaly detection.

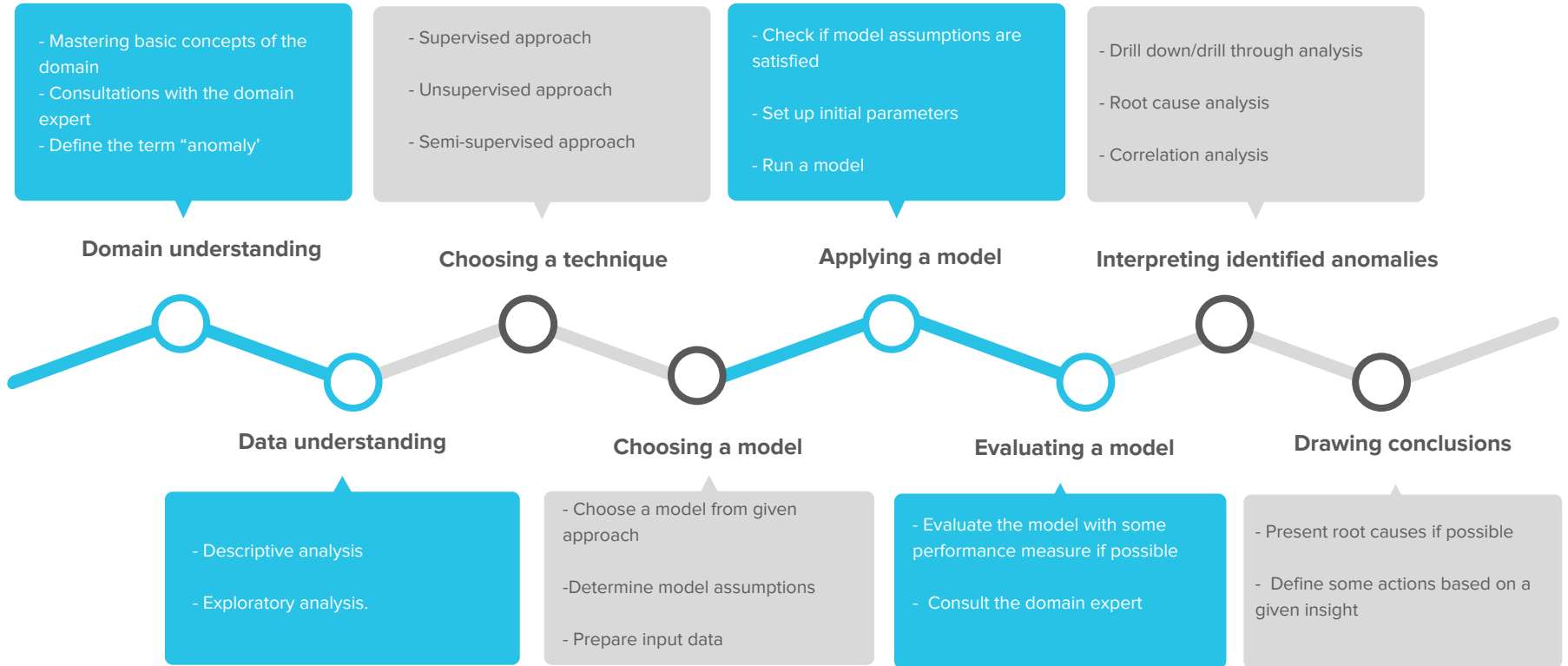
Unsupervised anomaly detection.

Semi-supervised anomaly detection.





