UNLOCKING THE POTENTIAL OF AI AND MACHINE LEARNING

FOR CANADA'S WATER SECTOR

PRESENTED BY:
Chris Gerrits, M.Sc., P.Eng.

CROZIER
CONSULTING ENGINEERS

ABOUT US

Crozier is an Ontario-based consulting engineering firm in the land development and building industry

We are committed to growing careers and building communities by delivering multidisciplinary engineering services to the private sector.

19+

Years of organic and consistent growth since being founded in 2004.

300+

Workforce of entrepreneurial, energetic, and caring employees.

5

Offices in key Ontario markets: Collingwood, Milton, Toronto, Bradford and Guelph.

ABOUT US

Chris Gerrits, M.Sc., P.Eng.

- B.Sc. in Water Resource Engineering (2000).
- M.Sc. In Water Resource Engineering (2001).
- MECP Licensed Well Technician.
- Over 30 years of experience in water well exploration, construction and testing.

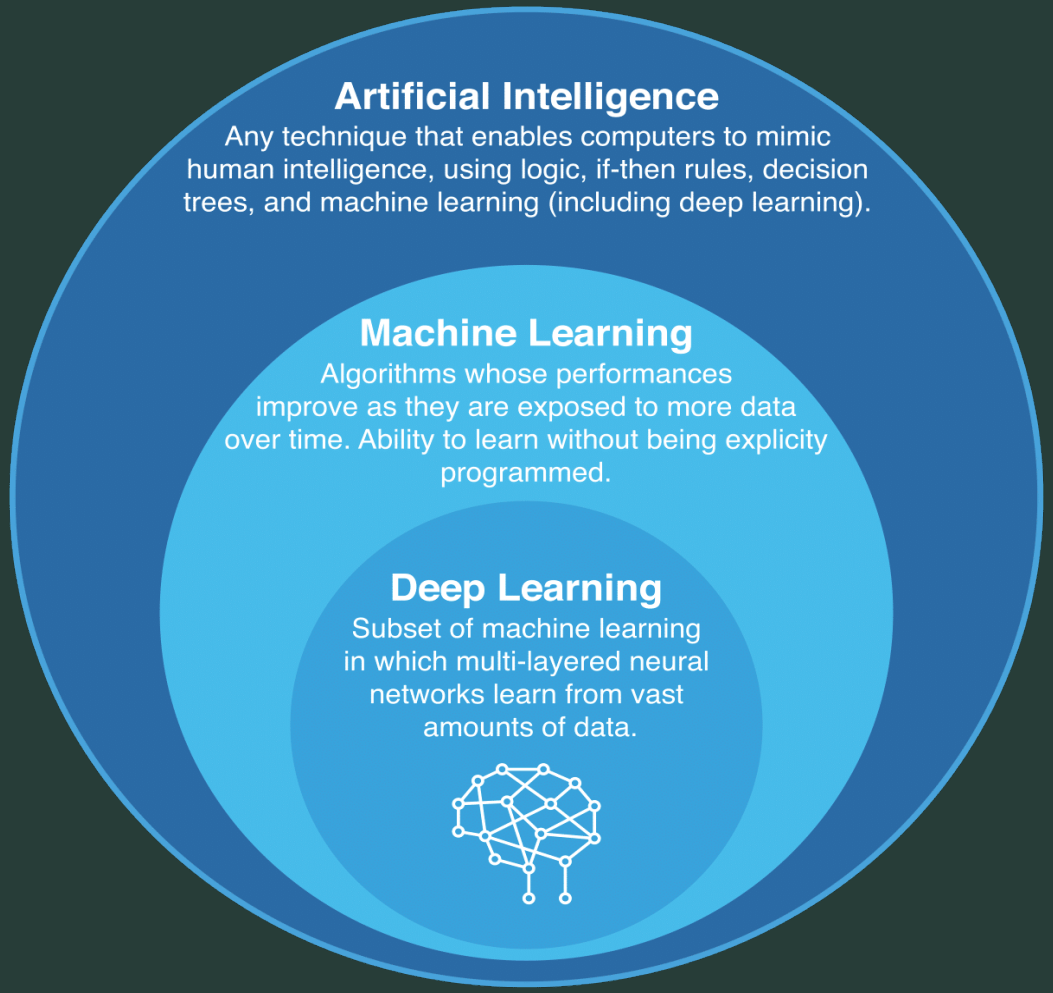
WHAT ARE ARTIFICIAL INTELLIGENCE & MACHINE LEARNING?

Artificial Intelligence (AI) is a field, which combines computer science and robust datasets, to enable problem-solving.

- **Strong AI vs Weak AI**
  - Weak AI – trained for specific tasks (Siri, Alexa, autonomous vehicles).
  - Strong AI – theoretical form of AI where a machine has intelligence equal to humans (AGI) or greater than humans (ASI).

- **Subfields of AI include Machine Learning (ML) and Deep Learning (DL)**
  - ML uses statistical methods to train algorithms to make predictions via data mining.
  - DL doesn't require a labeled dataset (ie. Supervised model), it can determine the different categories of data without human intervention.

WHAT ARE ARTIFICIAL INTELLIGENCE & MACHINE LEARNING?



- **Artificial Intelligence:** AI is like teaching machines to think and make decisions like humans, using data and algorithms.
- **Machine Leaning:** ML is a part of AI where computers learn from data to improve their tasks, without being directly programmed.
- **Deep Learning:** DL is a more advanced type of ML, using structures called neural networks to process data in complex, layered ways, similar to how the human brain works.

