Naval Security

ACOUSTIC DETECTION OF UNDERWATER MINES

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

2024/02 (v4.0) xtractis.ai

PROBLEM DEFINITION

GOAL

Design an AI-based decision system that accurately and instantly detects underwater mines from sonar echoes to equip vessels, submarines, and drones with a detector making rational automated decisions.

PROS & BENEFITS

- Identify the frequency bands involved in the detection of underwater mines and enhance knowledge by helping submarine staff and acoustic experts understand the causal relationships between specific frequency bands, their combination, and the presence of a mine.
- Help to design a virtual "**Golden Ear**" (expert in underwater acoustics) operating 24/7/365 with the same quality of decision, or to design by simulation undetectable mines.
- Assist the military profession in making an earlier and more reliable decision, thanks to rapid, systematic, and explainable detection process with limited sensors.

REFERENCE DATA

Source:
Terry Sejnowski, R. Paul Gorman, University of California - San Diego, Allied-Signal Aerospace Technology Center - Columbia

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 identifies detected object: **ROCK** | **MINE**

Predictive Variables 60 Potential Predictors are measures included in [0 ; 1] characterizing the energy in a specific frequency band, integrated over a period: Energy in frequency band 1, 2, 3...,60

Observations 208 sonar echoes obtained by bouncing sonar signals off obstacles, at various angles and under various conditions.
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: 176 signals | 84.62% 80% for Training, 20% for Validation	
ROCK	MINE
82 | 46,59%	94 | 53.41 %

External Test Dataset: 32 signals | 15.38%	
ROCK	MINE
15 | 46,87%	17 | 53.13%

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 **23 gradual rules without chaining, each rule uses some of the 29 variables that XTRACTIS identified as predictors.** Moreover, 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).

☑ **Efficient AI System** — 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	ROCK / MINE: Automated Decision (detect mine)
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 v12.2.43406

1. We launch 2,000 inductive reasoning strategies; each strategy is applied to 40 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 200 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 6,000 induced CVEs, the top-CVE with the best predictive performance remains complex: 1,989 rules sharing 60 predictors.

Given the small number of reference cases in the reference dataset, the XTRACTIS **CVE→IVE** Reverse-Engineering process is necessary to get a robust AND intelligible model:

4. We build a synthetic dataset composed of 35,200 new cases simulated by deduction from the top-CVE, around the 176 original learning cases but distinct from them.
5. We apply 2,000 induction strategies to the same single 34% Training | 33% Validation | 33% Test partition of this new dataset: XTRACTIS induces 2,000 IVEs.
6. The top-IVE selected is as efficient as the top-CVE, but more intelligible: 23 rules sharing 29 predictors.

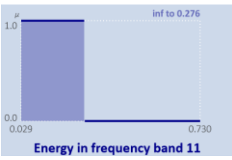
Total number of induced unitary models	Criterion for the induction optimization	Validation criterion for the top-model selection	Duration of the process (Induction Power FP64)
402,000 IVEs	F_1-Score	F_1-Score	~25 hours (1 Tflops)

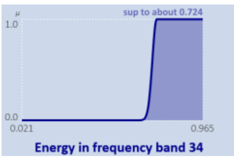
TOP-MODEL STRUCTURE

The top-IVE model has an excellent intelligibility as it has 23 rules aggregated into 2 disjunctive rules and combining the 29 predictors that XTRACTIS automatically selected out of 60 variables. The 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

- 29 energy measures in frequency bands (out of 60)
- Ranked by impact significance (1 strong, 9 medium & 19 weak signals): #1 Energy in frequency band 11, #2 Energy in frequency band 34…
- Labeled by fuzzy and binary classes
  Examples: **binary interval** "inf to 0.276";
  fuzzy interval "sup to about 0.724"





RULES

- 23 connective fuzzy rules without chaining (aggregated into 2 disjunctive fuzzy rules)
- 2 to 6 predictors per rule (on av., 3.9 predictors per rule)
- Example: **fuzzy rule R1** uses 3 predictors and concludes "ROCK". 22 other fuzzy rules complete this model, including 1 binary rule.

