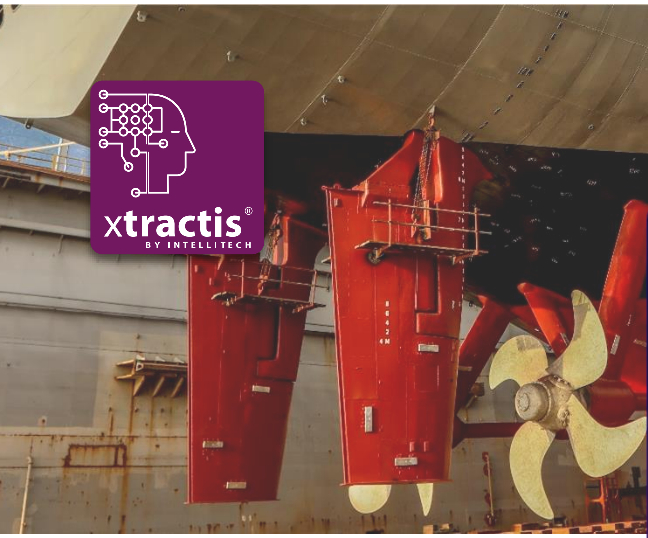


Maintenance

PREDICTION OF THE DEGRADATION OF A NAVAL PROPULSION UNIT

Benchmark vs. Random Forest, Boosted Tree & Neural Network

UC#02 — 2024/07 (v6.2) xtractis.ai

PROBLEM DEFINITION

GOAL

Design an AI-based decision system that accurately predicts the functional degradation of a naval propulsion unit compressor, given the hyper-complexity of the phenomenon (strongly nonlinear behavior) in order to rationally plan explainable maintenance operations.

PROS & BENEFITS

- Allow business experts and maintenance managers to understand the causal relationships between some turbine parameters and its future state of degradation.
- Find the truly influential parameters for assessing the state of degradation and thus reduce measurement and maintenance costs.
- Carry out turbine-specific maintenance actions to avoid critical damage, thanks to rapid and systematic diagnostics, while justifying each intervention.

REFERENCE DATA

Source:
DITEN / DIBRIS Departments of the University of Genova, Genoa

Dataset
Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science

Variable to Predict: The model predicts the **Gas Turbine Compressor Decay State Coefficient Strength** which is a continuous variable in the range [0.95 ; 1]. The lower it is, the higher the degradation (e.g., 0.97 means a compression degradation of 3 percentage points).

Potential Predictors: **14 parameters characterize the turbine**: lever position, ship speed, gas turbine shaft torque, gas turbine rate of revolutions, gas generator rate of revolutions, starboard propeller torque, port propeller torque, high pressure turbine exit temperature, gas turbine compressor inlet air temperature, gas turbine compressor outlet air temperature, …

Observations: **11,934 digital simulation points of a frigate gas turbine**, each is associated with a value of decay state coefficient. Data are divided into:

- a Learning Dataset for model induction using Training and Validation Datasets: 10,143 cases | 85% (82% for Training, 18% for Validation), and
- an External Test Dataset to check the top model's performance on real data and for benchmarking: 1,791 cases | 15%

MODEL TYPE | **Regression** ▲ | Multinomial Classification | Binomial Classification | Scoring

XTRACTIS-INDUCED DECISION SYSTEM

☑ **Intelligible Model, Explainable Decisions**

- The top-model is a decision system composed of 428 gradual rules without chaining.
- Each rule uses from 1 to 10 predictors among the 12 variables that XTRACTIS automatically identified as significant (out of the 14 turbine parameters).
- The model is fairly intelligible despite the large number of rules given the high complexity of the studied phenomenon: one can still undertand and explain how the system makes its decision because only a few rules are triggered at a time to compute the result.

☑ **High Predictive Capacity** It has an excellent Real Performance (on unknown data).

☑ **Ready to Deploy** It computes real-time predictions up to 70,000 decisions/second, offline or online (API).