ML MODEL

MACHINE LEARNING MODEL

We train different models such as Random Forest, Xgboost, and Artificial Neural Network

MODEL EVALUATION

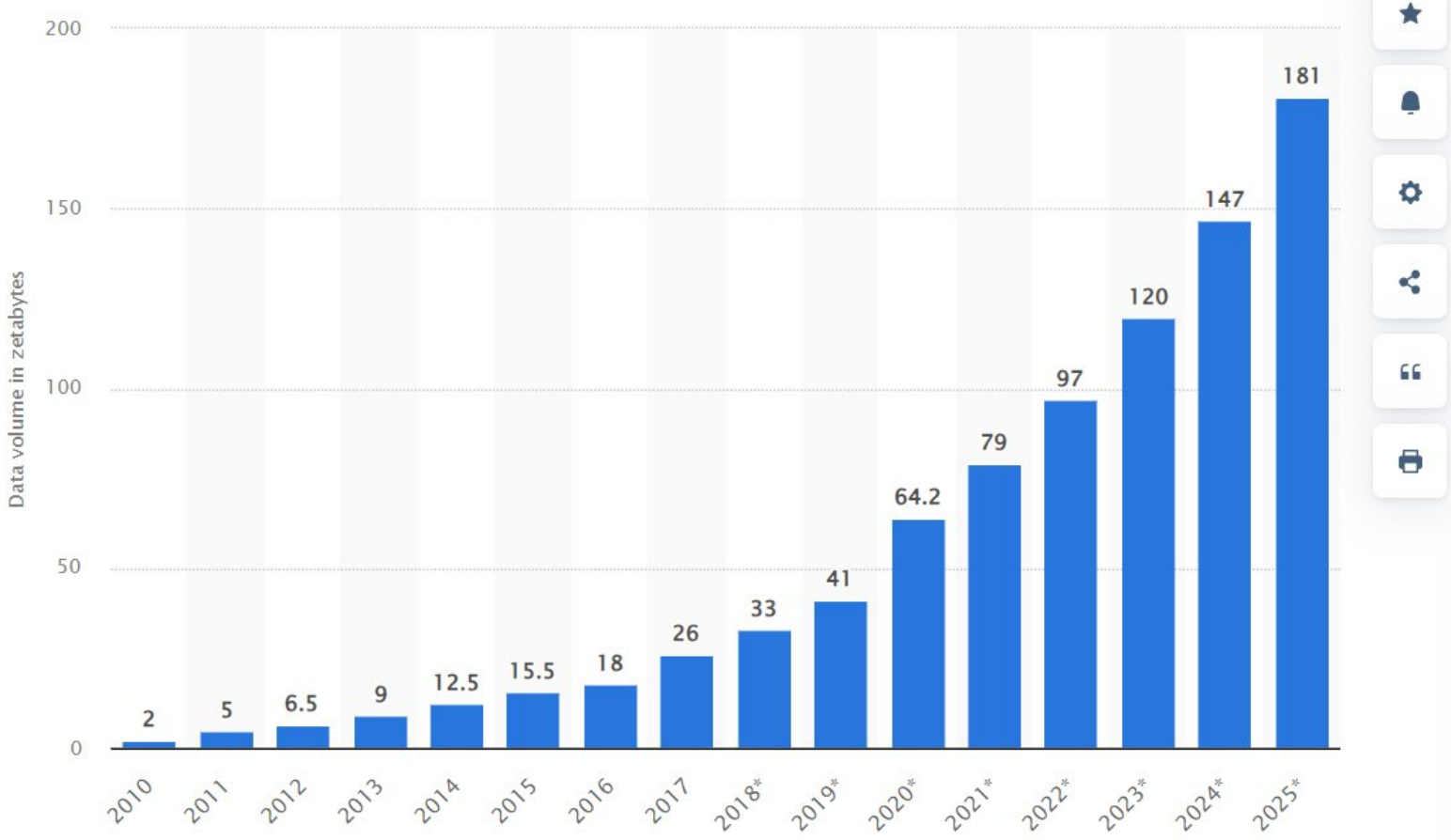
Is a process of assessing the performance and effectiveness of a model by measuring various metrices

OPTIMIZATION

of model's predictive accuracy and generalization capabilities using various methods

WORLD DATA

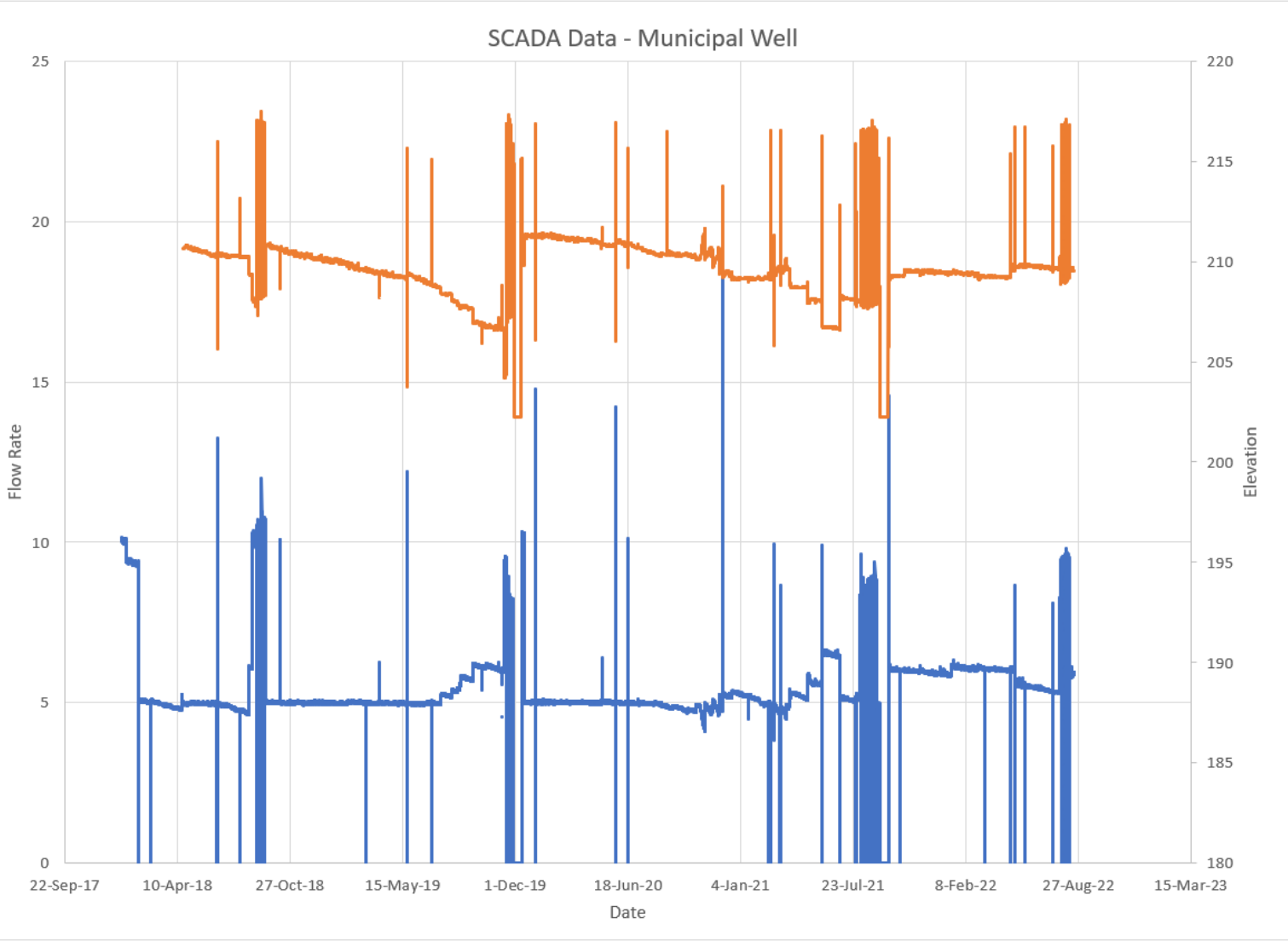
- By 2025 there will be 181 zetabytes of data floating around the world.
- It would take a single user 181 years to download all that data.
- Canada as 269 data centres across the country (2021 stats).



