Maintenance

PREDICTION OF THE RUPTURE OF A FLEXIBLE UNDERWATER PIPE

Benchmark vs. Logistic Regression, Random Forest, Boosted Tree & Neural Network

2025/06 (v4.0) xtractis.ai

PROBLEM DEFINITION

GOAL

Design an AI-based decision system that accurately predicts the upcoming risk of underwater pipes rupture considering the apparent complexity of the phenomenon, to plan rational and explainable maintenance operations.

PROS & 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 risk of 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

Source: Technip, FLEXIFRANCE

Variable to Predict The model predicts the Pipe State: **NO-RUPTURE | RUPTURE**.

Predictive Variables **74 Potential Predictors characterize each pipe**: composition, loading constraints, physical characteristics: number of layers, material type, measurements …
66 variables are numeric, 8 are nominal.

Observations 1,444 reference cases from 44 experiments, based on testing pipes at various pressures.

Data are divided into a Learning Dataset for model induction using Training and Validation Datasets, and an External Test Dataset to check the top-model's performance on real data and for benchmarking.

Learning Dataset*: 1,201 | 84.1% pipes from 37 experiments 80% for Training, 20% for Validation	
NO-RUPTURE	RUPTURE
870 | 72.44%	331 | 27.56%

External Test Dataset**: 243 | 15.9% pipes from 7 experiments	
NO-RUPTURE	RUPTURE
206 | 84.77%	37 | 15.23%

**25,262 | 28.42% missing values*
***17,982 | 31.85% missing values*

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 27 gradual rules without chaining.
- Each rule uses from 1 to 14 predictors among the 20 variables that XTRACTIS automatically identified as significant (out of the 74 ones characterizing each pipe).
- Only a few rules are triggered at a time to compute the decision.

☑ **High Predictive Capacity** It has a good 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	Reference Data	**INDUCTION**	XTRACTIS Top-Model	New Cases	**DEDUCTION**	NO-RUPTURE / RUPTURE Automated Decision (predict rupture)
SOFTWARE ROBOTS	XTRACTIS® **REVEAL**			XTRACTIS® **PREDICT**		
OUTCOME	the Decision System + its Structure & Performance Reports			the decision + its Explainability Report unfolding the deductive reasoning		

TOP-MODEL INDUCTION

INDUCTION PARAMETERS & PROCESS

Powered by:

XTRACTIS® REVEAL v13.0.45039

1. We launch 300 inductive reasoning strategies. Due to the small number of reference cases, each strategy is applied to 20 different 5-fold-partitions of the Learning Dataset to get a reliable assessment of the descriptive and predictive performances, respectively from Training and Validation Datasets.
2. 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).
3. Among the 900 induced CVEs, the top-CVE with the best predictive performance remains complex: 1,228 rules share 56 predictors.

Given the small number of cases in the reference dataset, the XTRACTIS **CVE→IVE** Reverse-Engineering process is necessary to induce a unitary intelligible model through a single split cross-validation, from a large synthetic reference dataset:

4. We build a synthetic dataset composed of 36,030 new cases simulated by deduction from the top-CVE, around the 1,201 original learning cases but distinct from them.
5. We apply 2,000 induction strategies to the same single partition of this new dataset (34% Training | 33% Validation | 33% Test): XTRACTIS induces 2,000 IVEs.
6. The top-IVE selected is the one that has the best performance and with the best intelligibility, i.e., the fewer predictors and rules.

Total number of induced unitary models
32,000 IVEs

Criterion for the induction optimization
F_1-Score

Validation criterion for the top-model selection
F_1-Score

Duration of the process @ Induction Speed FP64
~6 days @ 1.13 Tflops

Environmental footprint

86.4 kWh	4.84 kg of CO_2

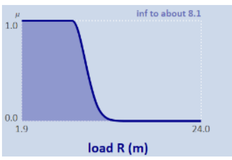
TOP-MODEL STRUCTURE

The top-IVE model has a very good intelligibility as it has **27 rules** combining **20 predictors**, with 5.4 predictors per rule on average. Its Structure Report reveals all the internal logic of the decision system and ensures that the model is understandable. It is a transparent model that can be audited by the expert and, if applicable, certified by the regulator before deployment to end-users.

