Evaluating Distribution Transformer Utilisation for Flexibility and Enhanced Observability using Multiple Sources of Data

SP Energy Networks

Jelena Ponoćko

Special thanks to: Rebecca Threlfall, Josephine O'Brien, Shengji Tee, Russell Bryans and Malcolm Bebbington

CIGRE UK Post Session Technical Conference
Birmingham

26/11/2024

Developing the Network of the Future

Our networks are in a period of vast and rapid change to meet our customers' Net Zero requirements.

Eight million LCTs by 2050

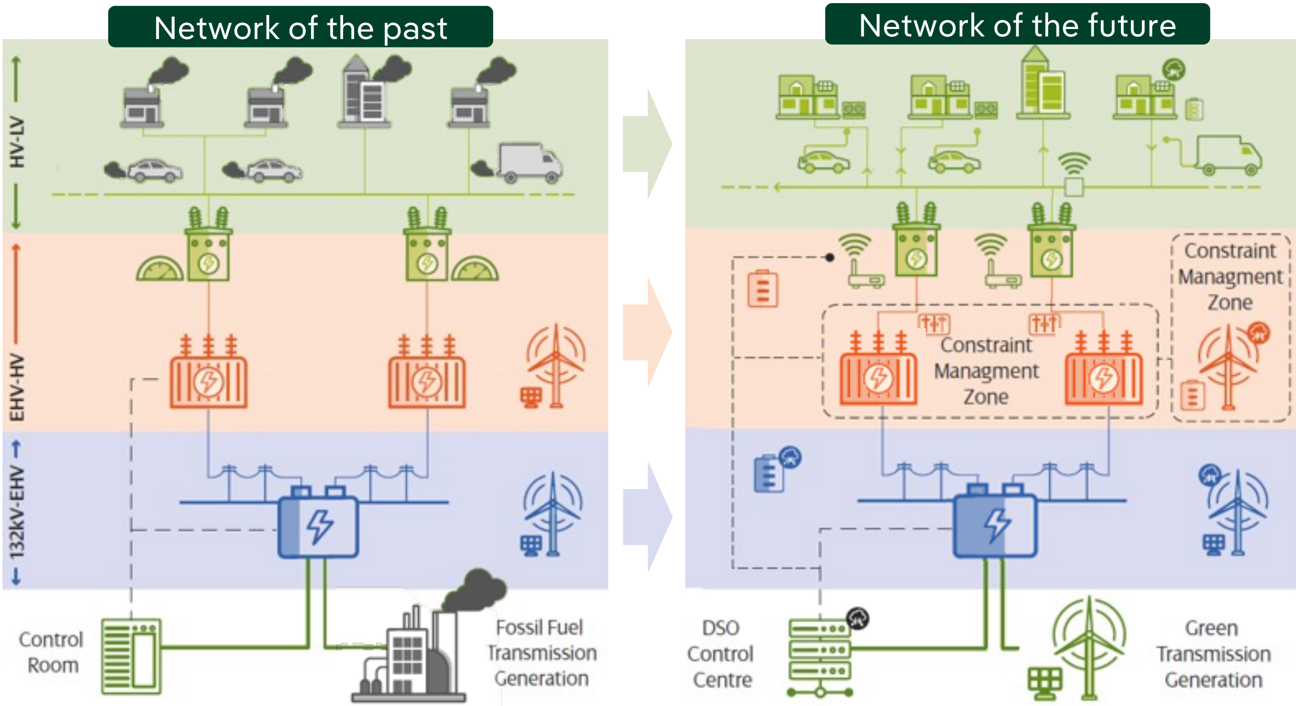
Five times more distributed generation by 2050

Double peak demand by 2050

Passive LV network with limited visibility, with manually intensive and operational processes.

High levels of generation (approx. 6GW) have been connected, with the network now approaching limits.

Closure of large fossil fuel generation.



Dynamic LV network with full visibility and control with widescale automation.

