Data-driven Predictive Maintenance in Industry 4.0: Concepts and Challenges

Rita P. Ribeiro - rpribeiro@fc.up.pt

University of Porto and INESC TEC
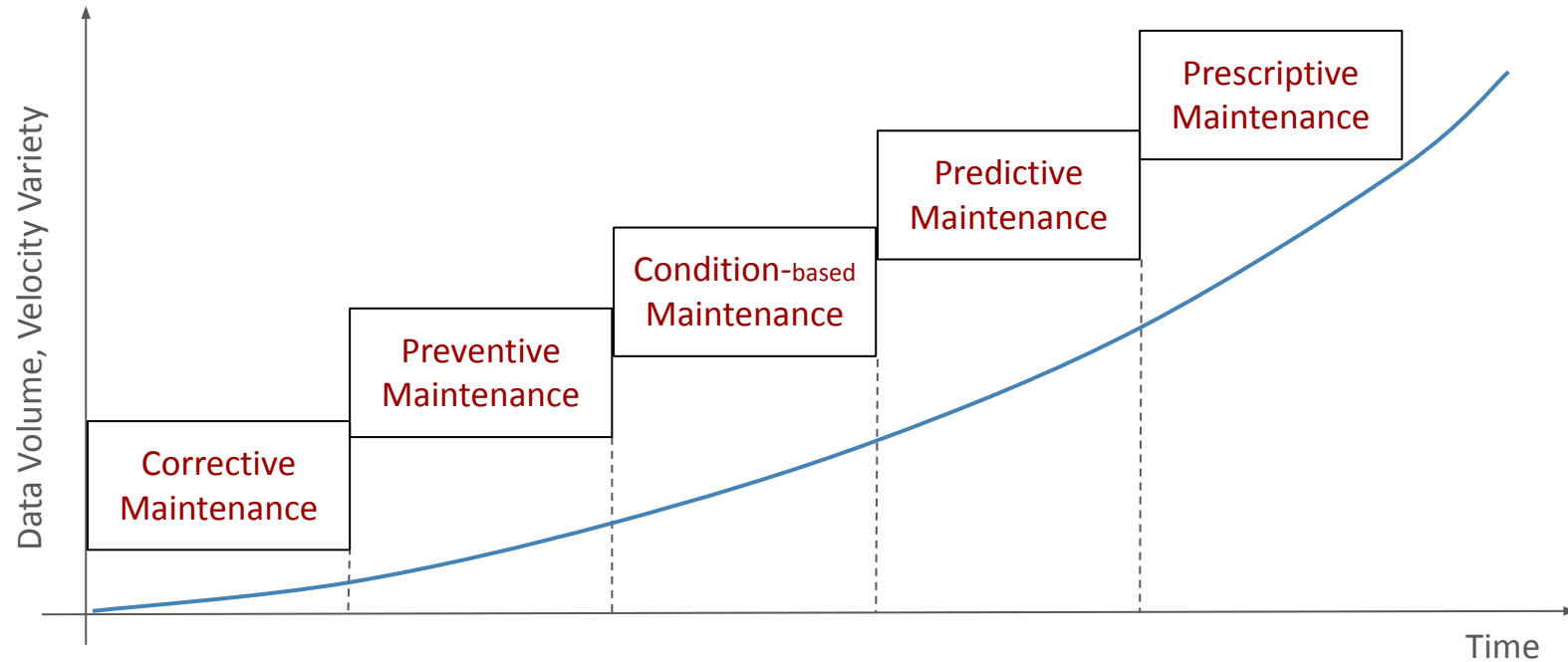
U. PORTO   INESCTEC

Source: bossar.com

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023

Maintenance Paradigms

Maintenance Paradigms

Maintenance	Corrective	Preventive	Condition-based	Predictive	Prescriptive
schedule	non-fixed	fixed	non-fixed	non-fixed	non-fixed
repair actions	when it breaks	on predefined time intervals	based on degradation level	based on early-stages of evolving failures	based on early-stages of evolving failures
effectiveness	technician's experience	technician's experience	technician's experience and monitoring data	monitoring data	monitoring data and maintenance systems
costs / savings	downtime repairs	possible unnecessary repairs	repair when necessary	repair when necessary	optimum repair timing

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023

Predictive Maintenance: Vision and Goals

- *"Techniques designed to determine the condition of equipment and estimate when to do maintenance"*
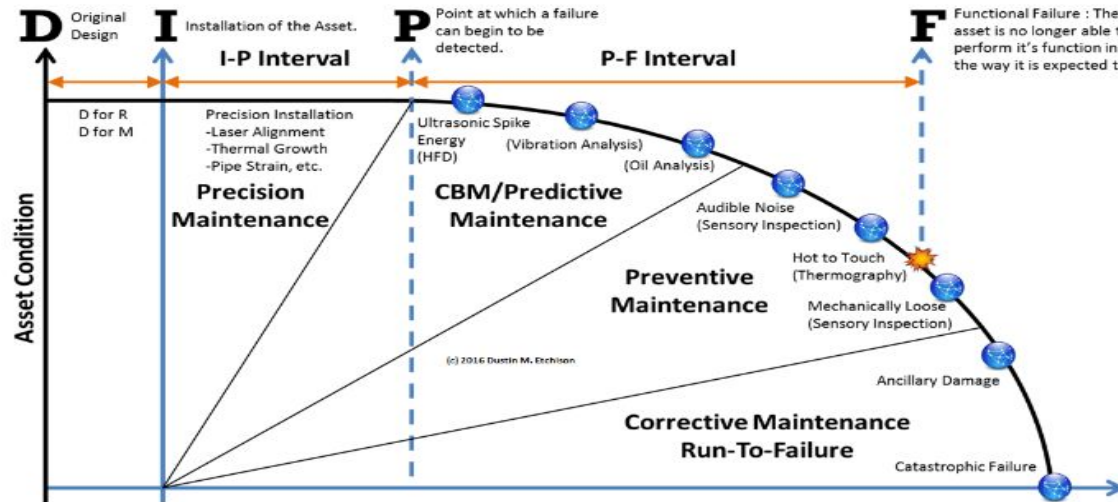
Source: production-technology.org

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023

Predictive Maintenance: Vision and Goals

Asset Condition

- **Health indicators** are quantifiable characteristics of a population

- **Normal** or Healthy is acceptable/desired operation state

- **Fault** is an unpermitted deviation from the acceptable operating condition.

- **Failure** is a permanent interruption of a system ability to perform a required function.

- **Component degradation** or wear can be observed as a motion (in sensor space)

Predictive Maintenance: Vision and Goals

- **Predictive Maintenance (PdM):** important and active area in Industry 4.0
  - keep mechanical pieces working as long as possible
  - the goal is to ensure their normal functioning under any circumstances
  - avoid disruption of everyday operations or a chain reaction on other equipment.

- Predictive Maintenance (PdM) can resort to
  - Physics Model-based Approaches
  - Data-driven Approaches
  - Hybrid Approaches

Data-driven Predictive Maintenance

- IoT embedded in machines and production lines is now a reality.