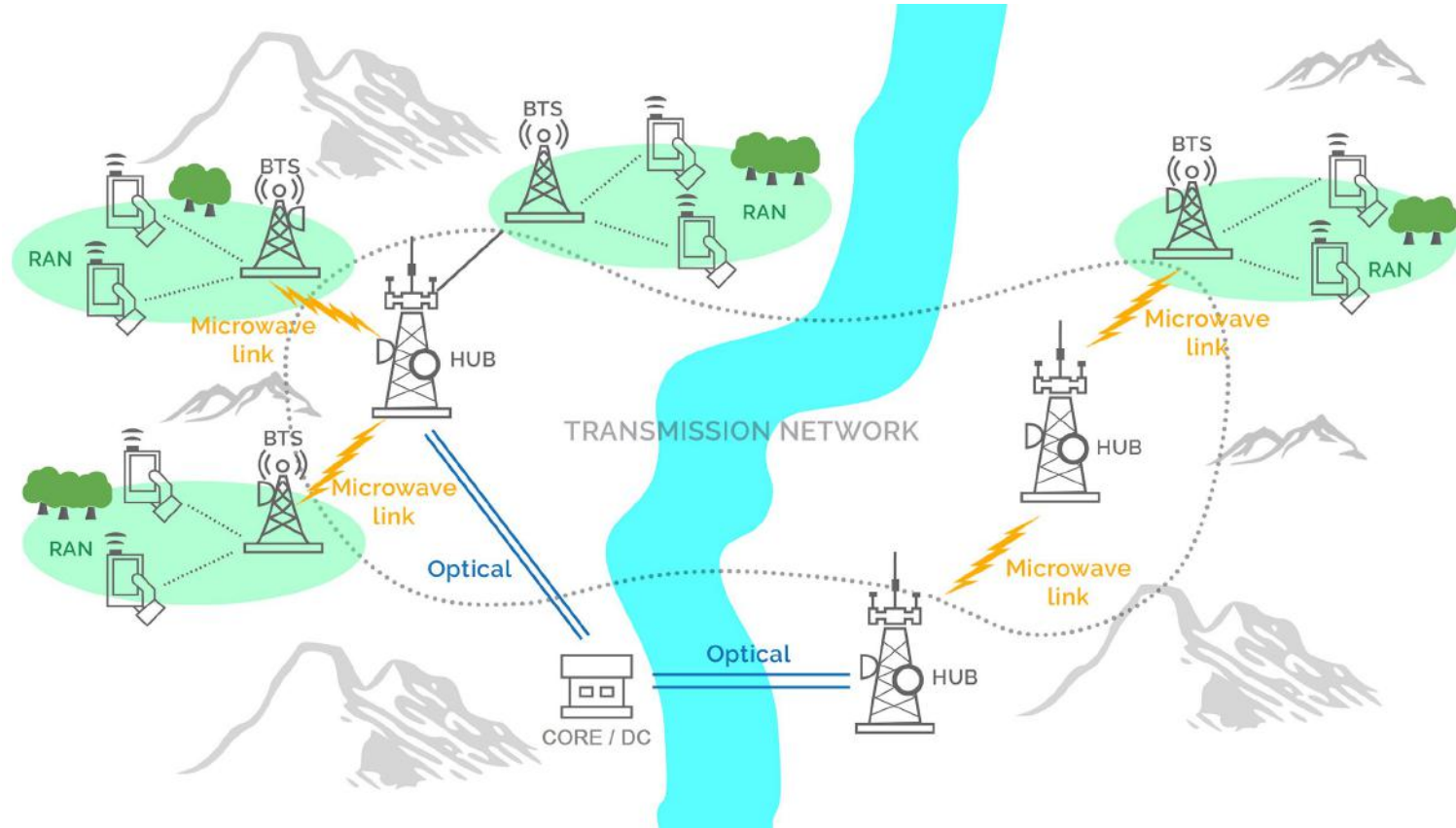
# The process of anomaly detection





Telecommunication network

# Telecommunication network.





Problem definition

# Problem definition.

“Understanding the **project objectives** and **requirements** from a domain perspective and then converting this knowledge into a **data mining problem** definition with a **preliminary plan** designed to achieve the objective.” (CRISP-DM)

- ❑ Site behaviour through time (100+ KPIs)
  - ❑ packet loss
  - ❑ packet delay
  - ❑ transmission success rate
  - ❑ ...
- ❑ Labels not available
- ❑ Based on the available KPIs, **automatically identify anomalous sites** in a given period of time

Technique to be used: **unsupervised anomaly detection**

Available models : **Isolation Forest / Autoencoders**





Benefits

# 24/7

Health checks



Enables continuous checks over the network parameters, and fires a signal when anomalous behaviour is detected.

# Efficient

Problem diagnosis



Contains information which KPIs have been flagged as anomalous, and in which point of time.

# Process

Automation



Helps in process automation, 100+ KPIs are simultaneously analysed, to automatically detect and report anomalous behaviour.

An aerial photograph of a dense forest, likely a coniferous forest, showing a variety of green shades and some dead, grey trees. A solid cyan horizontal banner is overlaid across the middle of the image. The text "Isolation Forest" is written in white, sans-serif font on the banner.