XTRACTIS PROCESS

STEPS	Turbine parameters Reference Data	INDUCTION	XTRACTIS Top-Model	Turbine parameters New Cases	DEDUCTION	State of the turbine Automated Decision (alert about the turbine degradation)
SOFTWARE ROBOTS	XTRACTIS® REVEAL Delivers the decision system + its Structure & Performance Reports			XTRACTIS® PREDICT Delivers the decision + the Prediction Report explaining its reasoning		

TOP-MODEL INDUCTION

INDUCTION PARAMETERS

Powered by: XTRACTIS® REVEAL v12.2.43064

1. We launch 700 inductive reasoning strategies; each strategy is applied to the same single partition of the learning dataset (82% Training / 18% Validation) to get a reliable assessment of the descriptive and predictive performances, respectively from Training and Validation Datasets.
2. Each strategy thus generates one unitary model called **Individual Virtual Expert** (IVE).
3. Among the 700 induced models, the top-IVE selected is the one that has the best predictive performance, close to its descriptive performance, and with the best intelligibility, i.e., with the fewer predictors and rules.

Total number of induced unitary models
700 IVEs

Criterion for the induction optimization
RMSE

Validation criterion for the top-model selection
RMSE

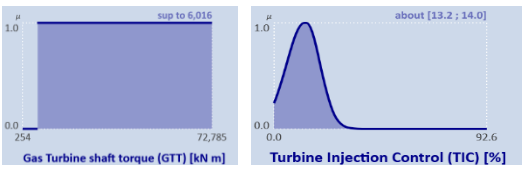
Duration of the process @ Induction Speed FP64
~17 days @ 1Tflops

TOP-MODEL STRUCTURE

The top-model has a poor intelligibility as it has **428 rules aggregated into 36 disjunctive rules** and combining **12 predictors** with 3.6 predictors per rule on average. But it remains intelligible as its Structure Report reveals all the internal logic of the decision system and ensures that the model is understandable by the human expert. It is a transparent model that can be audited and certified before deployment to end-users.

PREDICTORS

- 12 parameters out of 14
- Ranked by individual contribution (10 strong and 2 medium signals):
  #1 GT Compressor outlet air temperature
  #2 GT Compressor outlet air pressure…
- Labeled by fuzzy and binary classes
  Examples: **binary interval** "sup to 6,016";
  fuzzy interval "about [13.2 ; 14.0]"



RULES

- 428 connective fuzzy rules without chaining (aggregated into 36 disjunctive fuzzy rules)
- 1 to 10 predictors per rule (on average, 3.6 predictors per rule)
- Example: Fuzzy rule **R330** uses 6 predictors and concludes {0.988}. 427 other rules complete this model.

IF	Gas Turbine shaft torque [kN m]	IS	sup to 6,016
AND	Gas Turbine rate of revolutions (GTn) [rpm]	IS	inf to ~1,369
AND	Gas Generator rate of revolutions (GGn) [rpm]	IS	inf to ~6,656
AND	GT Compressor outlet air temperature (T2) [C]	IS in	~ [567.805 ; 570.087]
AND	Turbine Injection Control (TIC) [%]	IS in	~ [13.2 ; 14.0]
AND	Fuel flow (mf) [kg/s]	IS	Sup to ~0.30
THEN	GT Compressor decay state coefficient	IS	0.988

Literally, the Compressor is decaying by 1.2 percentage points if the Gas Turbine shaft torque is above 6,016 kN m, and its rate of revolutions is under approximately 1,369 rpm, and the Gas Generator rate of revolutions is below approximately 6,656 rpm, and the Gas Turbine Compressor outlet air temperature (T2) is between approximately 567.805°C to 570.087°C, and the Turbine Injection Control is between approximately 13.2% and 14.0%, and the Fuel flow (mf) is above about 0.30kg/s.

TOP-MODEL PERFORMANCE

The top-IVE performances, measured in Training / Validation, then in External Test on reference data, guarantee the model's predictive and real performances.

