

Study Committee B3

Substations & Electrical Installations

Paper ID 284

Practical Machine Learning Applications

Tom RHODES¹, Imene MITICHE², Tony MCGRAIL³, Falk WERNER³, Philip BOREHAM³

1: Duke Energy, 2: Glasgow Caledonian Uni., 3: Doble Engineering

Introduction

- Investigate practical issues with machine learning (ML)
- ML implementations may overstate expectations
- Practical results need to support decisions
- Practical implementations need to be able to justify analyses in an auditable manner: glass box, not black box

Aims for ML at Duke Energy

- Manage replacement of aging fleet – reduce emergent issues
- Targeted maintenance to extend asset life
- Reduce customer minutes interrupted from transmission asset failures
- Capture Expertise of SMEs many of whom are retiring

Issues faced in ML

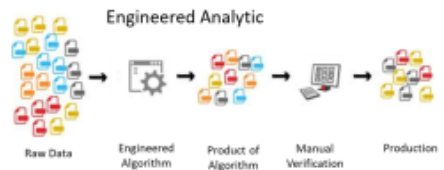
- “We tried it and it doesn’t work for power transformers” Duke Energy Engineer, 2017
- Data needs to be cleaned and maintained
- Expecting the ML to ‘sort out’ the validity and consistency issues with data did not work
- Need to have data scientists and IT professionals manage the *methods*, but Transformer SME’s manage the data *interpretations*
- Failure modes may not follow expected linear progressions; transformer populations may not follow a ‘bathtub curve’ for asset mortality
- “Your data must be wrong because if we use it our models don’t work.” Data scientist, 2019

Motivations for ML at Duke Energy

- Support Health and Risk Management analyses
- 10,000 transformers >7.5MVA
- 25,000 circuit breakers
- Lots of data, decreasing O&M dollars
- Decreasing number of people to evaluate and manage data, and experienced personnel retiring

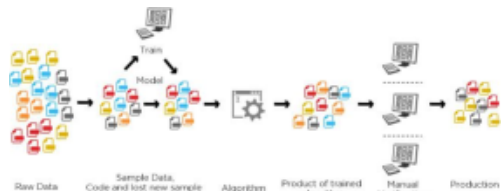
Engineering Analytic Application

- Standard analyses, IEEE etc, based on decades of experience
- Note use of SME for *manual verification*



Supervised ML at Duke Energy

- Lots of data, classified by SMEs
- Known results based on maintenance/intervention
- Note SME still performs *manual verification* of outputs



Study Committee B3

Substations & Electrical Installations

Paper ID 284

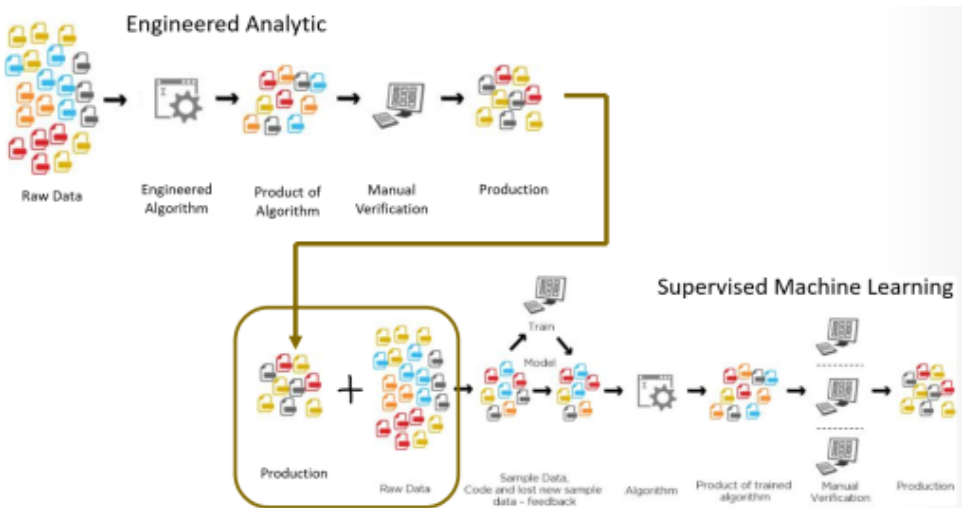
Practical Machine Learning Applications

Tom RHODES¹, Imene MITICHE², Tony MCGRAIL³, Falk WERNER³, Philip BOREHAM³

1: Duke Energy, 2: Glasgow Caledonian Uni., 3: Doble Engineering

Hybrid ML Approach - implementation

- Note the Engineered analytics have not 'gone away' both the raw data and the analytic outputs are input into the ML
- The system is good at identifying those cases which may be borderline or at an incipient stage