- Sensors provide a vast set of information on a equipment

    - Vibration, Temperature, Pressure, Acoustic Signals, Optical Signals, Speed, etc

- Large-scale stream processing is needed for real-time data

- Very important for industries for competitive reasons

- Data-driven Predictive Maintenance (PdM) became one of the central answers

Data-driven Predictive Maintenance

Dynamic Data Sources:

- sensor data recording various signals (environmental, physical, electrical or mechanical)

- logs data recording sequence of status

- service data with details of performed services (equipments, costs, technicians)

Source: Wang, Z. (2015)

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023

Data-driven Predictive Maintenance

- Dynamically define when a machine is okay or needs to be maintained.

- Use of advanced statistical methods, such as machine learning.

- Look at patterns across all sensors and make one multivariate prediction model.

- The more data sources and data available the better are the predictions.

- Predictive maintenance models get better at predicting future breakdowns over time.

- Find complex indications for breakdowns that are nearly impossible for humans to spot.

Data-driven Predictive Maintenance

- Equipment is becoming more and more complex
  - We need better diagnostic and fault detection systems

- Building specialised diagnostic functions is expensive
  - It is unrealistic to expect that for each individual fault

- Real usage data shows actual problems to look for
  - Sometimes what is "known" turns out to be wrong

- What is important to focus on changes with age
  - We need continuous monitoring and lifelong learning

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023

Data-driven Predictive Maintenance

Objectives:

- minimize maintenance costs
- reduce unplanned downtime
- avoid cost for failure recovery
- avoid consequential costs for unplanned stops

Opportunities:

- shorter workshop visits
- better planning and diagnostics
- reduce number of incorrect repairs
- optimization of maintenance intervals

Source: upkeep.com

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023
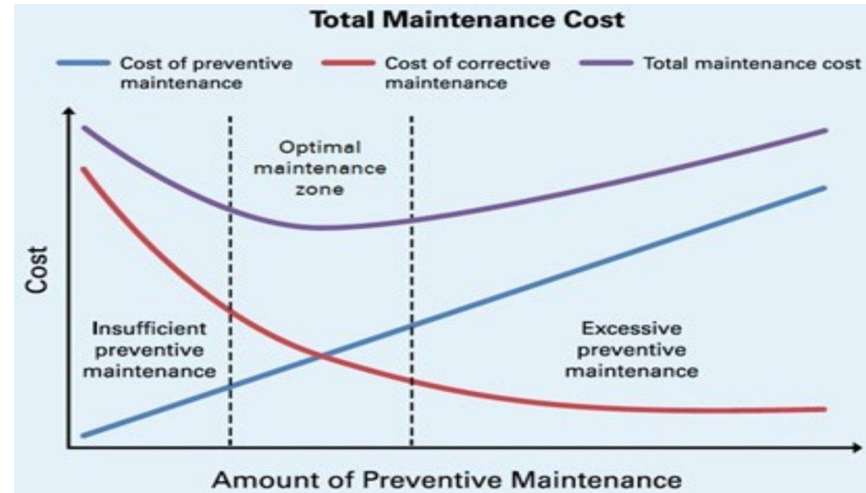
Some Application Domains

- **Energy Area**

    - Failure prediction of connected heating systems.

    - Anomaly detection in electricity meters.

- **Industry Area**

    - Predictive maintenance of complex equipments in high-tech manufacturing.

    - Failure prediction to minimize environmental hazards caused by equipment that performs extraction and refining of oil and gas.

- **Transport Area**

    - Monitoring sensor data from planes to increase passengers safety (e.g. turbofan engine).

    - Prediction of breakdowns in public transports to avoid trip cancellation

Domain Challenges

What data to collect, so to identify the suitable equipment condition indicators?

- Internal conditions
    - many components that build up a complex system that is the equipment

- External conditions

    - affect the functioning of the equipment (weather, geographical position)
    - but, equipment context is sometimes unknown
    - external conditions and usage mixed with fault symptoms

- Low-quality data

    - tangential to the relevant processes
   - low measurement accuracy and frequency

Domain Challenges

What data to collect, so to identify the suitable equipment condition indicators?

- Collecting large amounts of (reasonably reliable) normal operation data is (relatively) easy

- However, this data is usually far from being fully representative!

- We need faulty data!

  ○ Failures are usually rare events that take time to develop
  ○ Run-to-failure is expensive, and diagnostics is difficult
  ○ Fault injection does not embrace all situations
  ○ It is difficult to know all the possible types of faults

Domain Challenges

What data to collect, so to identify the suitable health indicators?

- Even if we gather faulty data …

- The appropriate time window to consider, depends on the domain

  - width
    - how far back is data relevant?

  - increment
    - how often should be learned new model?

  - position
    - what is latest time and duration of human intervention?
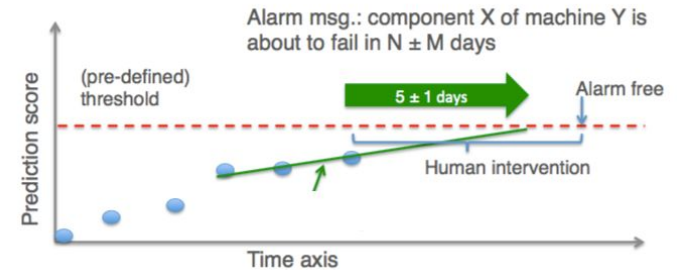
Wang, Z. (2015)

Machine Learning Challenges

How to create meaningful features from raw data?

Data acquisition

- raw sensor data
- high dimensionality
- low quality
    - missing values
    - outliers
    - constant or low variance values
    - null values
- feature engineering
    - temporal aggregation
    - capture the pattern of the functioning of the equipment

Features

Machine Learning Challenges

What about the target variable?

- all observations have a target value that indicates the equipment condition

    - it is not always easy to obtain; it may be dependent on the human expert

- some observations report the "normal" functioning of the equipment

    - it is difficult to collect every possible normal operating conditions

- none observations have information on the type of operational condition

   - every outlier can be considered a failure

Machine Learning Challenges

What configures a failure?

- The abnormal functioning of an equipment represents an anomaly

  - few or none examples of such event on the systems past

- One cannot always rely on historical fault data

  - new versions are deployed in new conditions

- How can one discover new, unexpected issues?

  - analysing "normal" behaviour, and detecting deviations
  - not every change in behaviour is a problem

- Inherent challenge to interpret discovered anomalies

Machine Learning Challenges

Why is a difficult learning task?

