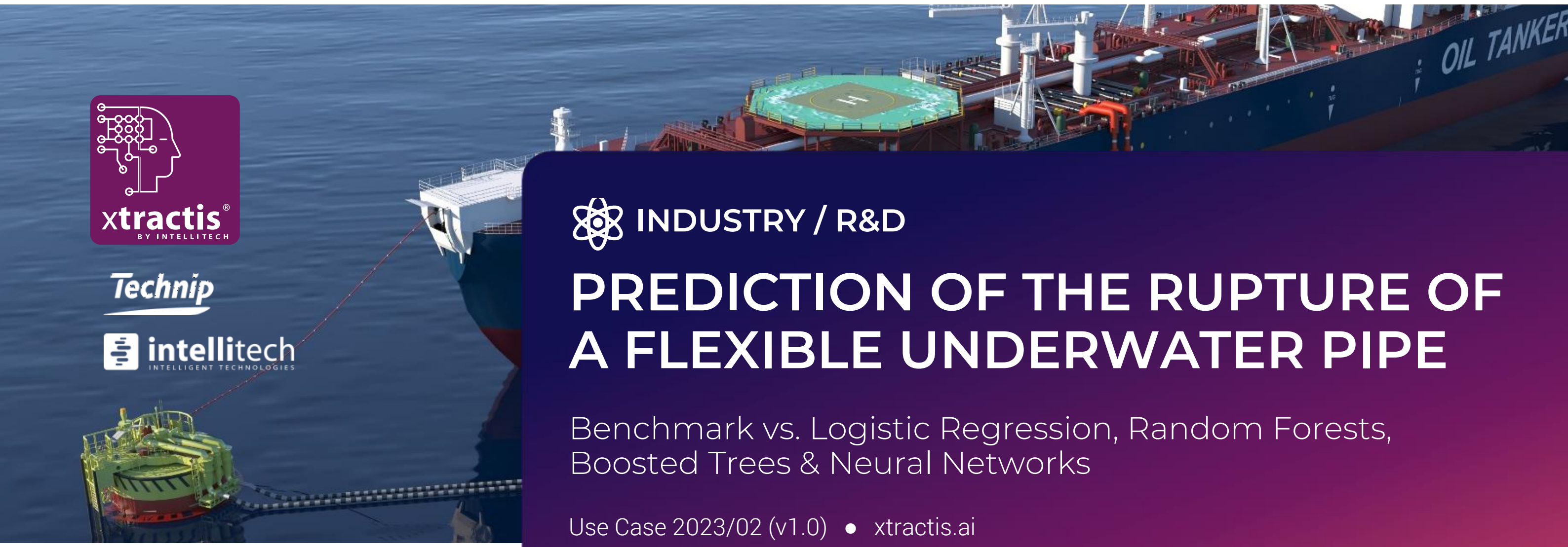



INDUSTRY / R&D

# PREDICTION OF THE RUPTURE OF A FLEXIBLE UNDERWATER PIPE

Benchmark vs. Logistic Regression, Random Forests, Boosted Trees & Neural Networks

Use Case 2023/02 (v1.0) • xtractis.ai

## PROBLEM DEFINITION

**PROBLEM**

**How to successfully predict the rupture of underwater pipes given the apparent complexity of the phenomenon?**

**GOALS & BENEFITS**

- ☑ Identify the predictors involved in the rupture of a pipe and enhance technical knowledge by helping petroleum industry engineers understand the causal relationships between these predictors, their combination, and the rupture.
- ☑ Find the truly influential parameters for assessing the state of the pipe and thus reduce measurement and maintenance costs.
- ☑ Carry out maintenance action specific for each pipe in order to avoid critical damage, thanks to rapid and transparent decision-making.

**REFERENCE DATA**

- ▶ **Observations**: 1,444 learning points from 44 experiments, based on testing pipes at various pressures. 37 experiments for Training / Validation with No-Rupture (1,201 | 72.44%), and Rupture (331 | 27.56%), 7 experiments for External Test with No-Rupture (206 | 84.77%), and Rupture (37 | 15.23%). Source: Technip.
- ▶ **Predictive Variables**: 74 Potential Predictors characterizing each pipe [composition, loading constraints, physical characteristics: number of blankets, material type, measurements …]. 66 variables are numeric, 8 are nominal.
- ▶ **Variable To Predict**: prediction of the rupture of the pipe [No-Rupture / Rupture].