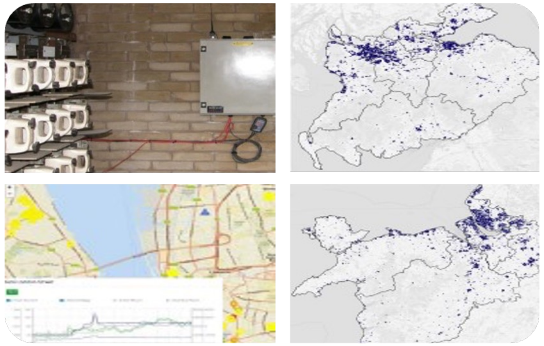
Transformation of planning and operational processes using our Engineering Net Zero Platform.

DSO Zones to enable flexibility, maximise utilisation and self-healing networks (using AI and automation).

Significant increase dynamic power flows, more complex planning & operations, and increasing whole system interactivity.

Network visibility plays a pivotal role

SP Energy Networks



LV visibility

Smart meter data, combined with LV monitoring at >14,000 substations (~75% of SPEN customers by 2028).

Fault level visibility

SPEN led ground-breaking research to measure network fault levels in real-time.

Machine learning

Using AI and machine learning to predict network state, condition, and requirements.

Increasing visibility to create a more dynamic, active, and intelligent network

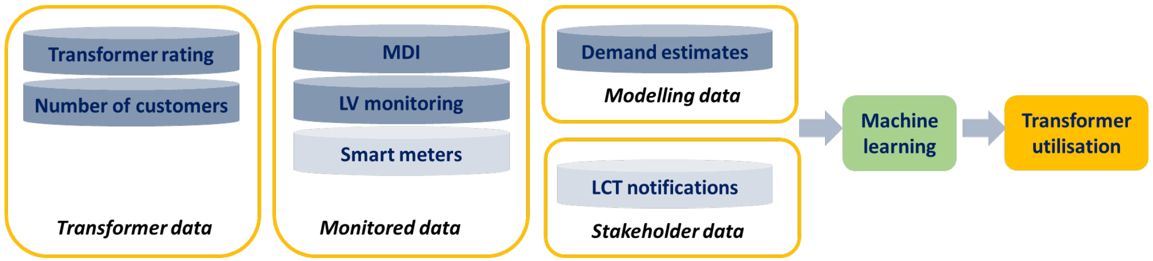
Facilitating Net Zero

	GB	SPEN
		
Electric Vehicles	~39m	~5m
Heat Pumps	~25m	~3m
Generation	~4x	~5x

Assessment of Secondary Transformer Utilisation

Machine Learning (ML) based assessment of transformer utilisation with incomplete data

- Secondary (11/0.4 kV or 6.6/0.4 kV) transformer utilisation (TU) is used in distribution network planning to evaluate the spare capacity of a transformer to supply additional demand.
- Transformer utilisation is used for:
  - Reporting network loading to the regulator
  - Identification of LV flexibility requirements
  - Stakeholders' open data requirements



- Data sources for TU assessment:
  - **Maximum Demand Indicator (MDI) data** – manually read, susceptible to missing data and prone to errors
  - **LV monitors** – high fidelity, half-hourly measurements
  - **Smart meters** aggregated consumption data – data privacy issues at lower aggregation levels

Even with 30% of missing data in our datasets (MDI measurements, LV monitor data, transformer parameters), ML allows us to estimate the utilisation of secondary transformers and **significantly improve visibility of the LV network**.

Assessment of Secondary Transformer Utilisation

SP Energy Networks

Areas with the highest priority for data analytics and machine learning

How does transformer utilisation relate to these areas?

Energy forecasting	Serves to determine the demand baseline.
Smart meter data analytics	Used as verification for aggregate smart meter data.
Asset management	Helps prioritise critical assets.
Network operation	Flags abnormal operating conditions and the need for optimised placement of new customers.
Customer segmentation	Flags abnormal operating conditions and the need for optimised placement of new customers.