IF	Energy in frequency band 11	IS	inf to 0.276
AND	Energy in frequency band 34	IS	sup to ~0.724
AND	Energy in frequency band 45	IS in	~[0.088 ; 0.312]
THEN	Detected Object	IS	ROCK

Literally, the detected object is Rock if the energy in frequency band 11 is under 0.276, and in band 34 is above approximately 0.724 and in band 45 is between approximately 0.088 and 0.312.

TOP-MODEL PERFORMANCE

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	99.87%	99.86%	99.82%	93.75%
Classification Error	0.14%	0.15%	0.19%	6.45%

EXPLAINED PREDICTIONS FOR 3 UNKNOWN CASES

Powered by: XTRACTIS® PREDICT v12.2.43406

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

DEDUCTIVE INFERENCE OF RULES

AUTOMATED DECISION

ECHO #133

actual value = MINE

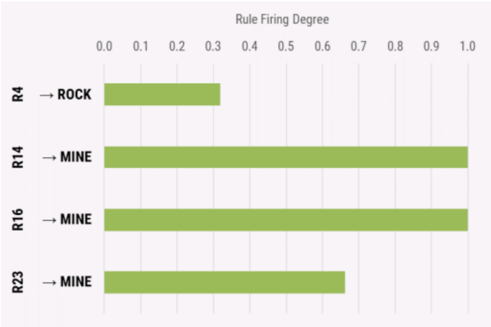
Energy in frequency band 8	0.422
Energy in frequency band 9	0.574
Energy in frequency band 11	0.349
Energy in frequency band 12	0.170
Energy in frequency band 15	0.422
Energy in frequency band 16	0.531
Energy in frequency band 19	0.191
Energy in frequency band 20	0.832
Energy in frequency band 21*	1.000
Energy in frequency band 22	0.408
…	…
Energy in frequency band 47	0.133
Energy in frequency band 49	0.106
Energy in frequency band 53	0.0081
Energy in frequency band 54	0.0303
Energy in frequency band 55	0.0190

Real Time

For this signal, 4 rules are triggered:

R14 and **R16** at 1.000, and **R23** at 0.662, to conclude {MINE}, and **R4** at 0.318 to conclude {ROCK}.

The other 19 rules are not activated.



NUMBER OF TRIGGERED RULES
4 / 23

FUZZY PREDICTION
{ MINE | 1.000,
ROCK | 0.318 }

FINAL PREDICTION
{ MINE }

The system delivers a correct detection compared to the acoustic expert / experiment:

MINE

ECHO #74

actual value = ROCK

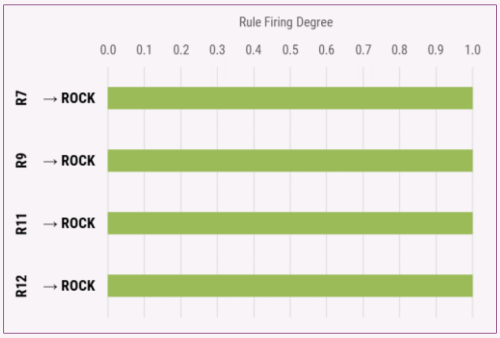
Energy in frequency band 8	0.095
Energy in frequency band 9	0.079
Energy in frequency band 11	0.126
Energy in frequency band 12	0.083
Energy in frequency band 15	0.151
Energy in frequency band 16	0.140
Energy in frequency band 19	0.299
Energy in frequency band 20	0.325
Energy in frequency band 21	0.380
Energy in frequency band 22	0.566
…	…
Energy in frequency band 47	0.090
Energy in frequency band 49	0.045
Energy in frequency band 53	0.0072
Energy in frequency band 54	0.0113
Energy in frequency band 55	0.0012

Real Time

For this signal, 4 rules are triggered:

R7, R9, R11 and **R12** at 1.000 to agree on {MINE}.

The other 19 rules are not activated.