Performance Type	DESCRIPTIVE	REAL	REAL
Dataset	82% Training	18% Test	External Test
RMSE	3.7×10^{-4}	3.7×10^{-4}	5.2×10^{-4}
Correlation	1.000	1.000	0.999

EXPLAINED PREDICTIONS FOR 2 UNKNOWN CASES

Powered by: XTRACTIS® PREDICT v12.2.43064

CASE
(from the External Dataset,
i.e., not included in the Learning Dataset)

DEDUCTIVE INFERENCE OF RULES

AUTOMATED DECISION

TURBINE # 11866

actual value = 1.000	
Lever position (lp)	4.16
Gas Turbine shaft torque (GTT) [kN m]	14,718
Gas Turbine rate of revolutions (GTn) [rpm]	1,547
Gas Generator rate of revolutions (GGn) [rpm]	7,716
Port Propeller Torque (Tp) [kN]	114
HP Turbine exit temperature (T48) [C]	623
GT Compressor outlet air temperature (T2) [C]	600.199
HP Turbine exit pressure (P48) [bar]	1.66
GT Compressor outlet air pressure (P2) [bar]	8.9
Gas Turbine exhaust gas pressure (Pexh) [bar]	1.0230
Turbine Injection Control (TIC) [%]	16.4
Fuel flow (mf) [kg/s]	0.33

Real Time

#Rule	Conlusion	Firing degree
R151 →	0.968	2.87E-04
R162 →	0.970	4.29E-04
R209 →	0.975	5.94E-04
R210 →	0.975	5.81E-04
R283 →	0.983	0.003
R309 →	0.986	0.007
R310 →	0.986	0.002
R344 →	0.990	0.006
R343 →	0.990	0.002
R342 →	0.990	0.001
R379 →	0.993	0.096
R419 →	1.000	1.000
PREDICTION=	**0.999**	

NUMBER OF TRIGGERED RULES
12 /428

FUZZY PREDICTION
{ 1.000 | 1.000,
0.993 | 0.096,
0.986 | 0.007,
0.990 | 0.006,
… }

FINAL PREDICTION
0.999

The decision system delivers the correct prediction compared to the actual value.

VERY LOW DEGRADATION

TURBINE #6391

actual value = 0.977	
Lever position (lp)	1.14
Gas Turbine shaft torque (GTT) [kN m]	988
Gas Turbine rate of revolutions (GTn) [rpm]	1,366
Gas Generator rate of revolutions (GGn) [rpm]	6,651
Port Propeller Torque (Tp) [kN]	8
HP Turbine exit temperature (T48) [C]	469
GT Compressor outlet air temperature (T2) [C]	548.340
HP Turbine exit pressure (P48) [bar]	1.13
GT Compressor outlet air pressure (P2) [bar]	6.1
Gas Turbine exhaust gas pressure (Pexh) [bar]	1.0190
Turbine Injection Control (TIC) [%]	2.2
Fuel flow (mf) [kg/s]	0.09

Real Time

#Rule	Conlusion	Firing degree
R18 →	0.950	0.137
R17 →	0.950	0.004
R122 →	0.964	0.072
R120 →	0.964	0.071
R141 →	0.966	0.344
R139 →	0.968	0.005
R156 →	0.969	0.005
R183 →	0.971	0.139
R218 →	0.975	0.361
R255 →	0.980	0.108
R317 →	0.986	0.486
R371 →	0.992	0.203
R372 →	0.992	4.14E-04
R426 →	1.000	0.005
PREDICTION=	**0.976**	

NUMBER OF TRIGGERED RULES
14/428

FUZZY PREDICTION
{ 0.986 | 0.486,
0.975 | 0.361,
0.966 | 0.344,
0.992 | 0.203
… }

FINAL PREDICTION
0.976

The decision system delivers the correct prediction compared to the actual value.