- Rare cases correspond to failures

- But these are the most relevant for the domain expert

- Imbalanced domains pose problems to standard learning tasks

- Imbalanced Domain Learning configures two assumptions

  - the representativeness of the cases on the training data is not uniform;

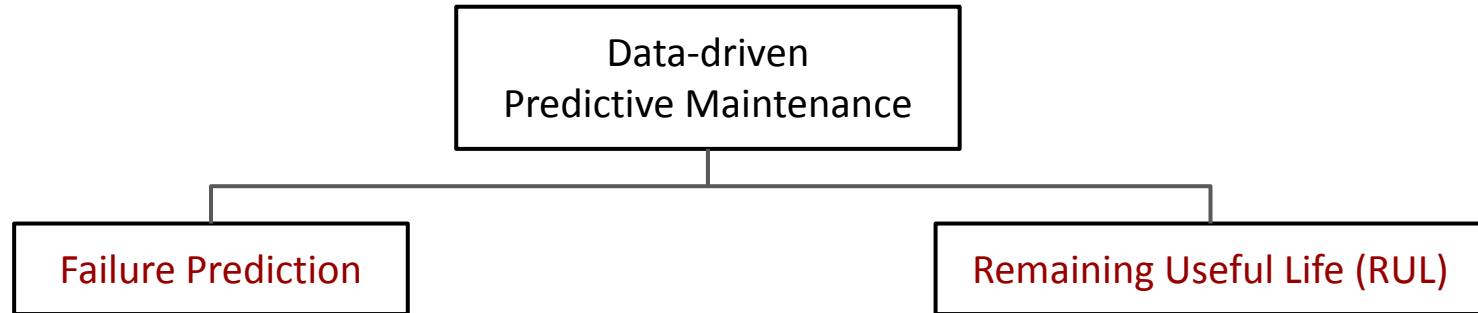 - the underrepresented cases are the most relevant ones for the domain.
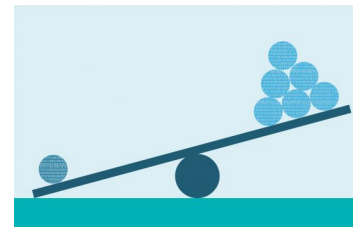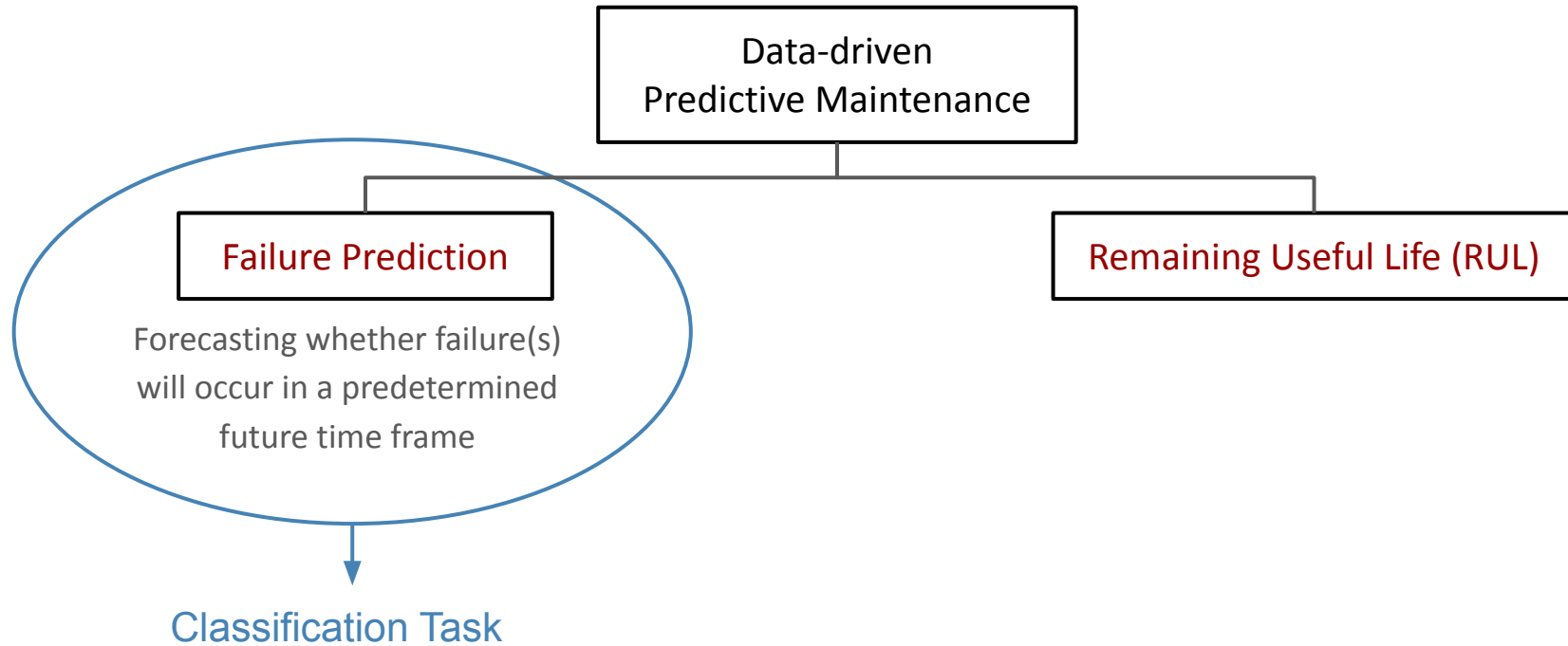
https://datascience.aero/

Machine Learning Challenges

Why is a difficult learning task?

- Rare cases correspond to failures and are relevant

- Definition of failures is dependent on the application domain knowledge

- Define every possible normal / abnormal behaviour is hard.

- Boundary normal / abnormal behaviour is often not precise.

- Hard to distinguish real evolving failures from just random noise in data

- The abnormal behaviour will evolve with time.

