Edward Hermann Haeusler
Informatics - PUC-Rio

Project Goal and scenario What

Currently, detecting equipment faults rely on real-time alarm triggers, scheduled maintenance, and statistics on half-life;

Failures in power transformers cost tens of millions of Reais in **Variable Portion (PV) payments – fines imposed by the regulatory agency due to equipment problems or unplanned shutdowns**;

Goal of the project: To minimize the risk of transformer stoppings and faults, proactive actions are taken through predictive analysis with **Cloud Computing and Machine Learning (ML)**;

Sample Complexity Estimation (D.Haussler)

	Classifier	*VC* dimension
1	Linear	$VC = dim(E) + 1$
2	SVM	$VC = min(dim(E), \frac{D^2}{M^2})$
3	Neural Nets	$VC = dim(E) \times \|N\|$
4	Decision trees T	$VC \approx \#nos(T)$
5	Bayesian Nets B	$VC \approx \#params(T)$

For lower VCdim classifiers, $VCdim \leq 40$, $dim(E) = 13$, $acc \geq .95$ and $conf \geq .95$ the sample should have at least 4300 trafo's failures. Furnas reports at most 2 by year in 30 years.

Our Approach

- Design of two indicators, CAI and EFRI, to assist preventive maintainance

- Use of tranfer learning for CAI, chromatography, and;

- EFRI, eletrical failures indicators and dataset, used more than 7 analog and digital datasets stored on respective systems in ELETROBRAS. It is the result of a series of balancing techniques in Data Science and ML;

Methodology:CAI

CLASSICAL *DGA* METHODS AND *Random Forest* MODEL PERFORMANCE IN THE TEST SET

Method	Accuracy	F1-score
Rogers	35.1%	24.2%
Rogers (refined)	46.8%	27.8%
Doernenburg	13.6%	14.5%
NBR 7274	51.7%	43.3%
IEC Ratio	51.2%	41.0%
IEC (refined)	64.4%	56.6%
Duval's Triangle	60.5%	38.8%
Doernenburg + Durval	69.7%	52.3%
Doernenburg + IEC (Ibrahim)	73.2%	71.3%
Random Forest	**92.2%**	**92.1%**

Assesment of the lower CVdim classifiers: SVM, Random Forest, FURIA and Random trees indicated Random Forest as the most adequate

Methodology:IRDE

EVALUATION METRICS ON TRAINING AND TEST SETS

Metric	Training	Test
Accuracy	97%	95%
AUROC	97%	89%
F1-score	97%	52%
F1-score of *failure*	98%	6%
Recall	97%	89%
Recall of *failure*	99.9%	**83%**
Relative Risk	3061	**99.7**

The IRDE development included, lower variance features removal, selection of features via decision trees, RandonUnderSampler f.b. SMOTE on the minority class and CV=10

Conclusion

Both indicators show that Machine Learning can complement classical Data Analytics tools in maintenance decision-making processes;

The CAI method outperformed the accuracy of the best classical models by 19 percentage points, while in the case of EFRI, a Random Forest achieved 95% accuracy in testing;

The Dashboard deployment will be in 2024.

References

- On the use of Machine Learning for predictive maintenance of power transformers, SBIC2023, Best paper;

- ENSIGHTS: Intelligent Monitoring of Electric Power Transmission Assets, ECIAIR 2022;

-SNTPEE 2022;