NUMBER OF TRIGGERED RULES
4 / 23

FUZZY PREDICTION
{ ROCK | 1.000,
MINE | 0.000 }

FINAL PREDICTION
{ ROCK }

The system delivers a correct detection compared to the acoustic expert / experiment:

ROCK

ECHO #134

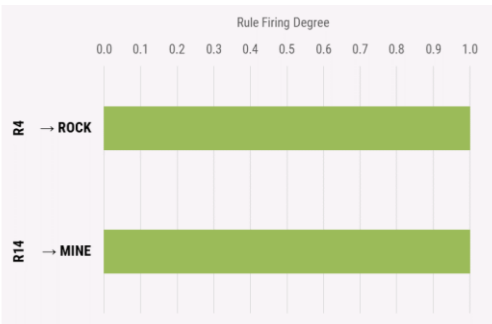
actual value = MINE

Energy in frequency band 8	0.268
Energy in frequency band 9	0.566
Energy in frequency band 11	0.5
Energy in frequency band 12	0.258
Energy in frequency band 15	0.452
Energy in frequency band 16	0.458
Energy in frequency band 19	0.533
Energy in frequency band 20	0.901
Energy in frequency band 21	0.994
Energy in frequency band 22	0.369
…	…
Energy in frequency band 47	0.077
Energy in frequency band 49	0.078
Energy in frequency band 53*	0.039
Energy in frequency band 54	0.0294
Energy in frequency band 55	0.0175

Real Time

For this situation, only 2 conflicting rules are triggered:

R4 and **R14** at 1.000.



NUMBER OF TRIGGERED RULES
2 / 23

FUZZY PREDICTION
{ ROCK | 1.000,
MINE | 1.000 }

FINAL PREDICTION
REFUSAL

The system cannot decide between the 2 classes so it refuses to decide; this is the only Refusal prediction from the External Test cases.

This Refusal could be a warning if the decision system runs in real-time.

More training data with situations near this echo profile should strengthen the model in this decision space area.

**Predictor value outside the variation range of the model (< 0.01 % OOR for case #133 and 8.92 % OOR for case #134) 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 situation of the "Refusal" prediction.*

TOP-MODELS BENCHMARK

MODELING PARAMETERS

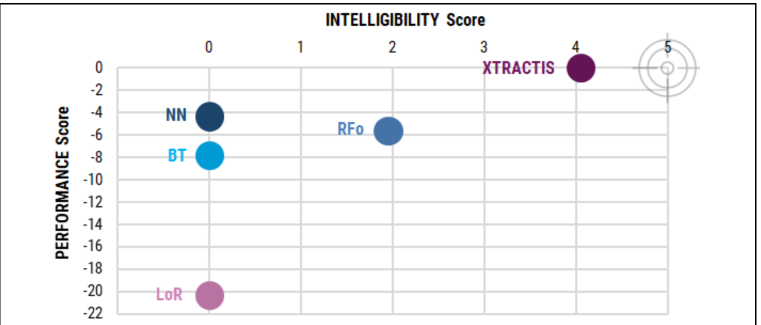
	XTRACTIS	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREES	NEURAL NETWORK
MODELS RELEASE	2022/09	2022/10	2021/12	2021/12	2022/02
ALGORITHM VERSION	XTRACTIS **REVEAL** 12.2.43406	Python 3.9.10 | Scikit-Learn 1.1.2	Python 3.6 | LightGBM 2.2.2	Python 3.6 | LightGBM 2.2.2	Python 3.6 | TensorFlow 2.6.2 | Keras 2.6.0
CROSS-VALIDATION TECHNIQUE	40×5 folds for each CVE model. Then 1-Split Validation for each IVE model: 34% Training | 33% Validation | 33% Test	40×5 folds for each CVE model	40×5 folds for each CVE model	40×5 folds for each CVE model	40×5 folds for each CVE model
NUMBER OF EXPLORED STRATEGIES(1)	2,000 induction strategies for the CVE on Training / Validation data. 2,000 induction strategies for the IVE on synthetic data	2,000 data analysis strategies on Training / Validation data	800 ML strategies on Training / Validation data	800 ML strategies on Training / Validation data	2,000 ML strategies on Training / Validation data
TOP-MODEL SELECTION(2)	Top-CVE among 6,000 CVEs, then Top-IVE among 2,000 IVEs	Top-CVE selected among 2,000 CVEs, then single model obtained by applying best CVE strategy on 100% of the Learning Dataset	Top-CVE selected among 800 CVEs, then single model obtained by applying best CVE strategy on 100% of the Learning Dataset		Top-CVE selected among 2,000 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 TREES	NEURAL NETWORK
NUMBER OF PREDICTORS (out of 60 Potential Predictors)	**29**	**51**	**55**	**58**	**60**
AVERAGE NUMBER OF PREDICTORS PER RULE / EQUATION	**3.9** per rule	**51.0** per equation	**4.1** per rule	**3.8** per rule	**48.2** per equation
STRUCTURE OF THE DECISION SYSTEM	**23** 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	**27** trees without chaining **369** binary rules	**1** chain of **48** trees **503** binary rules Tree #N corrects the error of the N-1 previous trees	**3** hidden layers | **74** hidden nodes **75** equations 74 unintelligible synthetic variables