MEDIUM DEGRADATION

TOP-MODELS BENCHMARK

MODELING PARAMETERS

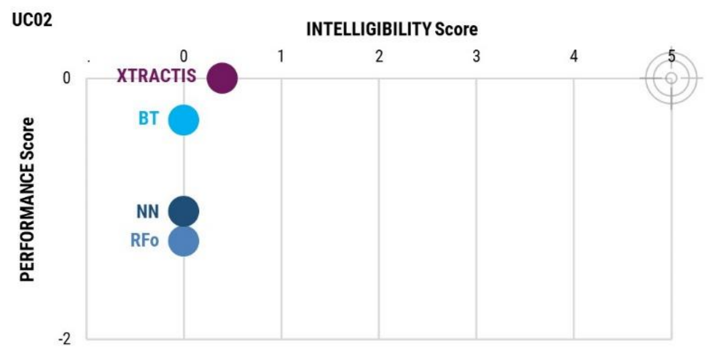
	XTRACTIS	RANDOM FOREST	BOOSTED TREE	NEURAL NETWORK
MODELS RELEASE	2022/08	2021/08	2021/05	2022/05
ALGORITHM VERSION	XTRACTIS **REVEAL** 12.2.43064	Python 3.6 | LightGBM 2.2.2	Python 3.6 | LightGBM 2.2.2	Python 3.6 | TensorFlow 2.6.0 | Keras 2.6.0
CROSS-VALIDATION TECHNIQUE	All explored strategies for all algorithms use the same single-split of the Learning Dataset: 82% Training | 18% Validation			
NUMBER OF EXPLORED STRATEGIES[1]	700 induction strategies	2,000 ML strategies	300 ML strategies	2,000 ML strategies
TOP-MODEL SELECTION[2]	Top-IVE among 700 IVEs	Top-IVE among 2,000 IVEs	Top-IVE among 300 IVEs	Top-IVE among 2,000 IVEs

TOP-MODEL STRUCTURE

	XTRACTIS	RANDOM FOREST	BOOSTED TREE	NEURAL NETWORK
NUMBER OF PREDICTORS (out of 14 Potential Predictors)	**12**	**14**	**14**	**14**
AVERAGE NUMBER OF PREDICTORS PER RULE / EQUATION	**3.6** per rule	**4.8** per rule	**7.9** per rule	**20.1** per equation
STRUCTURE OF THE DECISION SYSTEM	**428** fuzzy rules without chaining (aggregated into 36 disjunctive fuzzy rules) Only a few rules are triggered at a time to compute a decision	**119** trees without chaining **61,800** binary rules	**1** chain of **4,341** trees **366,487** binary rules Tree #N corrects the error of the N-1 previous trees	**2** hidden layers | **52** hidden nodes **53** equations 52 unintelligible synthetic variables in addition to the 14 original predictors

TOP-MODEL SCORES

	Random[3]	XTRACTIS	RFo	BT	NN
INTELLIGIBILITY Score[4]		0.39	0.00	0.00	0.00
IVE Real Perf. (RMSE $(\times 10^{-4})$) in External Test	*200.0*	0.75	2.00	1.07	1.77
Gap to IVE Leader in External Test		0.00	-1.25	-0.32	-1.02
PERFORMANCE Score[4]		0.00	-1.25	-0.32	-1.02



(1) For all algos: on the same Learning Dataset. All models are optimized according to their Validation RMSE.
(2) All top-models are selected according to their Validation RMSE while checking that it remains close to their Training RMSE.
(3) Baseline performances that models must exceed to perform better than chance (P-value = 0.001; 100,000 models generated by random permutation of the output values). The value of each performance criterion is generally achieved by a different random model.
(4) See Appendices for explanations and detailed results. Performance Scores are calculated on all available unknown data.