- Inherent lack of known labeled failures for training/validation of models.
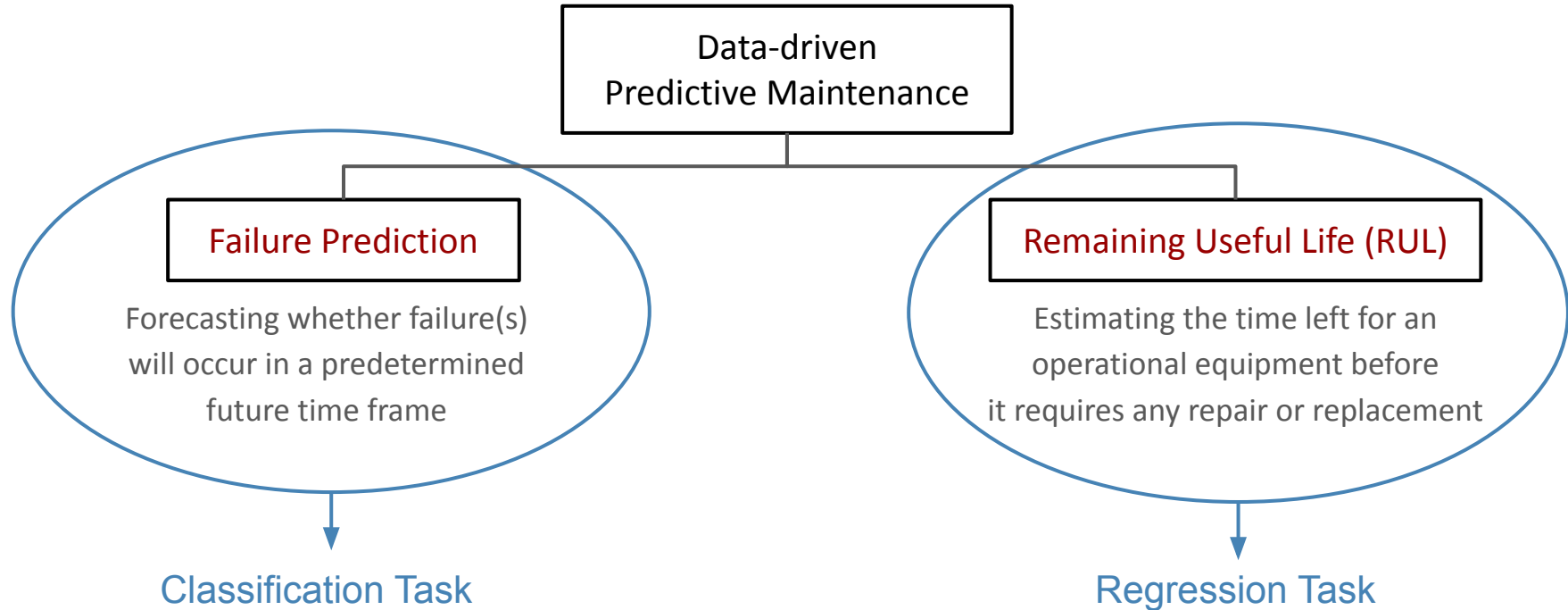
Predictive Learning Tasks

```
                    ┌─────────────────────────┐
                    │      Data-driven        │
                    │  Predictive Maintenance │
                    └─────────────────────────┘
                                 │
              ┌──────────────────┴──────────────────┐
   ┌──────────────────┐                  ┌──────────────────────────┐
   │ Failure Prediction │                │ Remaining Useful Life (RUL) │
   └──────────────────┘                  └──────────────────────────┘
```

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023

Predictive Learning Tasks

Data-driven Predictive Maintenance

Failure Prediction

Forecasting whether failure(s) will occur in a predetermined future time frame

Classification Task

Remaining Useful Life (RUL)

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023

Predictive Learning Tasks

Data-driven
Predictive Maintenance

Failure Prediction

Forecasting whether failure(s)
will occur in a predetermined
future time frame

Classification Task

Remaining Useful Life (RUL)

Estimating the time left for an
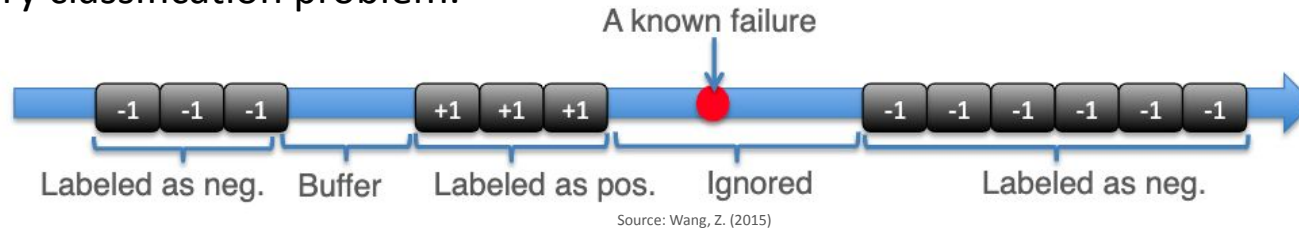operational equipment before
it requires any repair or replacement

Regression Task

Performance Assessment for Failure Prediction

Goal:

- Predict whether a failure will occur in a predetermined future time frame
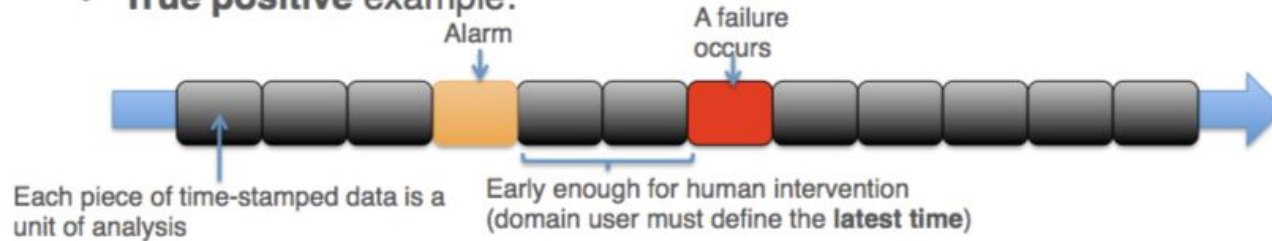
- Binary classification problem.

A known failure

-1 -1 -1   +1 +1 +1   ●   -1 -1 -1 -1 -1 -1

Labeled as neg.   Buffer   Labeled as pos.   Ignored   Labeled as neg.

Source: Wang, Z. (2015)

- An alarm is a positive prediction that a failure is about to happen.

- How effective is the prediction model?

	Alarm	**No Alarm**
Failure	TP	FN
Normal	FP	TN

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023

Performance Assessment for Failure Prediction

What is a True Positive?

- **True positive** example:

Alarm

A failure occurs

Each piece of time-stamped data is a unit of analysis

Early enough for human intervention (domain user must define the **latest time**)

Source: Wang, Z. (2015)

What is a False Positive?

- **False positive** example:

Too early to be treated success (domain user must define the **earliest time**)

Source: Wang, Z. (2015)

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023
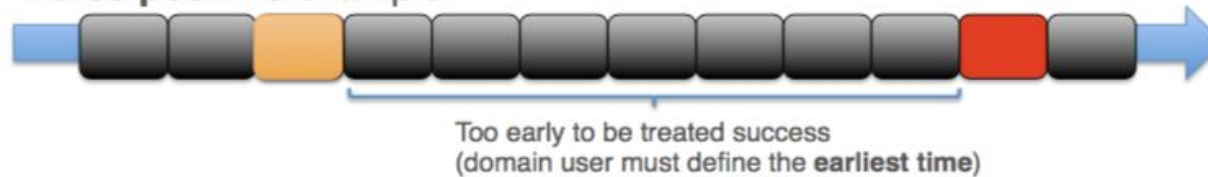
Performance Assessment for Failure Prediction

What is not a True Positive nor a False Positive?