**MODEL TYPE**

Regression | Multinomial Classification | **Binomial Classification** | Scoring

## XTRACTIS SOLUTION

**PROCESS**

| Reference Data | INDUCTION + Reverse-Engineering[1] | XTRACTIS Top-Model | New Cases | DEDUCTION | Automated Decision (supporting experts' diagnosis) |
|---|---|---|---|---|---|
| | | | | | No-Rupture / Rupture |

**SOFTWARE ROBOTS**

XTRACTIS® GENERATE

XTRACTIS® PREDICT + PREDICTion Report (for decision explainability)

**RESULTS**

- ☑ **Intelligible Predictive Top-Model**: Decision system composed of 27 unchained gradual rules, each rule using some of the 20 variables that XTRACTIS identified as predictors.
- ☑ **Robust Predictive Top-Model**: Good performance in External Test.
- ☑ **Operational Efficient System**: Real-time predictions up to 70,000 decisions/s., offline or online (API).


© Z. ZALILA & INTELLITECH [intelligent technologies]. 2002-2023. All Rights Reserved.

# TOP-MODEL INDUCTION

## INDUCTION PARAMETERS

We launch 300 inductive reasoning strategies; each strategy is applied to 20 different 5-fold-partitions of the Training/Validation dataset to get a reliable assessment of the descriptive and predictive performances. Each strategy thus generates 100 unitary models called **Individual Virtual Expert (IVE)**, whose decisions are aggregated with 3 possible operators into a **College of Virtual Experts (CVE)**. Among the 900 CVE, the top-CVE with the best predictive performance remains complex (56 predictors shared by 1,228 rules).

We then apply 2,000 induction strategies to the same single Training (34%)/Validation (33%)/Test (33%) partition of a synthetic dataset: 36,030 new cases simulated by deduction from the top-CVE, around the 1,201 cases but distinct from these original cases. This XTRACTIS Reverse-Engineering[1] process induces 2,000 IVE. The top-IVE selected is as efficient as the top-CVE, but intelligible (20 predictors shared by 27 rules).

| Total number of induced unitary models | Criterion for the induction optimization | Validation criterion for the top-model selection | Duration of the process (Induction Power FP64) |
|---|---|---|---|
| 32,000 IVE | $F_1$-Score | $F_1$-Score | 6 days (1 Tflops) |

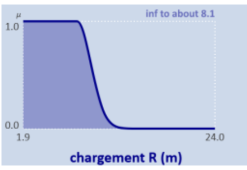
## STRUCTURE

→ Intelligibility

The top-IVE model combines the 20 predictors automatically selected by XTRACTIS into 27 rules. Its Structure Report reveals all the internal decision logic and ensures that the human expert understands the model. This decision system is a *white-box* model that can be audited by the domain expert and certified by the regulator before its deployment to end-users.

### PREDICTORS

- 20 features out of 74: 18 numeric, 2 nominal.
- Ranked by impact significance (4 strong signal, 10 medium signals, 6 weak signals):
  #1 P (bars) #2 Chargement R (m) #3 …
- Labeled by fuzzy classes, continuous or nominal.
  Example: **binary nominal** "{75 x 1,5 RILSAN}"
  **fuzzy interval** "inf to about 8.1"




### RULES

- 27 connective fuzzy rules without chaining
- 1 to 14 predictors per rule
  (on average, 5.4 predictors per rule)
- Example: **fuzzy rule R21** uses 1 nominal predictor, 3 numeric predictors, and concludes "Rupture".
  26 other fuzzy rules complete this model.

| | | | |
|---|---|---|---|
| IF | AWT sur voute | IS | {75 x 1,5 RILSAN} |
| AND | chargement R (m) | IS | inf to about 8.1 |
| AND | P/sEfl (MPa) | IS | inf to about 0.66 |
| AND | P (bars) | IS | sup to about 113 |
| THEN | Pipe State Prediction | IS | Rupture |

## PERFORMANCE

→ Robustness

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

| | Synthetic Data | | | Reference Data |
|---|---|---|---|---|
| Performance | DESCRIPTIVE | PREDICTIVE | REAL | REAL |
| Dataset | 34% Training | 33% Validation | 33% Test | External Test |
| $F_1$-Score | 99.58% | 99.58% | 99.30% | 91.18% |
| Classification Error | 0.23% | 0.23% | 0.38% | 2.47% |

**→ Xtractis Top-Model: Intelligible AND Good Predictive Capacity**


© Z. ZALILA & INTELLITECH [intelligent technologies]. 2002-2023. All Rights Reserved.