Secondary Transformers Dataset

SP Energy Networks

Predictors — Response

	Record No.	INPUT				OUTPUT
Training & Testing	1	$A1_1$	$A2_1$	$B1_1$	$C1_1$	$B2_1$
	2	$A1_2$	$A2_2$	$B1_2$	$C1_2$	$B2_2$
	…					
	280	$A1_{280}$	$A2_{280}$	$B1_{280}$	$C1_{280}$	$B2_{280}$
Validation	281	$A1_{281}$	$A2_{281}$	$B1_{281}$	$C1_{281}$	$B2_{281}$
	…					
	290	$A1_{290}$	$A2_{290}$	$B1_{290}$	$C1_{290}$	$B2_{290}$

A total of 290 transformers with the following data:

A. Transformer parameters:

A1: Nameplate rating
A2: Number of customers connected

B. Monitoring data:

B1: Average value of the MDI measurements from the last three years (2020, 2021 and 2022)
B2: Maximum value of one-year worth LV monitoring data (half-hourly samples)

C. Modelling data:

C1: Utilisation based on demand estimates

Machine Learning Methods for TU Assessment (1/2)

SP Energy Networks

- **K-nearest Neighbour**: used for data pre-processing, i.e., restoring missing samples in the data, by looking at the most similar records in the training data.

- **Multiple Linear Regression**: used for finding numerical relations between a numerical response attribute (e.g., TU) and numerical predictor attributes (e.g., transformer rating, number of customers connected to the transformer, etc.):

$$y = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$$

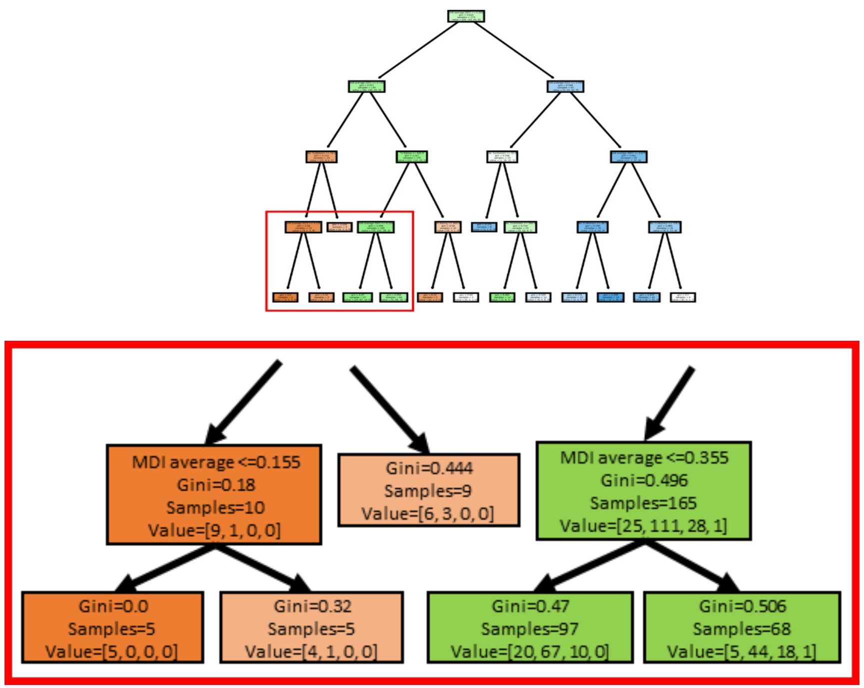
Weights (w_0 to w_n) are calculated based on the training data, i.e., a given set of examples of response values and corresponding predictors' values. Once calculated, these numerical weights can be used as predictors of the unknown TU values if the predictor attributes are known.

Machine Learning Methods for TU Assessment (2/2)