- **Invalid** prediction (neither TP nor FP):
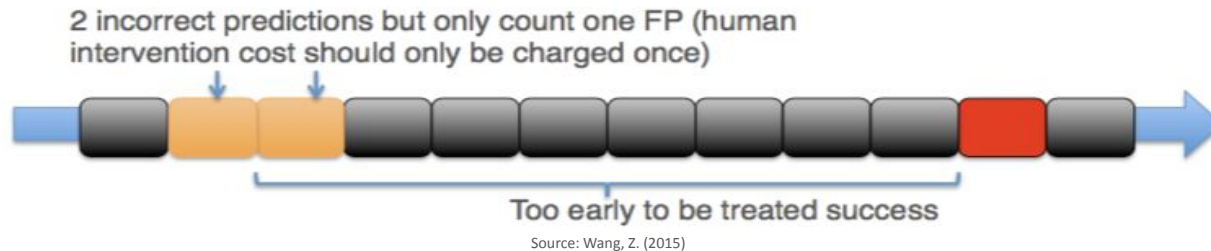
Algorithm captures a failure but too late for human intervention

Alarms should be muted in human intervention period

Source: Wang, Z. (2015)

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023
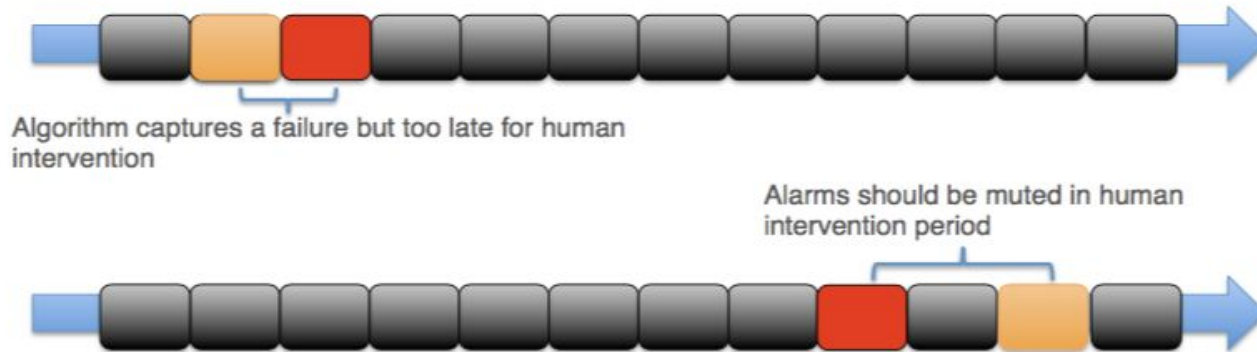
Performance Assessment for Failure Prediction

What should count as one True Positive?

3 correct predictions but only count one TP (one predicted failure should only be rewarded once)

A failure occurs

Enough time for human intervention

Source: Wang, Z. (2015)
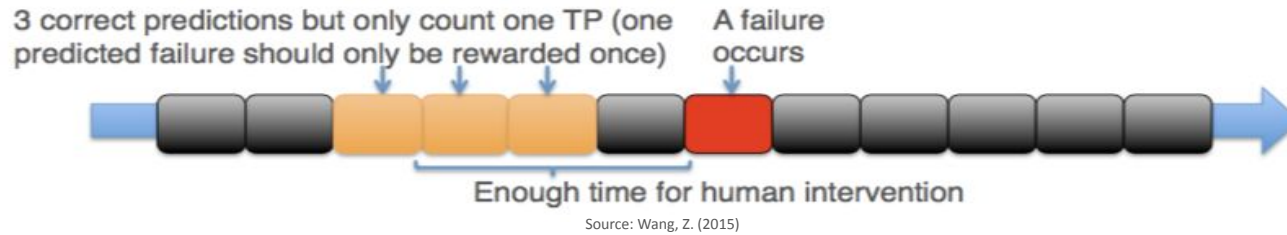
What should count as one False Positive?

2 incorrect predictions but only count one FP (human intervention cost should only be charged once)

Too early to be treated success

Source: Wang, Z. (2015)

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023

Performance Assessment for Failure Prediction

- Standard performance metrics (e.g. accuracy, error rate)
    - assume that all instances are equally relevant for the model performance

- Good performance estimation in models
    - that perform well on normal (frequent) cases and bad on outlier (rare) cases.

- Example

    - Dataset with 1% of labelled failures;
    - Model M predicts all cases as non-failures;
    - M has a estimated accuracy of 99%;
    - Yet, all the failures were missed!

Performance Assessment for Failure Prediction

- The focus is on a small subset of cases (e.g. failures)

- The average is not a good idea

- Standard metrics influenced by the performance on the most common cases

  - i.e. normal functioning of equipment

- Not useful for failure prediction

- What other performance metrics can we use?
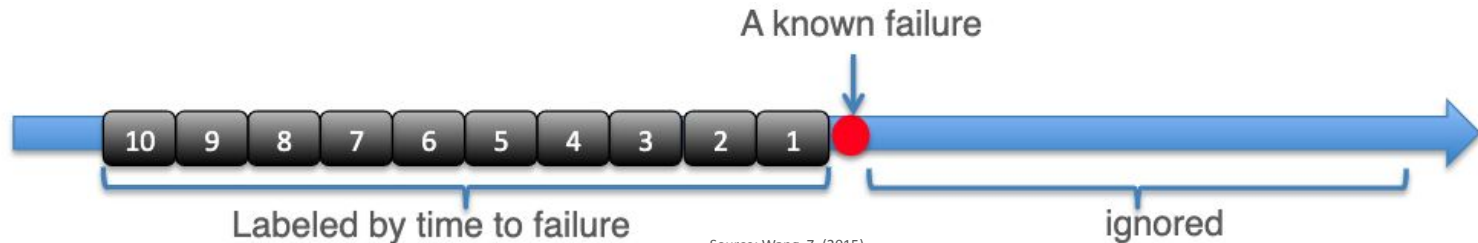
Performance Assessment for Failure Prediction

How effective is the failure prediction model?

- If we have domain information, we can use cost-sensitive metrics

  - What is the saving of capturing a failure in time to repair (TP)?

  - What is the inspection cost or the cost of a false alarm (FP)?

  - What is the replacement cost ot the cost of missing a failure (FN)?