SUPERVISORY CONTROL AND DATA ACQUISITION TECHNOLOGY

- SCADA Systems have been in use in water & wastewater for decades.
- Niagara Region:
  - $4.4 billions in W&W Infrastructure
  - 6 WTPs
  - 11 WWTFs
  - 178 Other Facilities & Pumping Stations including 614 km of below ground pipe
- 2020 stats - treated 55.56 BL of DW and 70.31 BL of WW.
- **THAT'S A LOT OF DATA!!**

DATA SCADA EXAMPLE



- Data from 01/01/2018 through 08/18/2022.
- 40583 data points per parameter.
- Water level, flow rate, pump speed other parameters collected.
- The data is used to some extent by the client.
- Could be used more to help with decision making.
- **PREDICTIVE MAINTENANCE.**

WHAT IS PREDICTIVE MAINTENANCE?

REACTIVE

Maintenance after breakdown

PREVENTIVE

Maintenance at regular intervals

PREDICTIVE

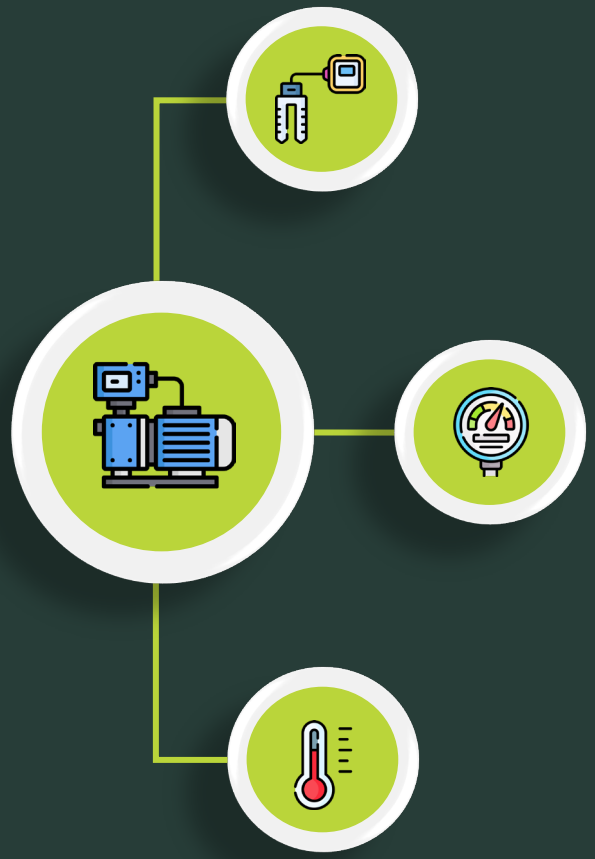
Predict breakdown before machine failure

PRESCRIPTIVE

Prescribes solutions

*"Predictive Maintenance is an **advanced technique** that helps us identify potential problems in machines before they completely stop working."*