More Use Cases: xtractis.ai/use-cases/

APPENDIX 1 — Calculation of the Intelligibility × Performance Scores

AI Technique #i	T_i	$i \in [1 ; n]$ n = number of AI Techniques benchmarked in terms of data-driven modeling = 5
Benchmark #k	B_k	$k \in [1 ; p]$ p = number of Benchmarks for the Use Case ∈ {1, 2, 3}

Remarks:

- In case of a small number of reference data, a CVE model (College of Virtual Experts) is generated by each explored strategy of T_i, generally via an N×K-fold cross validation. In this case, a Benchmark is led with the top-CVE on the External Test Dataset (ETD, composed of unknown reference cases). Then, a top-IVE model (Individual Virtual Expert) is generated from the top-CVE, through the XTRACTIS® reverse-engineering process, or for the other T_i, by applying the top-strategy, which has generated the top-CVE, on the Training and Validation Datasets. And a second Benchmark is led with this top-IVE on the same ETD.
- In case of a huge number of reference data, an IVE is generated by each explored strategy of T_i, via a 1-split validation. In this case, Benchmarks are led with the top-IVE on the Test Dataset (TD, composed of unknown reference cases) and on the available ETDs.
- Each Benchmark uses the latest versions of the following algorithms available at the date of the benchmark. XTRACTIS®: REVEAL; Logistic Regression: Python, Scikit-Learn; Random Forest & Boosted Tree: Python, LightGBM; Neural Network: Python, TensorFlow, Keras.
- Each B_k uses exactly the same TD and ETD for each T_i model.
- No Regression models can be obtained by Logistic Regression. So, this Data Analysis technique is benchmarked only for Classification or Scoring problems.
- The Holy Grail for critical AI-based decision systems is to obtain a model with the highest Performance <u>and</u> the highest Intelligibility scores (top-right corner of the graph).

PERFORMANCE Score

For each B_k, we calculate the values of the Performance Criterion (PC) on the same ETD for all the T_i top-CVEs; and on the same TD and ETDs for all the T_i top-IVEs. The PC is: RMSE in percentage for a Regression; F_1-Score for a Binomial Classification; Average F_1-Score or Average F_2-Score for a Multinomial Classification; Gini index for a Scoring.
Then, we compare the value of the PC of each T_i top-CVE (resp. top-IVE) to the best value of this PC reached by the best T_i top-CVE (resp. top-IVE) on ETD (resp. on TD and ETDs).

For Regression, we calculate for each T_i top-model (CVE and IVE): $PS(T_i, B_k) = Best\_PC(B_k) - PC(T_i, B_k)$.

For Classification and Scoring, we calculate for each T_i top-model: $PS(T_i, B_k) = PC(T_i, B_k) - Best\_PC(B_k)$.

Performance Score of T_i

$$\boxed{PS(T_i) = \text{Mean}\,(PS(T_i, B_k))_{k \in [1 ; p]}}$$

<u>Remark:</u>

- Each PS varies theoretically from -100 (Lowest Score) to 0 (Highest Score), but practically between -50 and 0.

INTELLIGIBILITY Score

We consider the T_i top-IVE. Its Intelligibility Score $IS(T_i)$ is valued from 0.00 to 5.00 regarding the structure of the model: number of predictors, classes, rules, equations, trees, synthetic variables, modalities to predict for classifications (or numeric variables to predict for regressions or scoring). The more compact the model, the higher its IS.

The IS of each T_i is obtained by accumulating the following five penalty values to the ideal IS value of 5.00 (each penalty has a null or a negative value):

- Penalty 1 (logarithmic penalty regarding the number of predictors):
  $$\text{Pen1}(T_i) = \min(0\,, 1 - \log_{10} number\ of\ predictors)$$
  Examples: *Pen1 = 0.00 for up to 10 predictors*
  Pen1 = –3.00 for 10.000 predictors

- Penalty 2 (linear penalty regarding the average number of rules or equations per modality to predict):
  $$\text{Pen2}(T_i) = \min\left(0\,, 0.01 - \frac{average\ number\ of\ rules\ or\ equations\ per\ modality\ to\ predict}{100}\right)$$
  Examples: *Pen2 = 0.00 for 1 rule or equation per modality to predict on average*
  Pen2 = –3.00 for 301 rules or equations per modality to predict on average