- If no domain information is available

  - e.g. Precision, Recall (TPR), False Alarm Rate (FNR), Impostor Pass Rate (FPR)
  - e.g. F-measure, AUC-ROC, AUC-PR

Performance Assessment for Remaining Useful Life (RUL)

Goal:

- Predict the time left until the end of the useful life of the equipement
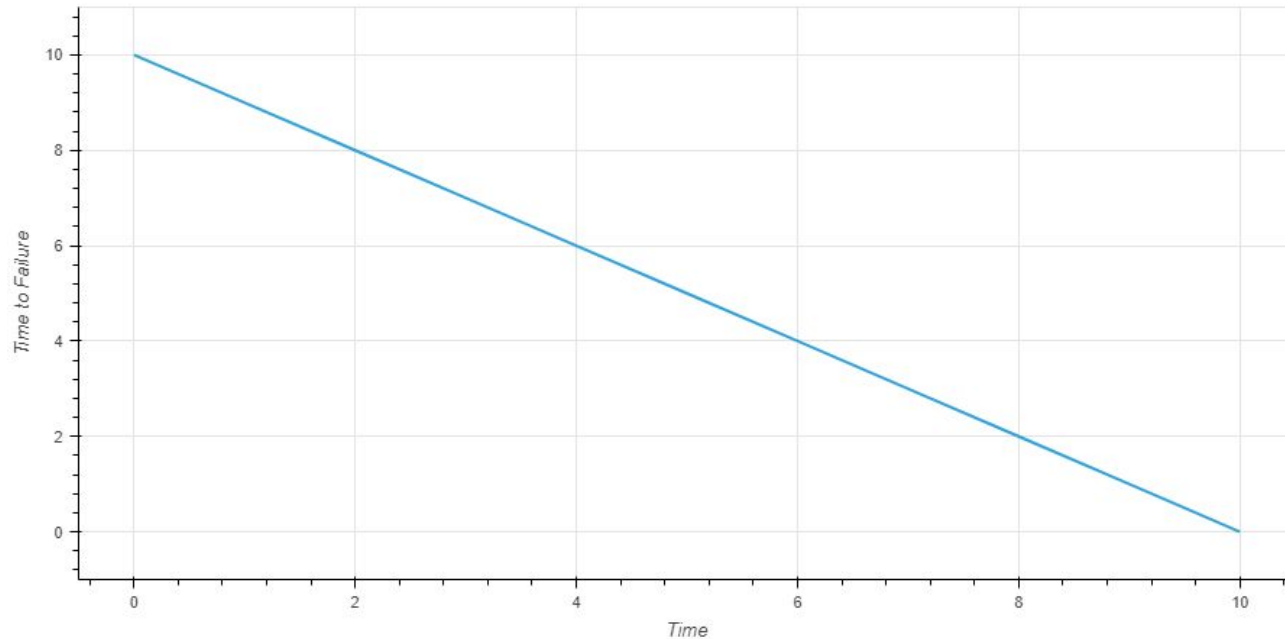
- Regression problem

Source: Wang, Z. (2015)

- The model uses more fine-grained information, not an abrupt change
  - e.g. gradual degradation of components

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023

Performance Assessment for Remaining Useful Life (RUL)

● **What is the Time-to-Failure?**

Source: Nowaczyk, S. (2021)

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023

Performance Assessment for Remaining Useful Life (RUL)

- ## What is the shape of RUL curves?
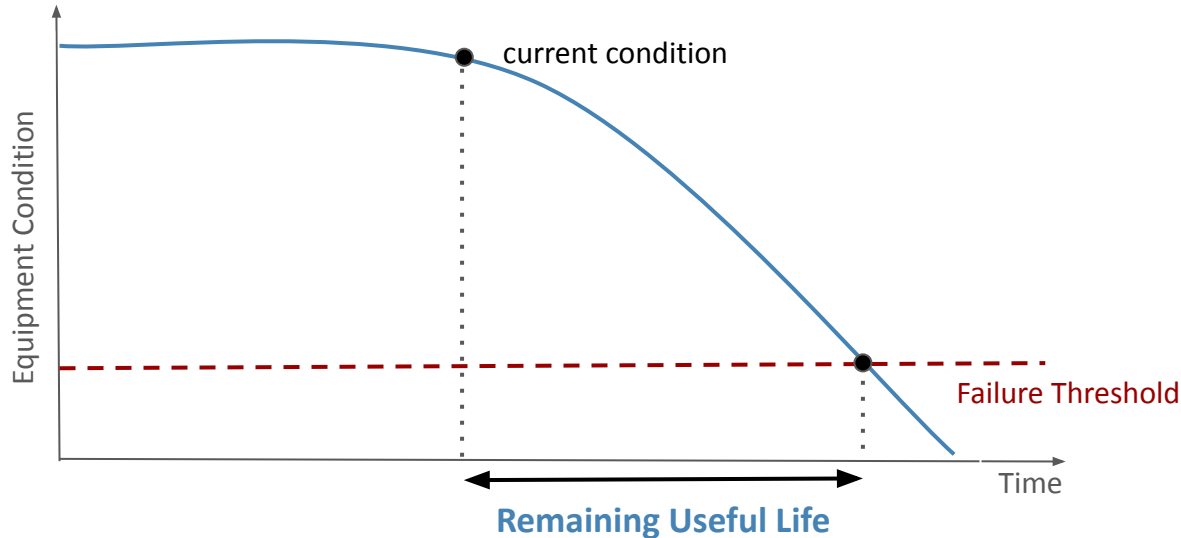
- ## Failure Threshold is the minimum acceptable condition estimated by

  - ### End-of-Life (EoL), MeanTime-to-Failure (MTTF) or Mean Time Between Failure (MTBF)

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023

Performance Assessment for Remaining Useful Life (RUL)

How effective is the RUL model?

- Standard Error Metrics
    - e.g. MAE, MSE, RMSE, NMSE, NRMSE, $R^{2,}$ MAPE

- But, once more, there is the question

    - are standard metrics appropriated for the requirements of this predictive task?

- Performance tends to be more critical as the system nears its end-of-life (EoL).

Performance Assessment for Remaining Useful Life (RUL)

How effective is the RUL model?

- A simple aggregate of performance is not a fair representative of overall performance

- Performance at specific times relative to the EoL can be a reasonable alternative.

- Criticality of predictions at different stages may be ranked differently.

- A robust metric should be capable of making an assessment at all stages.

- Some of metrics aim to alleviate these issues in evaluating prognostic performance

Performance Assessment for Remaining Useful Life (RUL)

How effective is the RUL model?

- **Exponential Transformed Accuracy (ETA)** - Nectoux et al (2021)

    - different hazard severity for under- and over-estimate of RUL

    - over-estimates are more penalized as can lead to more severe damage

    - higher error, worst RUL prediction.

    - ETA varies between 0 and 1, higher is better
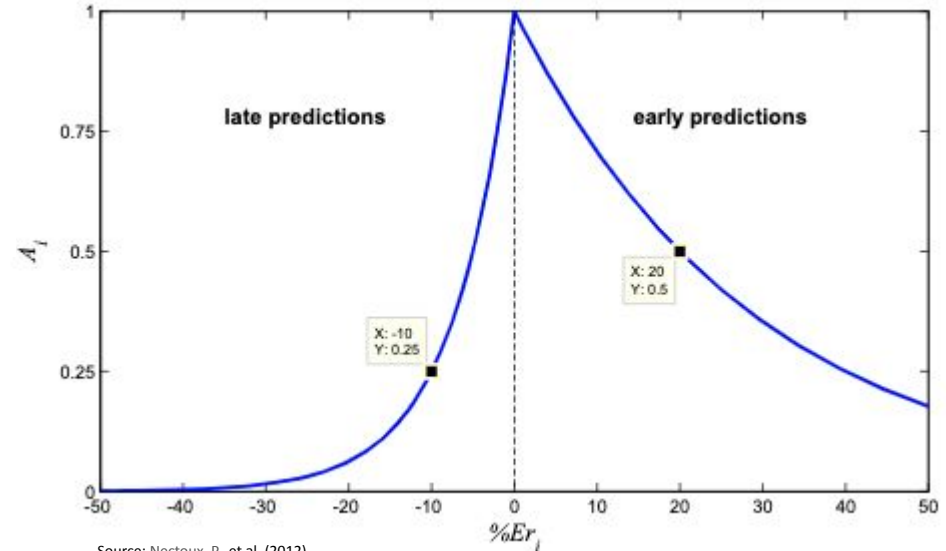
    - more than one version exists …

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023

Performance Assessment for Remaining Useful Life (RUL)

How effective is the RUL model?