DATA GENERATION



MACHINE WITH SENSORS

SCADA SYSTEM

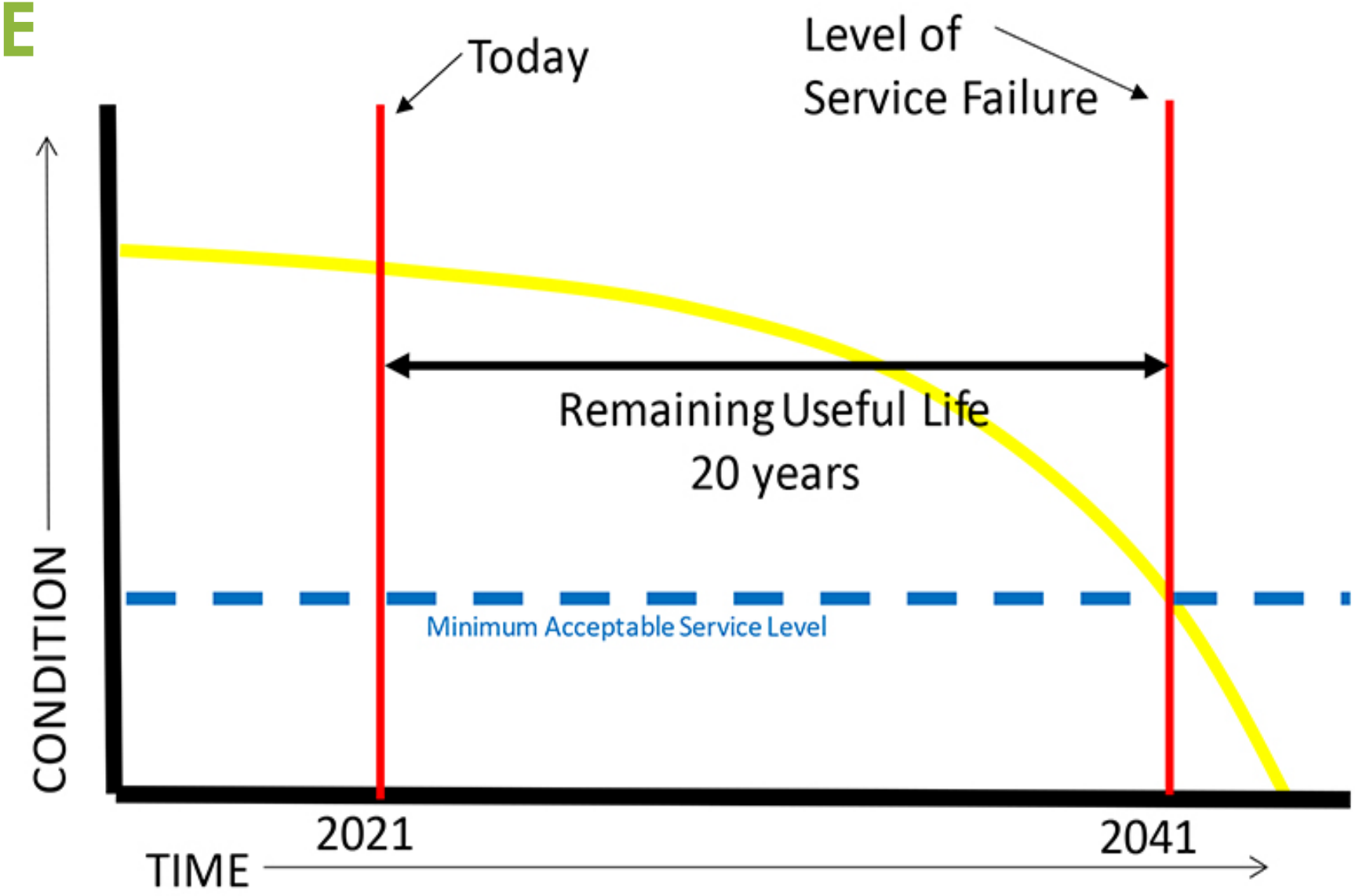
On Premises

INTERNET OF THINGS

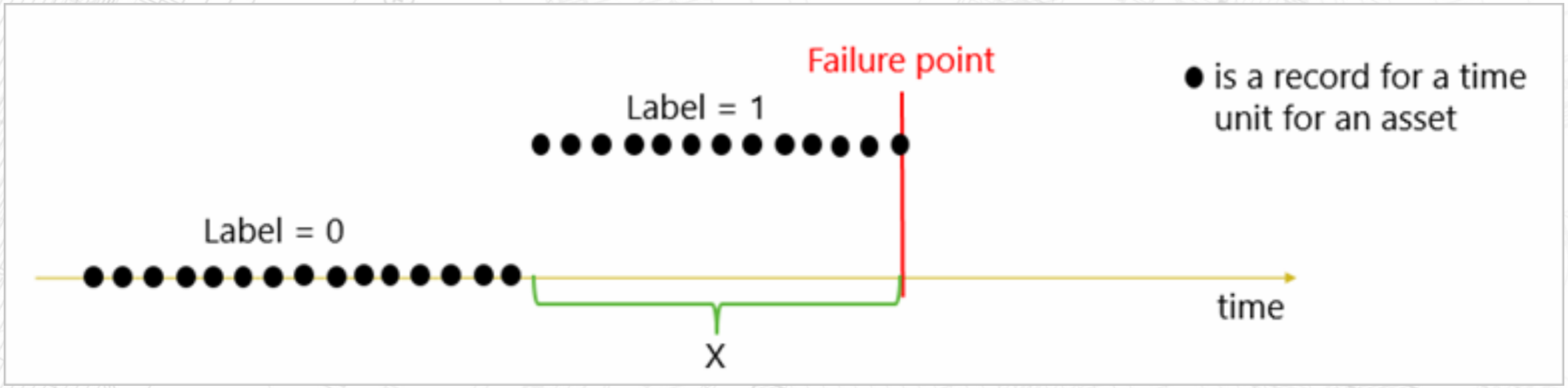
Cloud-base

ANALYTICS SYSTEM

ML FOR PREDICTIVE MAINTENANCE



ML FOR PREDICTIVE MAINTENANCE



CLASSIFICATION MODELS

Classification of different types of failures

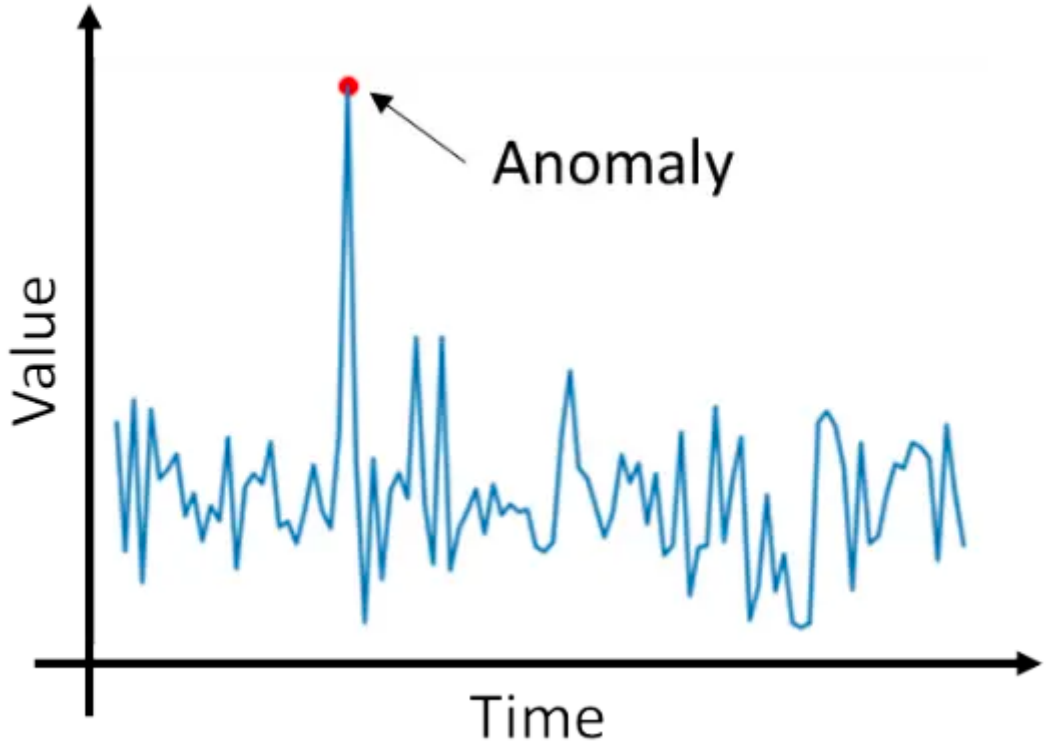
- Classification models help in predicting types of failures in a machine or predicting if asset will fail within a certain time frame
- Data needs to be accurately annotated

ML FOR PREDICTIVE MAINTENANCE

ANOMALY DETECTION

Flagging anomalous patterns

- Anomaly detection models flag outliers or abnormal behavior of a machine
- Target data is not required
- False negative prediction can cause a huge loss