PREDICTORS

- 20 features out of 74 (18 numeric, 2 nominal).
- Ranked by impact significance (3 strong signal, 11 medium signals, 6 weak signals):
  #1 P (bars) #2 load R (m) #3 …
- Labeled by nominal and fuzzy classes.
  Example:
  nominal "{75 × 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 on vault	IS	{75 × 1,5 RILSAN}
AND	load R (m)	IS	inf to ~8.1
AND	P/sEfl (MPa)	IS	inf to ~0.66
AND	P (bars)	IS	sup to ~113
THEN	Pipe State Prediction	IS	RUPTURE

Literally, the pipe is very likely to break if the AWT on vault equals 75 × 1,5 RILSAN, and the load R is under approximately 8.1 m and P/sEfl is under approximately 0.66 MPa and the Pressure is above approximately 113 bars.

TOP-MODEL PERFORMANCE

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

Perf. Type	Quality of CVE Copy				REAL
Dataset	34% Training (Synthetic Data)	33% Validation (Synthetic Data)	33% Test (Synthetic Data)	1,201 original points	External Test
F_1-Score	**99.58%**	**99.58%**	**99.30%**	**97.39%**	**91.18%**
Classification Error	0.23%	0.23%	0.38%	1.42%	2.47%

EXPLAINED PREDICTIONS FOR 2 UNKNOWN CASES

Powered by: XTRACTIS® PREDICT v13.0.45039

CASE

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

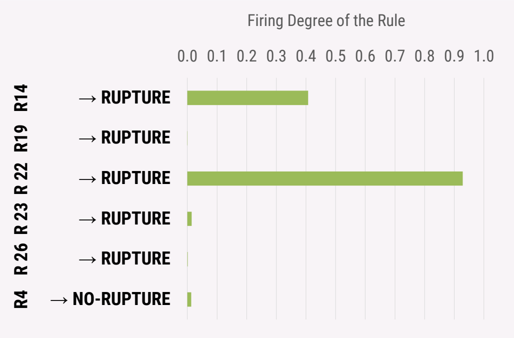
DEDUCTIVE INFERENCE OF RULES

REAL-TIME DECISION

Case #29b_44	
GunderA1 (mm)*	0.000 (2.63% OOR)
YMS (MPa) sheet 1	867
UTS (MPa) sheet 1	950
a (deg) sheet 2	30.2
a (deg) sheet 3	Missing Value
nb KV	3.00
Kv*	660.0 (2.63% OOR)
NbN*	2.00 (2.63% OOR)
N	67.0
Gapini (%)*	1.56 (4.53% OOR)
GapExp (%)	2.09
load R (m)	8.0
N1 sheet characteristics	67.0
sheet 1 IRad (mm4)	432
sheet 1 ITors (mm4)	79
N2 sheet 2 characteristics	69.0
P/sEfl (MPa)	0.79
P (bars)	250
Actual Value	RUPTURE

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 }

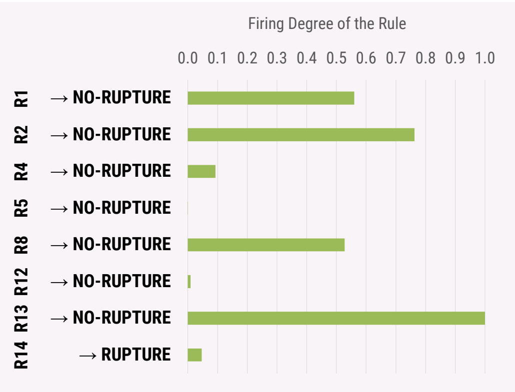
FINAL PREDICTION
{ RUPTURE }

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

RUPTURE

Case #29b_6	
GunderA1 (mm)*	0.000 (2.63% OOR)
YMS (MPa) sheet 1	867
UTS (MPa) sheet 1	950
a (deg) sheet 2	30.2
a (deg) sheet 3	Missing Value
nb KV	3.00
Kv*	660.0 (2.63% OOR)
NbN*	2.00 (2.63% OOR)
N	67.0
Gapini (%)*	1.56 (4.53% OOR)
GapExp (%)*	1.60 (1.64% OOR)
load R (m)	23.0
N1 sheet characteristics	67.0
sheet 1 IRad (mm4)	432
sheet 1 ITors (mm4)	79
N2 sheet 2 characteristics	69.0
P/sEfl (MPa)	0.79
P (bars)	20
Actual Value	NO RUPTURE