# EXPLAINED PREDICTIONS FOR 2 CASES FROM THE EXTERNAL TEST SET

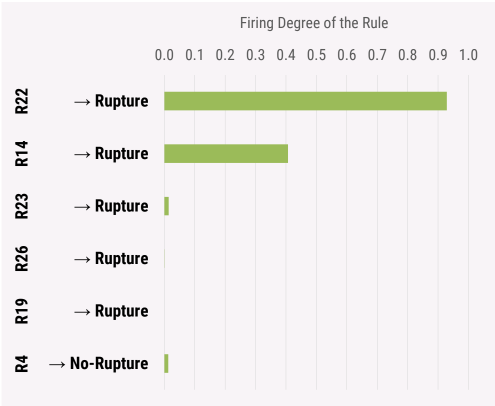
| CASE (not used in Training/Validation) | DEDUCTIVE INFERENCE OF RULES | AUTOMATED DECISION |
|---|---|---|

## CASE #29b_44 (actual value = RUPTURE)

| | |
|---|---|
| JsousA1 (mm) * | 0.000 |
| YMS (MPa) nappe 1 | 867 |
| UTS (MPa) nappe 1 | 950 |
| a (deg) nappe 2 | 30.2 |
| ... | ... |
| Mat | Fi09 |

Real Time

For this pipe, 6 rules are triggered:
**R22** is fired at 0.929, **R14** at 0.407, **R23** at 0.014, **R26** at 0.001 and **R19** at 4.79e.-5 to conclude {Rupture}.
**R4** is fired at 0.013 ton conclude {No-Rupture}.




NUMBER OF TRIGGERED RULES
6 / 27

FUZZY PREDICTION
{ Rupture | 0.929,
No-Rupture | 0.013 }

FINAL PREDICTION
{ Rupture }

The system delivers a correct diagnosis compared to that given by the technician:

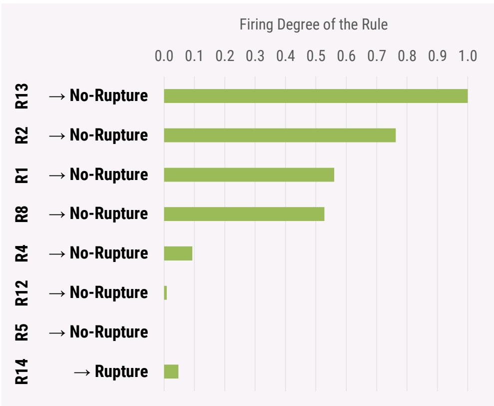
**RUPTURE**

## CASE #29b_6 (actual value = NO RUPTURE)

| | |
|---|---|
| JsousA1 (mm) * | 0.000 |
| YMS (MPa) nappe 1 | 867 |
| UTS (MPa) nappe 1 | 950 |
| a (deg) nappe 2 | 30.2 |
| ... | ... |
| Mat | Fi09 |

Real Time

For this pipe, 8 rules are triggered:
**R13** is fired at 1.000, **R2** at 0.763, **R1** at 0.560, **R8** at 0.528, **R4** at 0.093, **R12** at 0.009 and **R5** at 3.16e-04 to conclude {No-Rupture}.
R14 is fired at 0.047 to conclude {Rupture}.




NUMBER OF TRIGGERED RULES
8 / 27

FUZZY PREDICTION
{ No-Rupture | 1.000,
Rupture | 0.047 }

FINAL PREDICTION
{ No-Rupture }

The system delivers a correct diagnosis compared to that given by the technician:

**NO-RUPTURE**

**Predictor value outside the variation range of the model but inside the allowed extrapolation range. XTRACTIS will refuse to give a result for an extrapolation far from the allowed extrapolation range. It is one case of the "Refusal" prediction.*


© Z. ZALILA & INTELLITECH [intelligent technologies]. 2002-2023. All Rights Reserved.