CASE STUDY: Predict Remaining Useful Life of Municipal Water System Components

*We need to answer **when is the next rehabilitation** necessary for the well?*

WELL

- Rehabilitation is traditionally done on a REACTIVE basis
- Well Rehabilitation Diminishing Returns
  - Once the need for rehabilitation is established there is a long delay to schedule work

PUMPING SYSTEM

- Maintenance is traditionally done on a PREVENTIVE basis
  - After Maintenance Diminishing Efficiency

CASE STUDY: Predict Remaining Useful Life (RUL):

DATA COLLECTION

- Water Flow
- Pump Speed
- Water Elevation
- Specific Capacity
- (Calculated using Drawdown for each timestamp)

DATA CLEANING

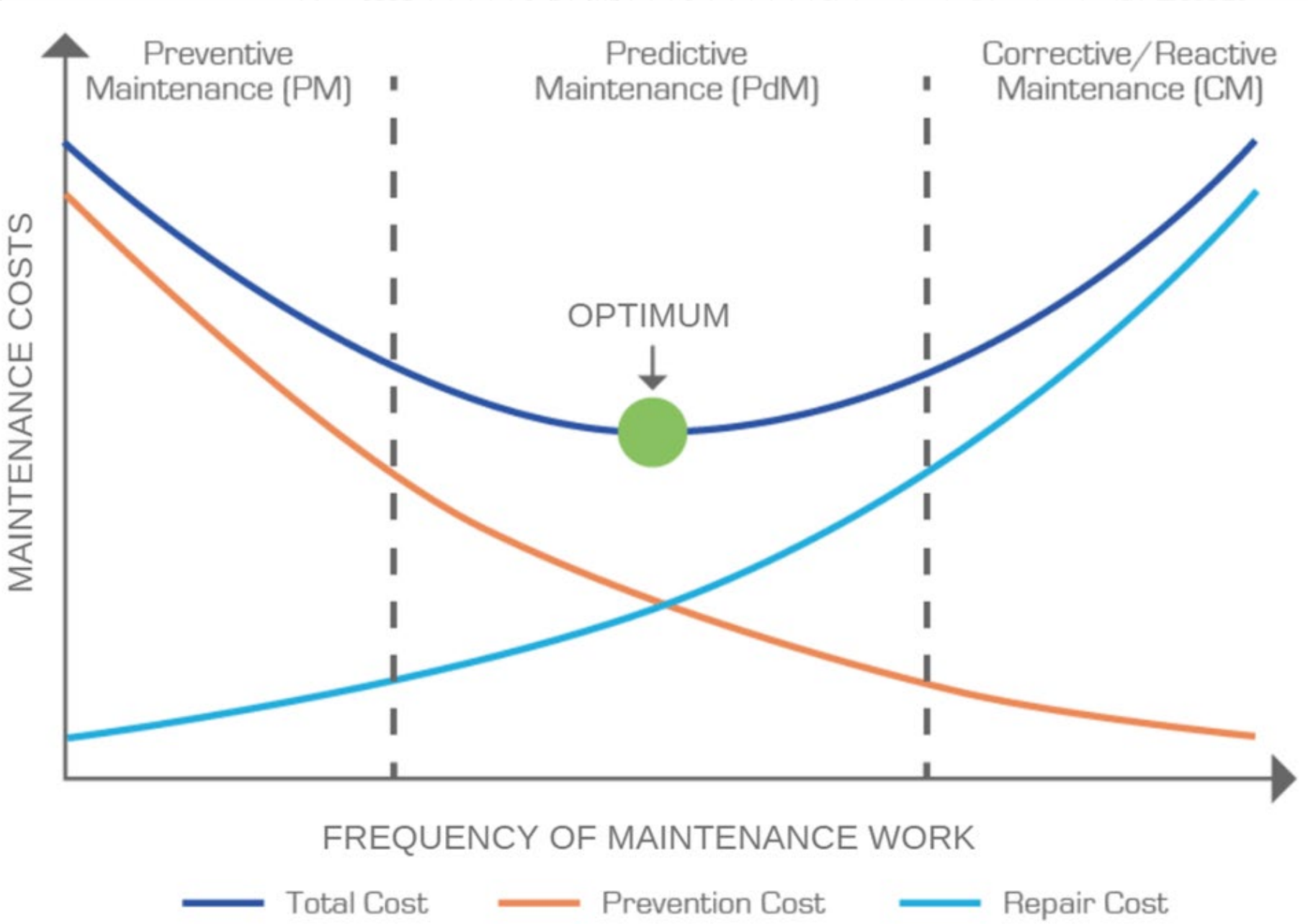
- Converted data with 5 minutes granularity to hourly granularity
- Replaced missing data points with rolling average of last 24 hours