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*

TOP-MODELS BENCHMARK

MODELING PARAMETERS

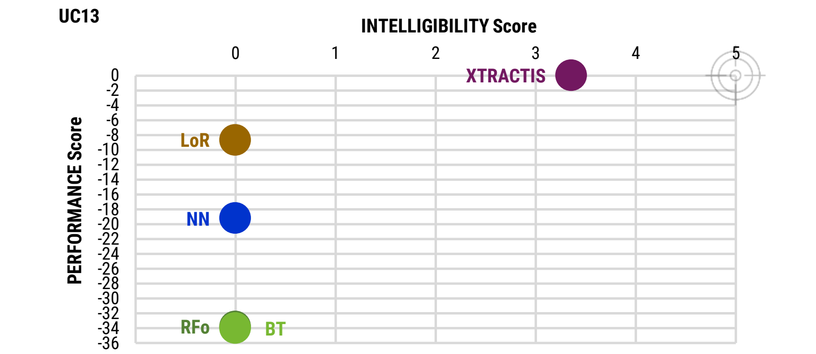
	XTRACTIS	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREE	NEURAL NETWORK
MODELS RELEASE	2023/02	2023/01	2023/01	2023/01	2023/01
ALGORITHM VERSION	XTRACTIS **REVEAL** 13.0.45039	XTRACTIS **BENCHMARK** module embedding Python 3.9.10 | Scikit-Learn 1.0.2 | LightGBM 3.3.2 | TensorFlow 2.10.0 | Keras 2.10.0			
CROSS-VALIDATION TECHNIQUE	20×5 folds for each CVE model. Then 1-Split Validation for each IVE model: 34% Training | 33% Validation | 33% Test	For LoR, RFo, BT & NN: 20×5 folds for each CVE model			
NUMBER OF EXPLORED STRATEGIES(1)	300 induction strategies for the CVE on Training / Validation data. 3 aggregation operators tested. 2,000 induction strategies for the IVE on synthetic data	300 data analysis strategies on Training / Validation data Aggregation operator: Absolute Majority	300 ML strategies on Training / Validation data. Aggregation operator: Absolute Majority		
TOP-MODEL SELECTION(2)	Top-CVE with Absolute majority aggregator among 900 CVEs, then Top-IVE among 2,000 IVEs	Top-CVE selected among 300 CVEs, then single model obtained by applying best CVE strategy on 100% of the Learning Dataset			

TOP-MODEL STRUCTURE

	XTRACTIS	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREE	NEURAL NETWORK
NUMBER OF PREDICTORS (out of 74 Potential Predictors)	**20**	**24** out of 97: 6 nominal predictors split into 29 binary variables	**28**	**36**	**95** out of 97: 6 nominal predictors split into 29 binary variables
AVERAGE NUMBER OF PREDICTORS PER RULE OR EQUATION	**5.4** per rule	**24.0** per equation	**4.6** per rule	**5.1** per rule	**61.8** per equation
STRUCTURE OF THE DECISION SYSTEM	**27** fuzzy rules without chaining (aggregated into 2 disjunctive rules) Only a few rules are triggered at a time to compute a decision	**1** linear equation	**13** trees without chaining **394** binary rules	**1** chain of **15** trees **477** binary rules Tree #N corrects the error of the N-1 previous trees	**3** hidden layers | **56** hidden nodes **57** equations 56 unintelligible synthetic variables, in addition to the 95 original predictors

TOP-MODEL SCORES

	Random(3)	XTRACTIS	LoR	RFo	BT	NN
INTELLIGIBILITY Score(4)		3.36	0.00	0.00	0.00	0.00
CVE Real Performance (F_1-Score) in External Test	*35.14*	92.86	82.50	59.62	68.42	75.95
Gap to CVE Leader in External Test	*-57,72*	**0.00**	**-10.36**	**-33.24**	**-24.44**	**-16.91**
IVE Real Performance (F_1-Score) in External Test	*35.14*	91.18	84.21	56.86	47.62	69.66
Gap to IVE Leader in External Test	*-56,04*	**0.00**	**-6.97**	**-34.32**	**-43.56**	**-21.52**
Average Real Performance in External Test	*35.14*	92.02	83.36	86.98	58.02	72.81
PERFORMANCE Score(4)		**0.00**	**-8.67**	**-33.78**	**-34.00**	**-19.22**



(1) For all algos: on exactly the same splits of the Learning Dataset. All models are optimized according to their Validation F_1-Score.
(2) All top-models are selected according to their validation F_1-Score while checking that it remains close to their Training F_1-Score.
(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

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 and 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

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

Remark:

- 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 variable or modality to predict):

  $$\text{Pen2}(T_i) = \min\left(0\,,\ \frac{1 - average\ number\ of\ rules\ or\ equations\ per\ variable\ or\ modality\ to\ predict}{40}\right)$$

  Examples: *Pen2 = 0.00 for 1 rule or equation per variable or modality to predict on average*
  Pen2 = −3.00 for 121 rules or equations per variable or 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

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

Remarks:

- 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	Classification Error	Min. Sensitivity Specificity	Sensitivity	Specificity	PPV	NPV	**F_1-Score**	Refusal
RANDOM MODEL								
Nb of Random Permutations (P-value) = 100,000 (0.001)								
Performance against chance	*19.75%*	*35.14%*					***35.14%***	
XTRACTIS TOP-MODEL								
CVE - Descriptive Performance (Training)	0.17%	99.40%	99.40%	100.00%	100.00%	100.00%	99.70%	0 (0.00%)
CVE - Predictive Performance (Validation)	0.17%	99.40%	99.40%	100.00%	100.00%	100.00%	99.70%	0 (0.00%)
CVE - Real Performance (External Test)	1.71%	89.66%	89.66%	99.51%	96.30%	98.55%	**92.86%**	9 (3.70%)
IVE - Descriptive Performance (Training)	0.23%	99.67%	99.67%	99.81%	99.49%	99.88%	99.58%	0 (0.00%)
IVE - Predictive Performance (Validation)	0.23%	99.56%	99.56%	99.85%	99.59%	99.84%	99.58%	0 (0.00%)
IVE - Real Performance (Test)	0.38%	99.22%	99.22%	99.77%	99.38%	99.71%	99.30%	0 (0.00%)
IVE - Real Performance (1,201 original points)	1.42%	95.77%	95.77%	99.66%	99.06%	98.41%	97.39%	0 (0.00%)
IVE - Real Performance (External Test)	2.47%	83.78%	83.78%	100.00%	100.00%	97.17%	**91.18%**	0 (0.00%)
LOGISTIC REGRESSION TOP-MODEL								
CVE - Descriptive Performance (Training)	2.83%	96.68%	96.68%	97.36%	93.29%	98.72%	94.96%	
CVE - Predictive Performance (Validation)	2.58%	96.68%	96.68%	97.70%	94.12%	98.72%	95.38%	
CVE - Real Performance (External Test)	5.76%	89.19%	89.19%	95.15%	76.74%	98.00%	**82.50%**	
IVE - Descriptive Performance (Training)	2.16%	96.07%	96.07%	98.51%	96.07%	98.51%	96.07%	
IVE - Real Performance (External Test)	4.94%	86.49%	86.49%	96.60%	82.05%	97.55%	**84.21%**	
RANDOM FOREST TOP-MODEL								
CVE - Descriptive Performance (Training)	0.17%	99.40%	99.40%	100.00%	100.00%	99.77%	99.70%	
CVE - Predictive Performance (Validation)	0.17%	99.40%	99.40%	100.00%	100.00%	99.77%	99.70%	
CVE - Real Performance (External Test)	17.28%	82.52%	83.78%	82.52%	46.27%	96.59%	**59.62%**	
IVE - Descriptive Performance (Training)	0.17%	99.70%	99.70%	99.89%	99.70%	99.89%	99.70%	
IVE - Real Performance (External Test)	18.11%	78.38%	78.38%	82.52%	44.62%	95.51%	**56.86%**	
BOOSTED TREE TOP-MODEL								
CVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
CVE - Predictive Performance (Validation)	0.08%	99.70%	99.70%	100.00%	100.00%	99.89%	99.85%	
CVE - Real Performance (External Test)	9.88%	70.27%	70.27%	93.70%	66.67%	94.61%	**68.42%**	
IVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
IVE - Real Performance (External Test)	13.58%	40.54%	40.54%	94.66%	57.69%	89.86%	**47.62%**	
NEURAL NETWORK TOP-MODEL								
CVE - Descriptive Performance (Training)	0.83%	98.49%	98.49%	99.43%	98.49%	99.43%	98.49%	
CVE - Predictive Performance (Validation)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
CVE - Real Performance (External Test)	7.82%	81.08%	81.08%	94.17%	71.43%	96.52%	**75.95%**	
IVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
IVE - Real Performance (External Test)	11.11%	83.78%	83.78%	89.81%	59.62%	96.86%	**69.66%**	