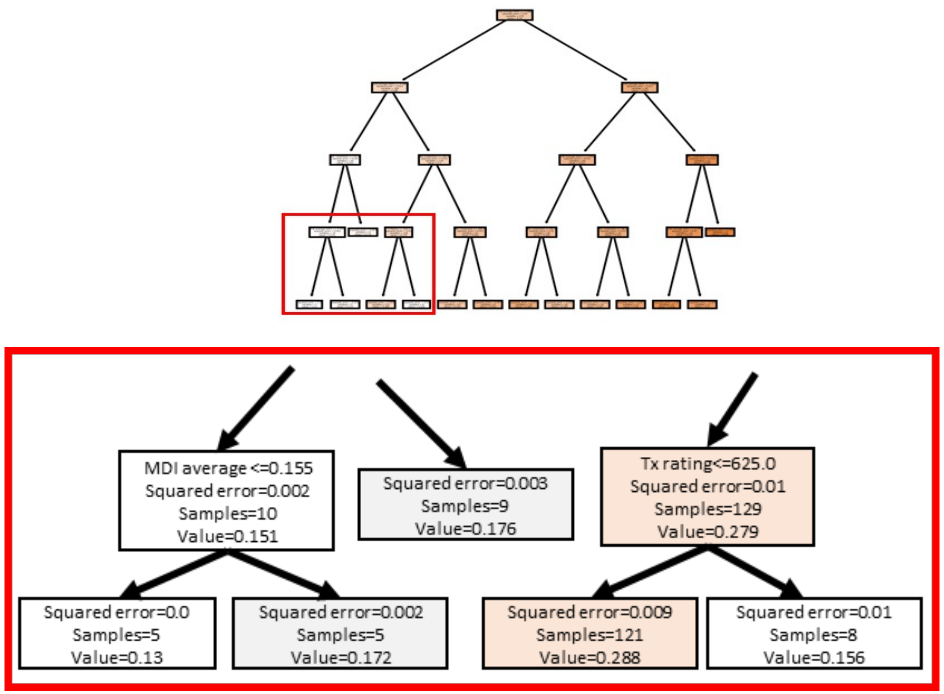
SP Energy Networks

Decision Trees, based on if-then rules:

Classification (Categorical) Trees, where the dataset is split into classes that belong to the response variable

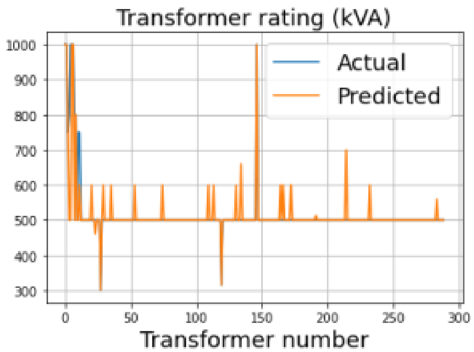


Regression Trees, where the response variable is continuous.

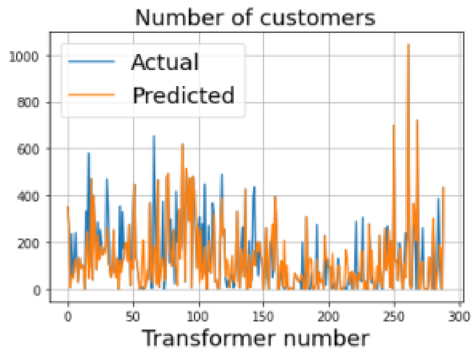


Missing Data Restoration using K-nearest Neighbour Method

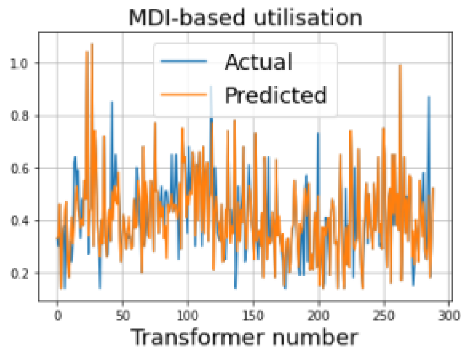
Typically, about 30% of data samples are missing or erroneous.