DATA ANNOTATION

- Assigned target values in terms of number of hours

Date Time	RUL (In Hours)
2014-01-01 1:00	2
2014-01-01 2:00	1
2014-01-01 3:00	0
2014-01-01 4:00	0
2014-01-01 5:00	523
2014-01-01 6:00	522

PROBLEM STATEMENT



ML MODEL

MACHINE LEARNING MODEL

We train different models such as Random Forest, Xgboost, and Artificial Neural Network

MODEL EVALUATION

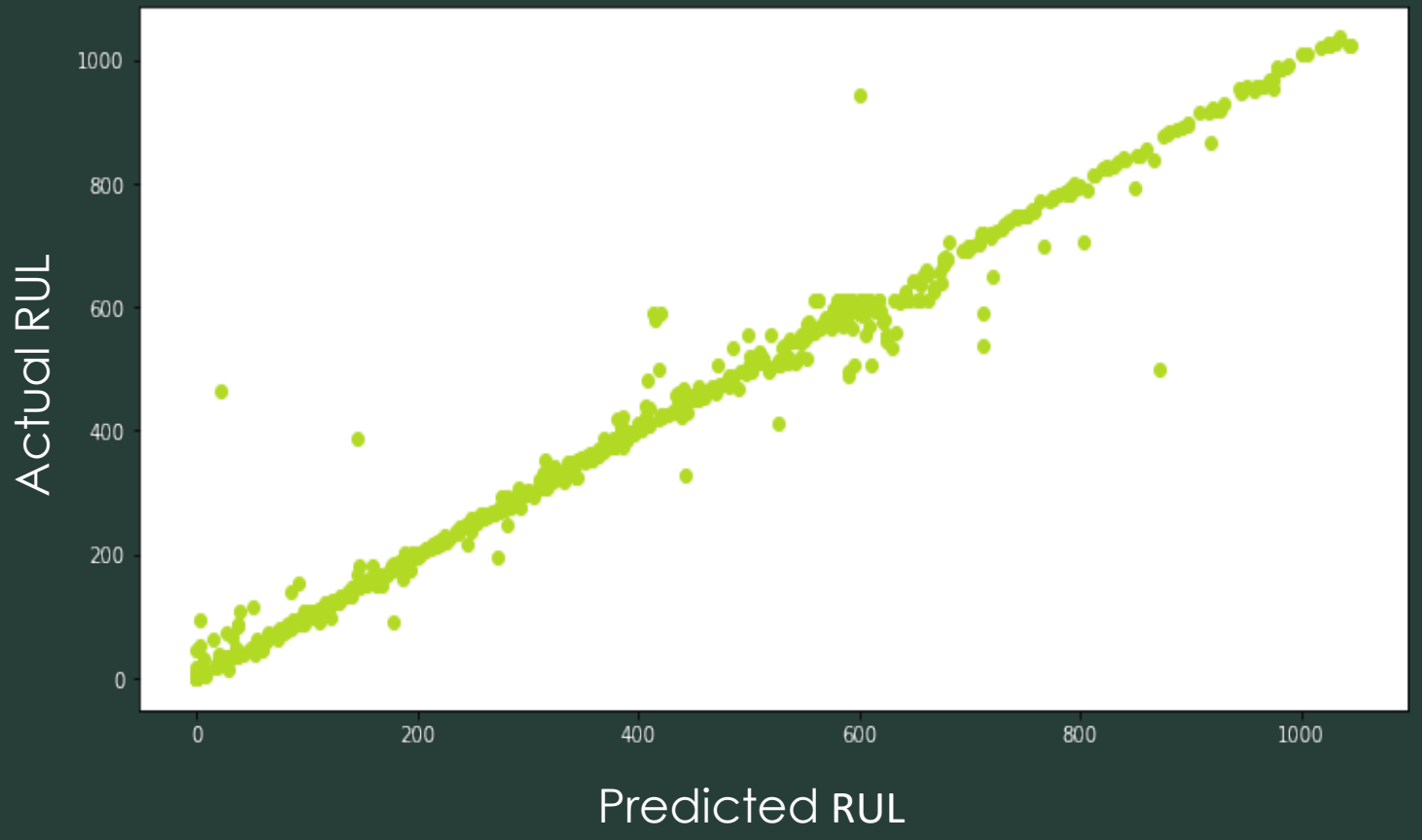
Is a process of assessing the performance and effectiveness of a model by measuring various metrices