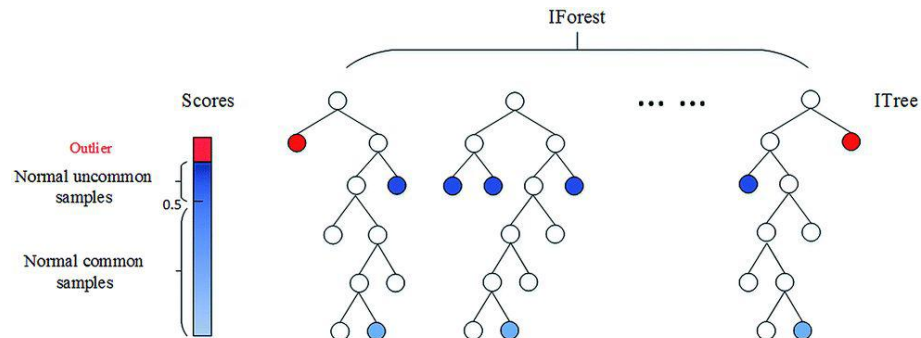
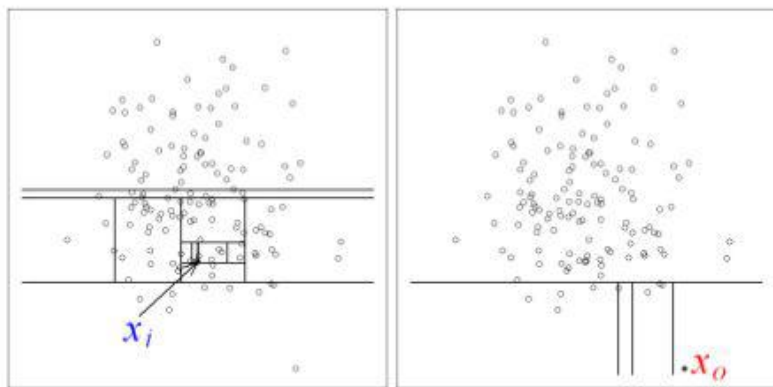
# Isolation Forest



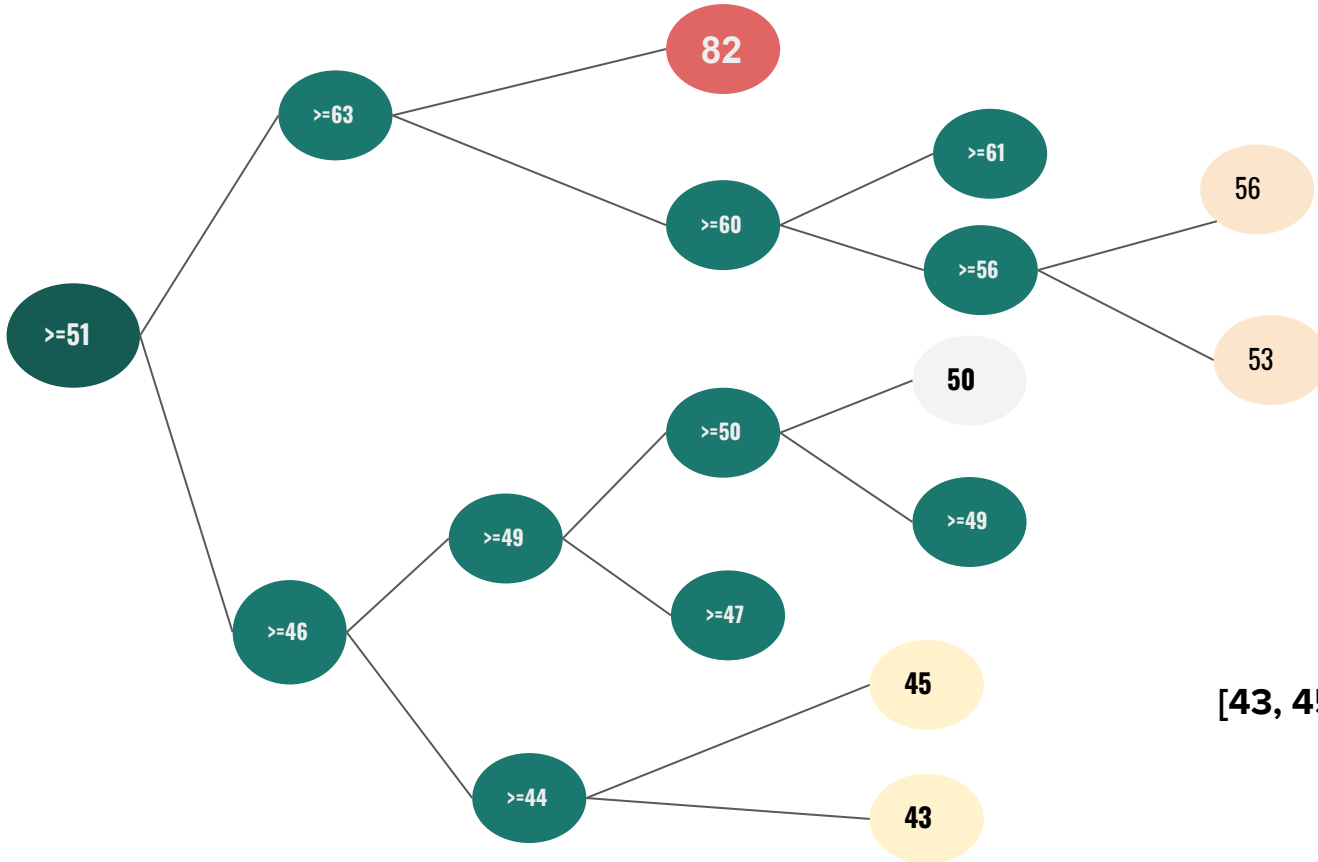
# Isolation Forest algorithm.