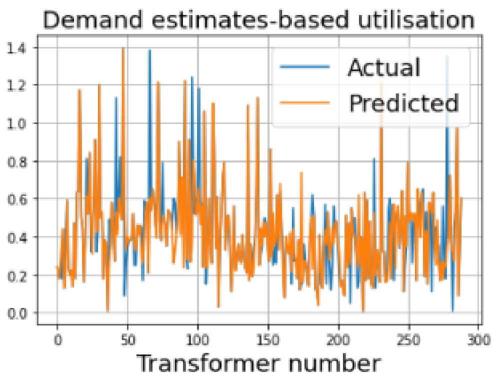
a



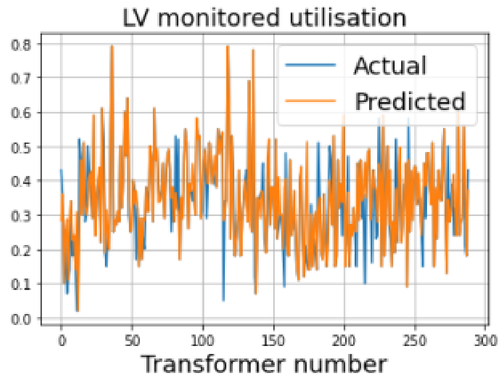
b



c



d

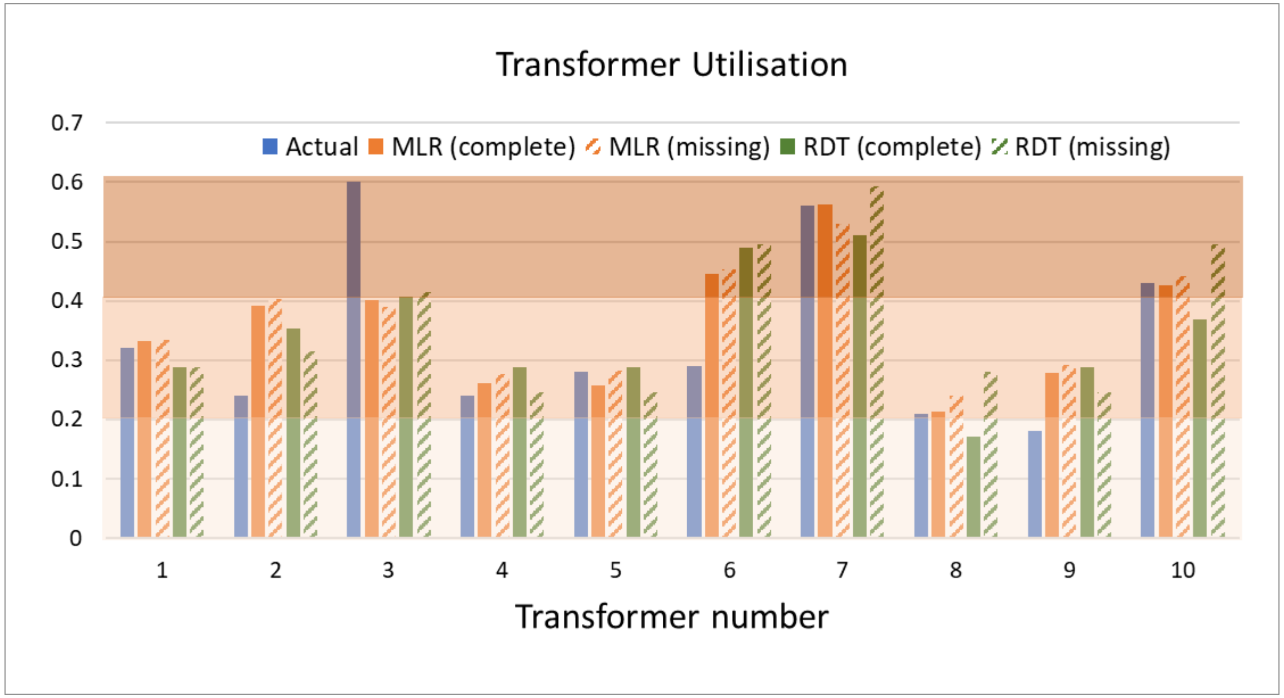


e

TU based on Linear Regression and Regression Trees

MLR – Multiple Linear Regression
RDT – Regression Decision Tree

- In most cases, **MLR and RDT result in the same band as the actual TU**.
- The lowest accuracy can be seen with transformer 3 - its TU is on the verge of category 4 (TU=0.6), while ML models estimate category 2 or 3.
- This points out the need for a larger dataset in the training process that would allow the model to identify higher values with better accuracy.