OPTIMIZATION

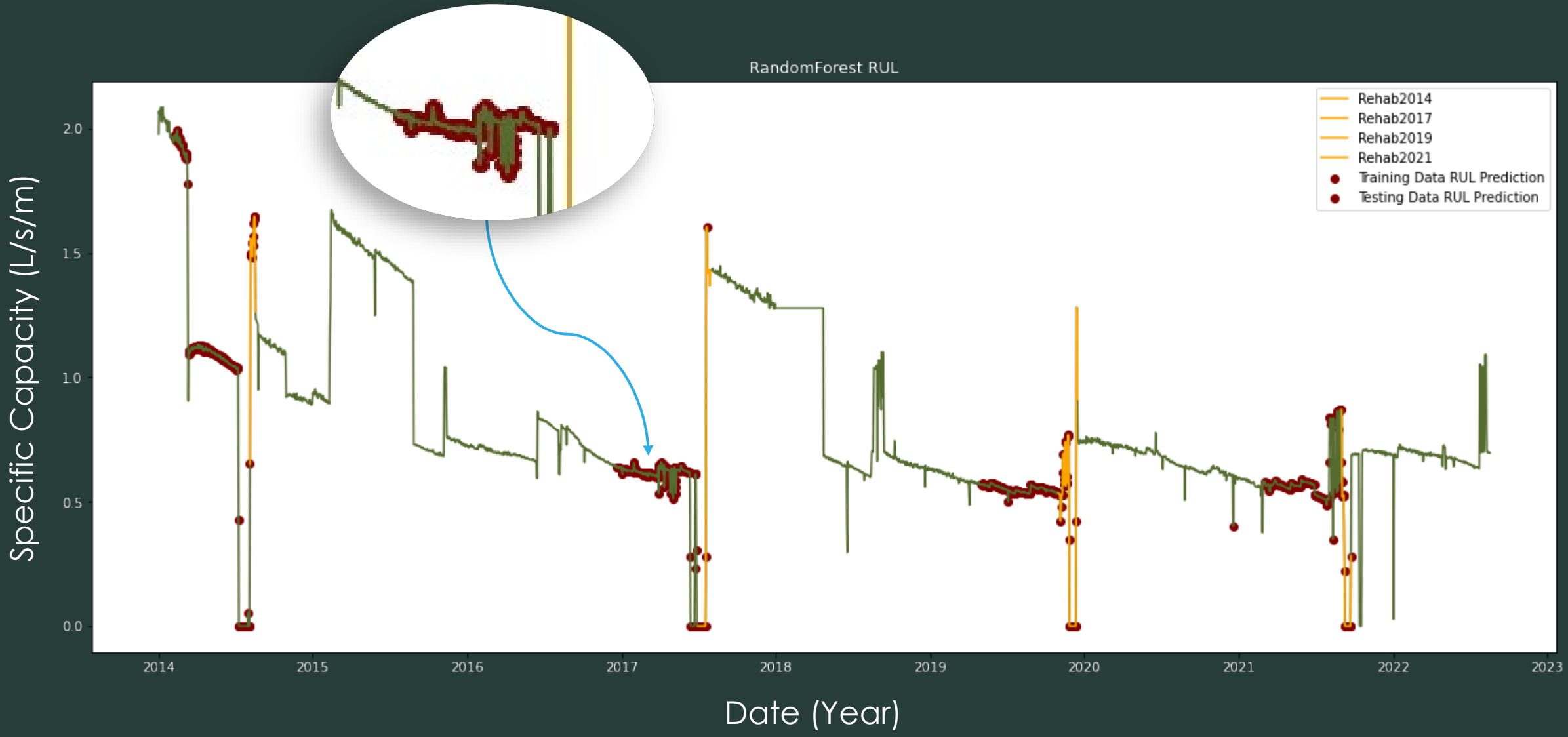
of model's predictive accuracy and generalization capabilities using various methods

RESULT

ML Regressor	RMSE	MAE	R2
XGBoost	1057.64	275.96	0.98
RandomForest	**980.6**	**200.81**	**0.98**
Neural Network	1329.96	577.64	0.96



PREDICTION OF REHAB BEFORE 180 DAYS



ADVANTAGES OF USING
ML IN PREVENTIVE MAINTANCE

ASSET MANAGEMENT

- Potentially increase the life of the asset

PROACTIVE MAINTENANCE

- Allow for advanced planning of work to reduce additional well losses

COST SAVINGS

- Longer asset life
- More time before well replacement is required

LIMITATION OF USING SUPERVISED ML MODELS

NO GENERALIZATION

- We cannot implement same model to assets with different characteristics

NO EXTRAPOLATION

- It can only interpolate
- Accurate prediction on seen data only

CONTINUOUS UPDATES

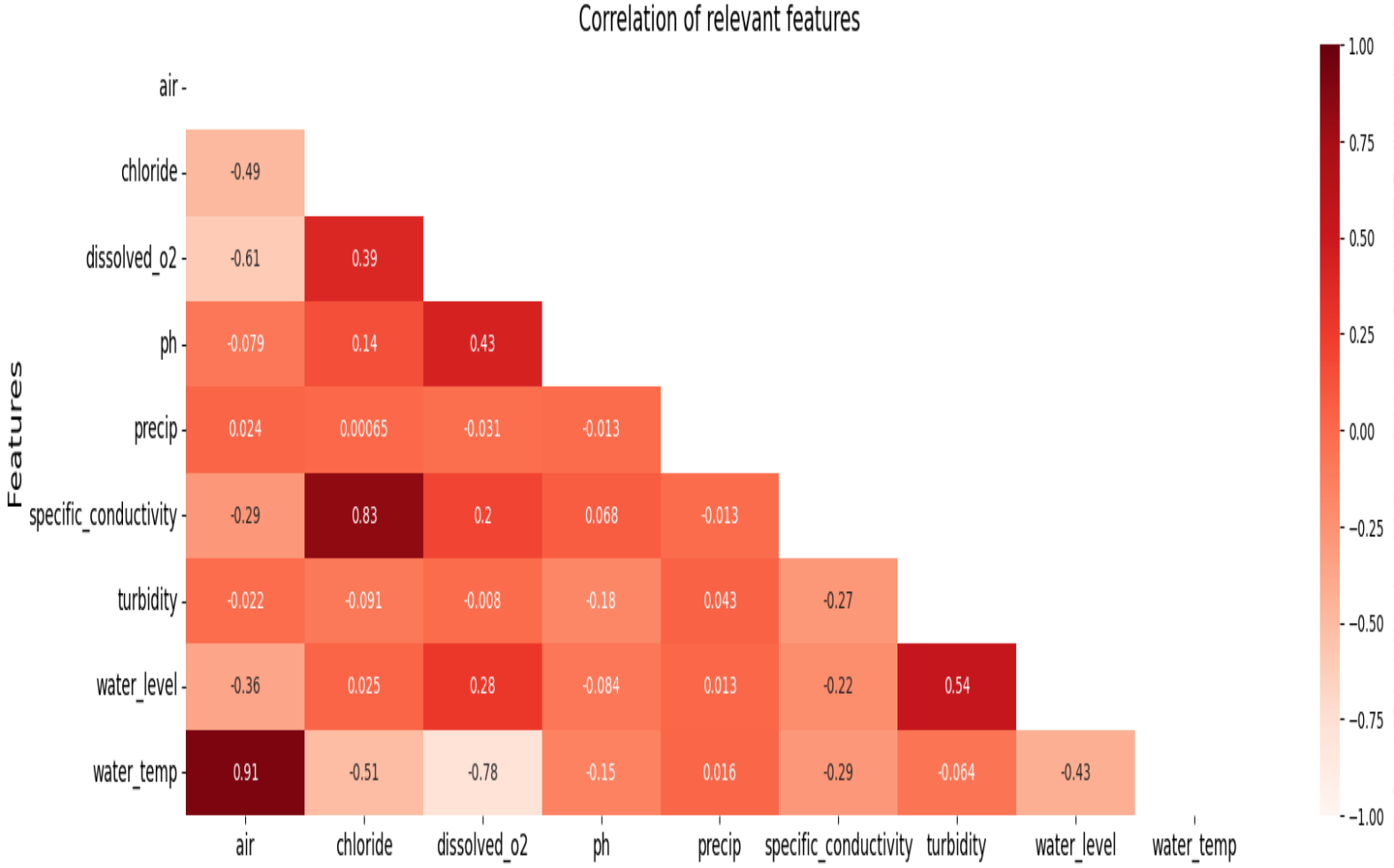
- Model needs to retrain once we have new data to fit