- ❑ tries to “isolate” an instance
- ❑ the isolation is performed by building an ensemble of trees
- ❑ trees are built by recursively selecting a random feature and performing random partitioning, until an instance is positioned in a terminal node, which means it is isolated
- ❑ a smaller number of partitions needed to isolate an instance indicates higher chances of anomaly
- ❑ Anomaly score is determined by the average path length to the terminal node a given instance has been placed into.

$x_o$  is more likely to be an anomaly than  $x_i$



# Isolation Forest - example.



[43, 45, 50, 50, 53, 53, 56, 82]



# Autoencoders

# Autoencoders.

Autoencoder is a unsupervised neural network. In most cases, the main goal of training an autoencoder is to provide an output that is the same as input.

They work by compressing the input into a latent-space representation, and then reconstructing the output from this representation. It is consisted of two parts: encoder and decoder.

Encoder is used for mapping from the original space to another space, of possibly higher or lower dimensionality.

Decoder is used for mapping back from this new space to the original one.

Anomaly score is determined by the error a decoder made in reconstruction phase.

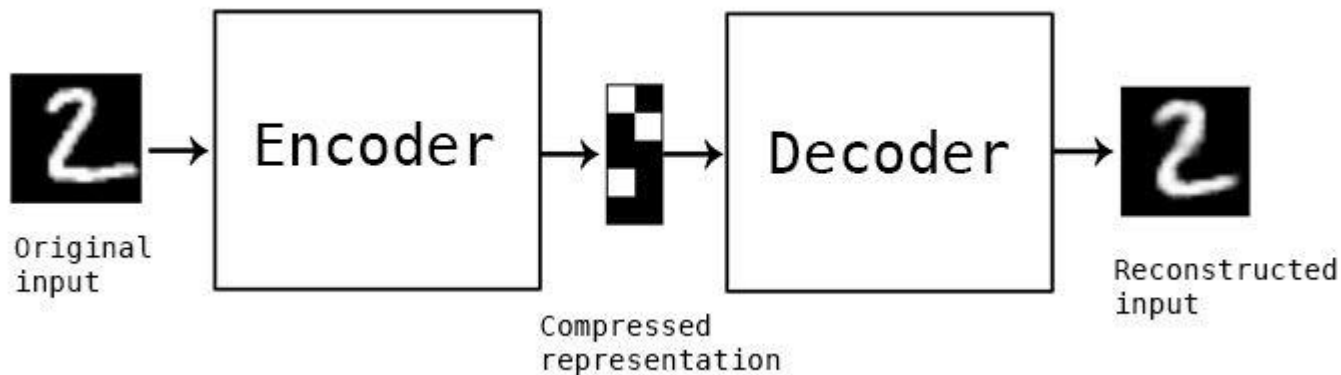
# Autoencoder - characteristics.

Compression and decompression functions are:

- ❑ Data-specific
- ❑ Lossy
- ❑ Automatically-learned

... which makes it pretty convenient for:

- ❑ Data denoising
- ❑ Dimensionality reduction
- ❑ Anomaly detection





# Comparative analysis

# Comparative analysis

		Isolation Forest	Autoencoder
1	Accuracy	72%	69%
2	Pros	<ul style="list-style-type: none"><li>- Very fast</li><li>- Good performances with redundant data</li><li>- Can work in both supervised and unsupervised mode</li></ul>	<ul style="list-style-type: none"><li>- Possible to extract error per each dimension</li><li>- Good for catching nonlinear dependencies</li><li>- Convenient for noise reduction</li></ul>
3	Cons	<ul style="list-style-type: none"><li>- Not possible to extract path lengths for individual dimensions</li><li>- Not possible to visualize i-trees</li><li>- Could have low performance when working with non or slightly deviant features</li></ul>	<ul style="list-style-type: none"><li>- Takes too much time with high-dimensional data</li><li>- Learns to capture as much information as possible rather than as much relevant information as possible</li></ul>

A top-down view of a white notebook with lined pages, a rose gold pen, and a white pen with gold dots, resting on a white surface. A blue horizontal bar is overlaid on the right side of the image.

# Q&A

Valentina Đorđević  
Data Scientist @Things Solver  
[valentina@thingsolver.com](mailto:valentina@thingsolver.com)