- **Exponential Transformed Accuracy (ETA)**

$$A_i = \begin{cases} exp^{-ln(0.5).(Er_i/5)} & \text{if } Er_i \leq 0 \\ exp^{+ln(0.5).(Er_i/20)} & \text{if } Er_i > 0 \end{cases}$$

$$\%Er_i = 100 \times \frac{ActRUL_i - \widehat{RUL_i}}{ActRUL_i}$$

late predictions   early predictions

X: -10
Y: 0.25

X: 20
Y: 0.5

Source: Nectoux, P. et al. (2012)

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023

Performance Assessment for Remaining Useful Life (RUL)

How effective is the RUL model?

- Prognostic Horizon (PH)

    - test if the predictions satisfy a given prognostic horizon;
    - difference when the prediction results first satisfy a specified criterion and EoL

- The PH may be specified by an allowable error bound (α) around the true EoL.

$$\text{PH}^{\alpha} \quad = \quad i^{EoL} - i^{\alpha} \quad \text{where} \quad i^{\alpha} = \min\{i : |y_i - f_i| \leq \alpha\}$$

- The choice of α depends on the estimate of time required to take a corrective action.

Performance Assessment for Remaining Useful Life (RUL)

How effective is the RUL model?

- If the model passes the PH Test …

- **α-λ accuracy** identifies if it performs:

  - within desired error margins (parameter α) of the actual RUL

  - at any given time instant (parameter λ) that may be of interest

- More metrics exist …
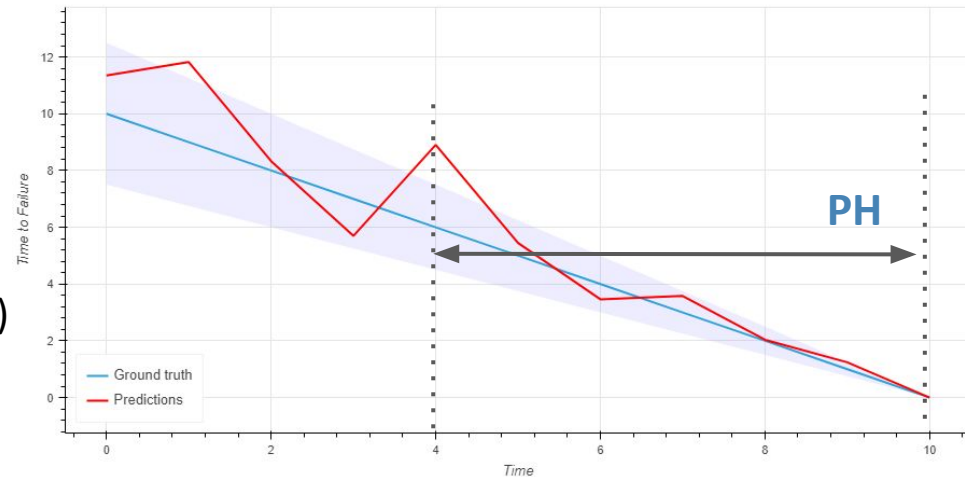
Source: Nowaczyk, S. (2021)

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023

Performance Assessment for Remaining Useful Life (RUL)

How effective is the RUL model?

- Remember: the goal is to help users make decisions
- Knowing the actual time to failure facilitates making detailed plans

- However, it can also be "too much information"
- In many situations, a binary "replace now" signal is better
- One can, of course, always turn RUL into "yes/no"

- Not all predictions are equally important
  - predictions closer to failure should be more precise
  - RUL over-estimation is more dangerous than under-estimation

Machine Learning Approaches

- Data acquisition
    - manual, expensive, but necessary

- Data preprocessing
    - boring, time-consuming and difficult

- Building a model
    - mostly automatic and reliable

- Evaluation & Exploitation
   - largely domain/scenario-specific

Machine Learning Approaches

- Data acquisition

    - manual, expensive, but necessary

- Data preprocessing

    - boring, time-consuming and difficult

Deep Learning can help!

- Building a model

    - mostly automatic and reliable

- Evaluation & Exploitation

    - largely domain/scenario-specific

Machine Learning Approaches

Why the need for Deep Learning?

- It allows for so called "end-to-end learning"
- Input raw data, arbitrarily complex, and get good results

- New paradigm: Big Data
- No need to "squeeze" every bit of useful clues from few data points

- Aim for unbounded complexity
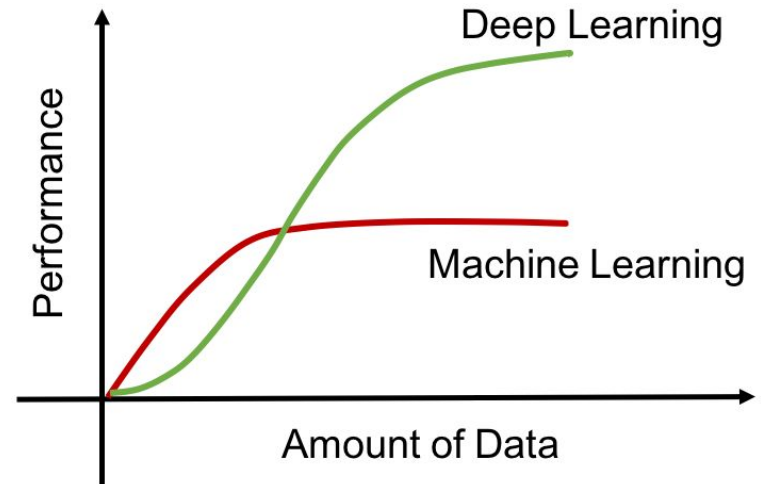
Deep Learning Approaches

- **Deep learning = Deep Neural Networks**
  - many layers
 - millions of neurons
  - efficient optimisation
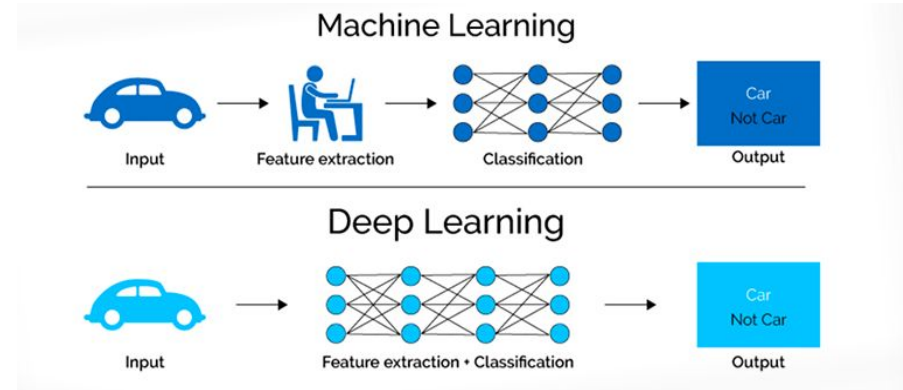
Source: ThinkWik

- **Deep Neural Networks (DNNs)**

  - can be directly applied to the raw signal, without computing first ad hoc features

  - learns a hierarchical representation of data and "creates" complex features

  - feature engineering is automatic!