TU categories (bands)

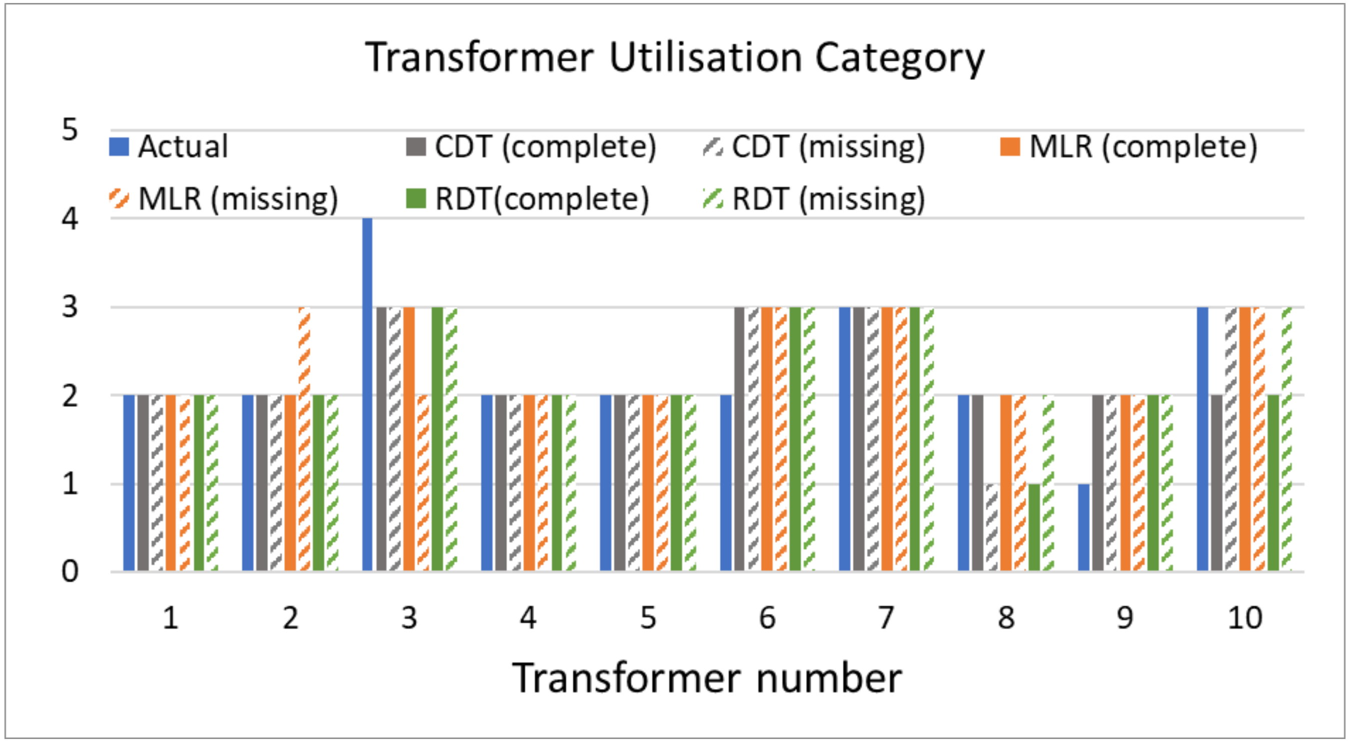
TU Category	1	2	3	4	5	6
TU Range	<0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1	>1

TU Bands based on the Three ML Methods

SP Energy Networks

MLR – Multiple Linear Regression
RDT – Regression Decision Tree
CDT – Categorical Decision Tree

- In most cases, **restored missing data** in the training dataset **will not influence the categorical estimate**.
- **Regression methods** (MLR and RDT) **perform better** and provide more comprehensive information as they estimate not only the TU category, but also its value, necessary for flexibility tenders.



TU categories (bands)

TU Category	1	2	3	4	5	6
TU Range	<0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1	>1

Conclusion

- We compared **three ML methods** for transformer utilisation assessment using multiple sources of data: multiple linear regression, regression decision trees and categorical decision trees.
- **Regression based methods** performed better, even after missing data restoration.
- As the **LV monitor rollout** progresses, complemented with smart meter data and LCT notifications, the accuracy of the methods will keep improving.


Thank You
SP Energy Networks
26/11/2024