- Penalty 3 (linear penalty regarding the average number of predictors per rule or equation):
  $$\text{Pen3}(T_i) = \min\left(0\,, \frac{9 - 3 \times average\ number\ of\ predictors\ per\ rule\ or\ equation}{7}\right)$$
  Examples: *Pen3 = 0.00 for up to 3.0 predictors per rule or equation on average*
  Pen3 = –3.00 for 10.0 predictors per rule or equation on average

- Penalty 4 (linear penalty regarding the number of trees per chain, here for BT only):
  $$\text{Pen4}(T_i) = \min(0\,, 1 - number\ of\ trees\ per\ chain)$$
  Examples: *Pen4 = 0.00 for 1 tree per chain*
  Pen4 = –3.00 for 4 trees per chain

- Penalty 5 (maximum penalty due to unintelligibility of synthetic variables, here for NN only):
  $$\text{Pen5}(T_i) = -5$$

Intelligibility Score of T_i

$$\boxed{IS(T_i) = \max(0.00\,, 5.00 + (Pen1+Pen2+Pen3+Pen4+Pen5))}$$

<u>Remarks:</u>

- For the difference between the Intelligibility and the Explainability of a model, please see the XTRACTIS® Brochure, page 7.
- The real complexity of the process/phenomenon under study is intrinsic, i.e., it could not be reduced or simplified, but only discovered; thus, the top-model will be complex if the process/phenomenon turns out to be complex [Zalila 2017]. Consequently, for some complex process/phenomenon, IS can be equal to 3.00 or less, even if Ti natively produces intelligible models (XTRACTIS, Random Forest).
- For similar structures, the Boosted Tree model is always less intelligible than the Random Forest one, as it is composed of chains of trees, instead of a college of trees (see Penalty 4).
- Neural Network model has always the lowest IS of 0.00, because it uses synthetic unintelligible variables (hidden nodes) in addition to all the potential predictors (see Penalty 5).

APPENDIX 2 — Use Case Results (all Performance criteria of all Top-Models)

Performance Criterion	Correlation	MAE ($\times 10^{-4}$)	**RMSE ($\times 10^{-4}$)**	Refusal
RANDOM MODEL				
Number of Random Permutations (P-value) = 100,000 (0.001)				
Performance against chance (External Test)	*0.074*	*160.0 (23.18%)*	*200.0 (28.67%)*	
XTRACTIS TOP-MODEL				
Descriptive Performance (Training)	a1.000	2.3 (0.33%)	3.7 (0.53%)	0 (0.00%)
Predictive Performance (Validation)	1.000	2.3 (0.33%)	3.7 (0.52%)	0 (0.00%)
Real Performance (External Test)	0.999	2.9 (0.41%)	**5.2 (0.75%)**	0 (0.00%)
RANDOM FOREST TOP-MODEL				
Descriptive Performance (Training)	0.998	5.9 (0.84%)	8.5 (1.22%)	
Predictive Performance (Validation)	0.996	8.4 (1.19%)	13.1 (1.87%)	
Real Performance (External Test)	0.996	8.9 (1.27%)	**14.0 (2.00%)**	
BOOSTED TREES TOP-MODEL				
Descriptive Performance (Training)	1.000	1.0 (0.14%)	1.5 (0.22%)	
Predictive Performance (Validation)	0.999	4.7 (0.67%)	6.9 (0.99%)	
Real Performance (External Test)	0.999	4.8 (0.69%)	**7.5 (1.07%)**	
NEURAL NETWORK TOP-MODEL				
Descriptive Performance (Training)	0.997	9.6 (1.37%)	12.1 (1.73%)	
Predictive Performance (Validation)	0.997	9.7 (1.39%)	12.1 (1.73%)	
Real Performance (External Test)	0.997	9.9 (1.42%)	**12.4 (1.77%)**	