Deep Learning Approaches

Deep Auto-encoders

- widely used in PdM

- motivated by large amount of data produced by sensors

- it learns to reconstruct the input

- it spots deviations from the input

- deviations can constitute failures
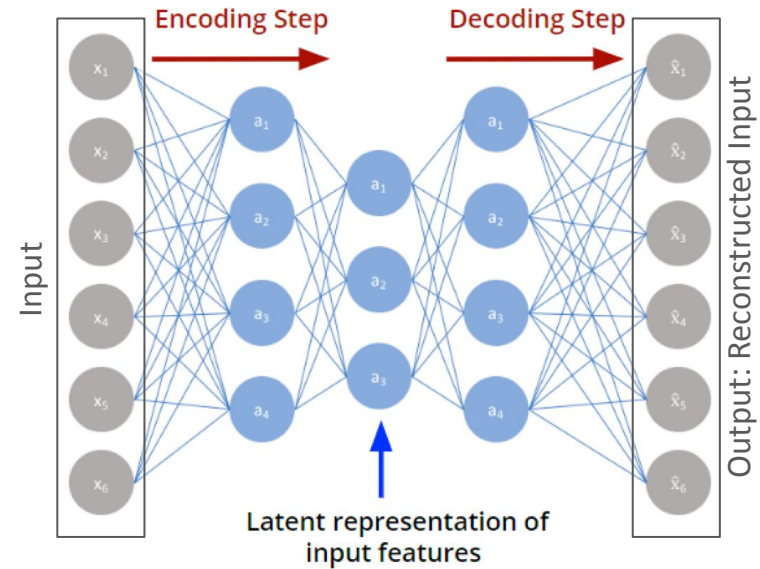
Source: DataScienceCentral

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023

Deep Learning Approaches

Long Short-Term Memory (LTM) - AE

- also, widely used in PdM

- large amount of data produced by sensors

- real-time multivariate time series

- sequences that are not easily reconstructed are anomalous sequences

- anomalous sequences can represent failures

Source: ResearchGate

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023
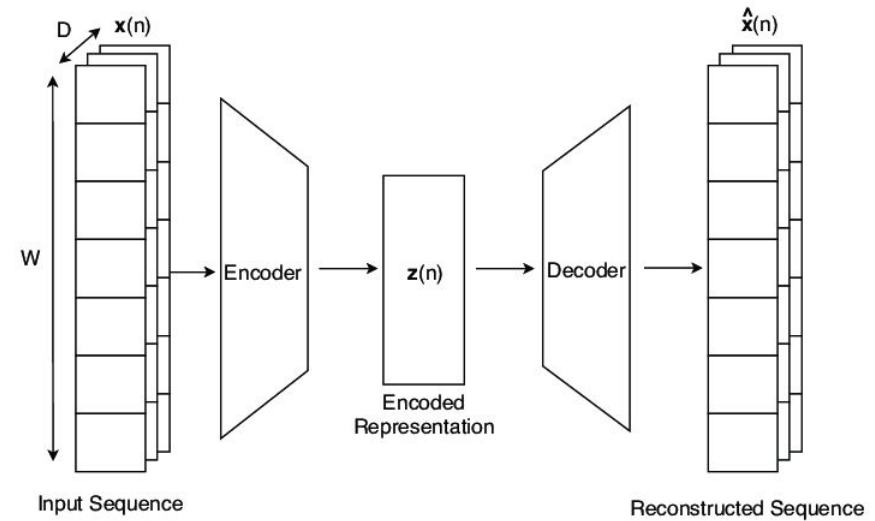
Conclusion and Discussion

Next steps …..

- **Explainable Predictive Maintenance**
    - AI is today better than human experts in predicting faults
    - It, of course, depends on the data and the complexity of the equipment
    - Why is the problem occurring, and what can be done to mitigate it?
    - How to balance different objectives and come up with maintenance plan?

- **Prescriptive Maintenance**
    - Predictive Maintenance + Prescriptive Analytics
    - Integrates asset management and maintenance systems
    - Determine the best time and actions plan for the maintenance intervention
    - Maximize the useful file of the equipment

Conclusion and Discussion

- PdM is an extremely interesting application area for ML research

- Offers a wide spectrum of problems with varying difficulty

- Many ML problem formulations are applicable to a PdM setting...
  - supervised/unsupervised (semi-, self-supervised, reinforcement ...)
  - active, transfer, online, federated, ensemble, concept drift, label noise, ...
  - feature extraction, selection and engineering
  - data fusion - integrate data from multiple sources and types (images, audio, video)
  - multi-output learning
  - XAI - explainable and interpretable models

References

- N. Davari, B. Veloso, G. A. Costa, P. M. Pereira, R. P. Ribeiro, J. Gama (2021): A Survey on Data-Driven Predictive Maintenance for the Railway Industry. Sensors 21(17): 5739

- M. S. Mouchaweh, J. Gama, R. P. Ribeiro, S. Nowaczyk, S. Pashami, B. Veloso (2021): Summer School on Data-Driven Predictive Maintenance for Industry 4.0. Co-located with DSAA 2021.

- Z. Wang (2015). Predictive maintenance (from a machine learning perspective). IEEE BigData 2015 Tutorial

- Y. Lei, N. Li, L. Guo, N. Li, T. Yan, J. Lin (2018): Machinery health prognostics: A systematic review from data acquisition to RUL prediction, Mechanical Systems and Signal Processing, Vol. 104, 2018, 799-834,

- P. Nectoux, Gouriveau R., Medjaher K., Ramasso E., Chebel-Morello B., Zerhouni N., Varnier C. (2012) PRONOSTIA : An experimental platform for bearings accelerated degradation tests. IEEE International Conference on Prognostics and Health Management, PHM'12.

- A. Saxena, J. Celaya, B. Saha, S. Saha, K. Goebel (2010): Metrics for offline evaluation of prognostic performance, Int. J. Prog. Health Manage. 1 (2010) 1–20.

Rita P. Ribeiro (rpribeiro@fc.up.pt) , MAP-i: KDD, 2022 / 2023