Hybrid ML Approach - discussion

- It was estimated that 95% of the benefit of the ML application could be achieved through data clean-up, consistency and systematic standards application
- The remaining 5%, however, are borderline cases identified early enough to support intervention in a *meaningful manner*
- The ML is required to 'do well' repeatedly on well-organized data, with well-defined analyses, and has gained acceptance within the organization

ML Outcome at Duke Energy

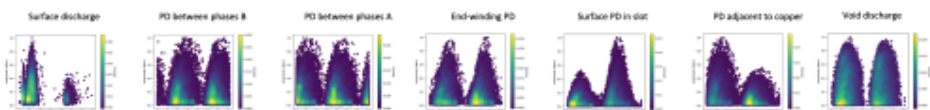
- Identification of units with incipient cooling system issues allowed for early intervention in cases which may have otherwise been missed
- Data clean up "is essential", covering 80% of initial time/effort in this application
- No 'black box' analytics allowed – SMEs can see how the ML works and what it 'considers' important

Phase Resolved Partial Discharge (PRPD)

- Partial discharge pulses recorded over multiple power system cycles, usually require expert analysis
- Some sources do have similar PRPD patterns

PRPD Implications

- Different PD sources may have different time to failure reflecting different failure modes
- Early diagnosis is often crucial to successful intervention



Study Committee B3

Substations & Electrical Installations

Paper ID 284

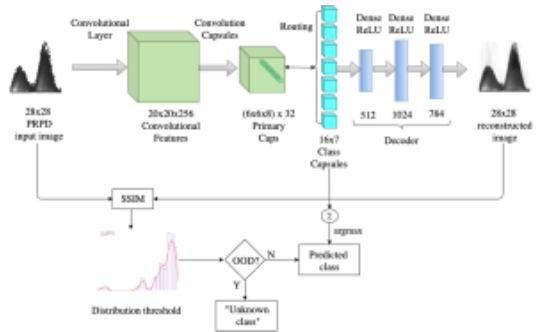
Practical Machine Learning Applications

Tom RHODES¹, Imene MITICHE², Tony MCGRAIL³, Falk WERNER³, Philip BOREHAM³

1: Duke Energy, 2: Glasgow Caledonian Uni., 3: Doble Engineering

PRPD Data Encoding/Decoding

- No point in classifying a PD pattern if it is not well represented in the training data set
- The data is 'compressed' and then reconstructed using the 'knowledge' within the AI based on the training data
- Poor reconstruction means the data is not well represented in the training set and may be an anomaly, an outlier, or a new data 'class' etc.
- "Out of distribution" (OOD) analysis helps ensure we don't inappropriately try classify spurious data



Reconstruction Error

- Three approaches used to calculate error: Mean Absolute (MAE), Mean Square (MSE) and Structural Similarity Index Matrix (SSIM)
- Set SSIM index to 0.60 as the discriminant

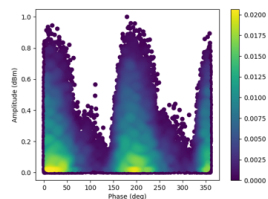
Reconstruction Error Statistics

- SSIM provides best overall accuracy

Method	Overall accuracy %	In-distribution accuracy %	OOD accuracy %
CapsNet-MAE	96.61	93.23	100
CapsNet-MSE	96.67	93.34	100
CapsNet-SSIM	97	94	100

106 MVA Steam Turbine Data - 1

- Raw data shown below: SSIM at 0.94
- Data in line with training data: classified as Phase A PD



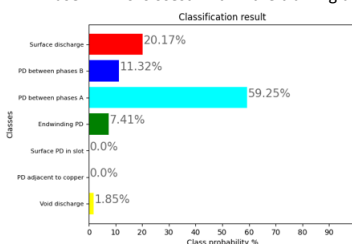
106 MVA Steam Turbine Data - 2

- Raw data shown below: SSIM at 0.38
- Data is OOD: not classified



Classification of Data - 1

- Data has similarity with a number of possible sources
- Phase A PD is 'closest' within the training data



Conclusion

- OOD detection method shows promise in identifying data which is not well represented in the original ML data set
- The SSIM metric of the original and the reconstructed PRPD images is a good indicator of OOD
- OOD detection and PD types classification in real-world data has been successful. Further work to 'generalize' has begun
- "Train a ML system to distinguish between cats and dogs... and the system can **also** learn to not try to classify an alligator as it doesn't look at all familiar..."