# TOP-IVE BENCHMARK

| | XTRACTIS | LOGISTIC REGRESSION | RANDOM FOREST | BOOSTED TREES | NEURAL NETWORK |
|---|---|---|---|---|---|
| **MODELS RELEASE** | 2023/02 | 2023/01 | 2023/01 | 2023/01 | 2023/01 |
| **ALGO VERSION** | XTRACTIS **GENERATE** 13.0.45039 | Python 3.9, Scikit-Learn 1.0.2 | Python 3.9, LightGBM 3.3.2 | Python 3.9, LightGBM 3.3.2 | Python 3.9, TensorFlow 2.10.0, Keras 2.10.0 |
| **CROSS-VALIDATION TECHNIQUE** | 20×5 folds for each CVE model<br>Then 1-Split Validation for each IVE model (for the reverse engineering of top-CVE): 34% Training; 33% Validation; 33% Test | 20×5 folds for each CVE model | 20×5 folds for each CVE model | 20×5 folds for each CVE model | 20×5 folds for each CVE model |
| **NUMBER OF EXPLORED STRATEGIES**[2] | 300 induction strategies for the CVE on Training / Validation data<br>2,000 induction strategies for the IVE on synthetic data | 300 data analysis strategies on Training / Validation data | 300 ML strategies on Training / Validation data | 300 ML strategies on Training / Validation data | 300 ML strategies on Training / Validation data |
| **NUMBER OF MODELS** | 900 CVE + selection of the top-CVE<br>2,000 IVE (for the reverse engineering of top-CVE) + selection of the top-IVE | 300 CVE + selection of the top-CVE<br>1 top-IVE | 300 CVE + selection of the top-CVE<br>1 top-IVE | 300 CVE + selection of the top-CVE<br>1 top-IVE | 300 CVE + selection of the top-CVE<br>1 top-IVE |

## TOP-IVE STRUCTURE

| | XTRACTIS | LOGISTIC REGRESSION | RANDOM FOREST | BOOSTED TREES | NEURAL NETWORK |
|---|---|---|---|---|---|
| **NUMBER OF PREDICTORS** (out of 74 Potential Predictors) | **20** | **24** | **28** | **36** | **95** (6 modal variables are decomposed into 29 binary variables) |
| **DECISION STRUCTURE** | System with **27** unchained fuzzy rules | **1** linear equation with **24** coefficients. | **13** trees; **394** binary rules | **15** chained trees; **477** binary rules | **3** hidden layers; **56** hidden nodes |
| **MODEL INTELLIGIBILITY** (& DECISION EXPLAINABILITY) | Only 5.4 predictors per rule on average<br>Only a few rules are triggered at a time | A few predictors and coefficients | Many predictors and too many rules | Tree #N corrects the error of the N-1 previous trees | Unintelligible synthetic variables |

## TOP-IVE REAL PERFORMANCE (External Test)

| | *Random*[3] | XTRACTIS | LOGISTIC REGRESSION | RANDOM FOREST | BOOSTED TREES | NEURAL NETWORK |
|---|---|---|---|---|---|---|
| **Classification Error** | *19.75%* | **2.47%** | **4.94%** | **18.11%** | **13.58%** | **11.11%** |
| Sensitivity | | 83.78% | 86.49% | 78.38% | 40.54% | 83.78% |
| Specificity | | 100.00% | 96.60% | 82.52% | 94.66% | 89.81% |
| PPV | | 100.00% | 82.05% | 44.62% | 57.69% | 59.62% |
| NPV | | 97.17% | 97.55% | 95.51% | 89.86% | 96.86% |
| **$F_1$-Score** | *35.14%* | **91.18%** | **84.21%** | **56.86%** | **47.62%** | **69.66%** |
| Refusals | N/A | 0.00% | N/A | N/A | N/A | N/A |
| **MODEL ROBUSTNESS** | | **#1** | **#2** | **#4** | **#5** | **#3** |

1 Given the small number of reference cases of this dataset, the XTRACTIS Reverse-Engineering (CVE→IVE) is necessary to get a robust AND intelligible model.

2 All CVE and IVE models are optimized according to their validation $F_1$-Score. The XTRACTIS top-CVE and top-IVE are selected according to their validation $F_1$-Score while checking that it remains close to their training $F_1$-Score. The ML/LoR CVE top-models are selected according to their $F_1$-Score mean value in validation. Each ML/LoR top-IVE is obtained by applying the respective ML/LoR top-CVE strategy on 100% of the Training/Validation data.

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).

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


© Z. ZALILA & INTELLITECH [intelligent technologies]. 2002-2023. All Rights Reserved.