CASE STUDY: Water Quality Trend Analysis

- Many Conservation Authorities and other regulatory bodies complete multi variable data collection
- Sample from one surface water location in a Southern Ontario stream
- Data collected in 15 minute intervals, millions of data points over the time period investigated

Features	Start Datetime	End Datetime	Data Range
Air Temperature (deg c)	2010-02-20 07:00:00	2022-12-31 08:00:00	98943
Chloride (mg/L)	2010-02-20 07:00:00	2022-12-31 08:00:00	101647
Dissolved Oxygen (mg/L)	2010-02-20 07:00:00	2022-12-31 08:00:00	99171
pH	2010-02-20 07:00:00	2022-12-31 08:00:00	97149
Precipitation (mm/hr)	2010-02-20 07:00:00	2022-12-31 08:00:00	112730
Specific Conductivity (uS/cm)	2010-02-20 07:00:00	2022-12-31 08:00:00	98336
Turbidity (NTU)	2010-02-20 07:00:00	2022-12-31 08:00:00	94462
Water Level (masl)	2013-09-13 13:00:00	2022-12-31 08:00:00	81362
Water Temperature (deg c)	2010-02-20 07:00:00	2022-12-31 08:00:00	104818

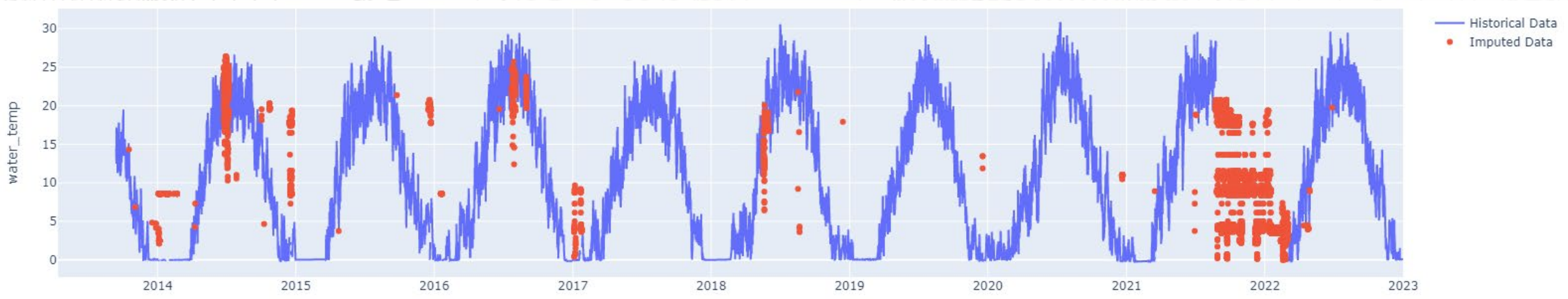
CASE STUDY: Water Quality Trend Analysis



The correlation matrix represents the correlation factor between features.

- High +ve correlations;
  - Water Temp - Air Temp (0.91)
  - Chloride - Conductivity (0.83)
- Moderate +ve correlations;
  - Water Level – Turbidity (0.54)
  - pH - Dissolved O2 (0.43)
  - Dissolved O2 - Chloride (0.39)
- High -ve correlation;
  - Dissolved O2 - Water Temperature (-0.78)
  - Dissolved O2 - Air Temperature (-0.61)
- Moderate -ve correlation;
  - Chloride - Water Temperature (-0.51)

CASE STUDY: Water Quality Trend Analysis



- Discarding high quality data because of data gaps is not optimal.
- Considering the seasonal nature of the data, replacing missing values with just the mean or median value is not suitable.
- We need to implement a sophisticated imputation method based on machine learning
- Here, we have applied the Light Gradient Boosted Machine algorithm to impute missing values in the data.
- The imputation process almost recreates similar data by considering the trend and seasonality. However, some smoothing of the predicted data will be required before fitting any machine learning model for data forecasting.

CONCLUSIONS

Globally there is an unfathomable amount of data that can be used to inform decision makers

System operators should leverage existing data at their disposal to their benefit

There is a lot to be gained by better understanding data trends

THANK YOU