INTELLIGIBILITY × PERFORMANCE SCORES (Performance Score is calculated on all available unknown data)

	Random(3)	XTRACTIS	LoR	RFo	BT	NN
INTELLIGIBILITY Score(4)		4.05	0.00	1.95	0.00	0.00
CVE Real Performance (F_1-Score) in External Test		91.43	72.22	88.24	88.24	88.24
Gap to CVE Leader in External Test		**0.00**	**-19.21**	**-3.19**	**-3.19**	**-3.19**
IVE Real Performance (F_1-Score) in External Test	*55.17%*	93.75	72.22	85.72	81.25	88.24
Gap to IVE Leader in Test		**0.00**	**-21.53**	**-8.03**	**-12.50**	**-5.51**
Top-IVE Average Real Performance	*55.17%*	92.59	72.22	86.98	84.75	88.24
PERFORMANCE Score(4)		**0.00**	**-20.37**	**-5.61**	**-7.85**	**-4.35**



(1) For all algos: on the same 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.

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

APPENDIX 1 — Calculation of the Intelligibility × Performance

AI Technique #i	T_i	i∈[1 ; n] n = number of AI Techniques benchmarked in terms of data-driven modeling = 5
Benchmark #k	B_k	k∈[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 & Boost Trees: 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 target is to obtain 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

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

<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):

  $Pen1(T_i) = \min(0\,, 1 - \log_{10} number\ of\ predictors)$

  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):

  $Pen2(T_i) = \min\left(0\,, 0.01 - \frac{average\ number\ of\ rules\ or\ equations\ per\ modality\ to\ predict}{100}\right)$

  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):

  $Pen3(T_i) = \min\left(0\,, \frac{9 - 3 \times average\ number\ of\ predictors\ per\ rule\ or\ equation}{7}\right)$

  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 chained trees, here for BT only):

  $Pen4(T_i) = \min(0\,, 1 - number\ of\ chained\ trees)$

  Examples: Pen4 = 0.00 for 1 tree Pen4 = –3.00 for 4 chained trees

- Penalty 5 (maximum penalty due to unintelligibility of synthetic variables, here for NN only):

  $Pen5(T_i) = -5$

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 Forests).
- For similar structures, the Boosted Trees model is always less intelligible than the Random Forest one, as it is composed of chains of trees, instead of a college of trees (cf. 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 (cf. 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	*47.90%*	*48.58%*					***55.17%***	
XTRACTIS TOP-MODEL								
CVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0 (0.00%)
CVE - Predictive Performance (Validation)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0 (0.00%)
CVE - Real Performance (External Test)	9.38%	86.67%	94.12%	86.67%	88.89%	92.86%	91.43%	0 (0.00%)
IVE - Descriptive Performance (Training)	0.14%	99.82%	99.91%	99.82%	99.84%	99.89%	99.87%	141 (1.18%)
IVE - Predictive Performance (Validation)	0.15%	99.84%	99.84%	99.87%	99.89%	99.81%	99.86%	139 (1.20%)
IVE - Real Performance (Test)	0.19%	99.77%	99.77%	99.85%	99.87%	99.74%	99.82%	152 (1.31%)
IVE - Real Performance (287 original points)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	2 (1.14%)
IVE - Real Performance (External Test)	6.45%	93.33%	93.75%	93.33%	93.75%	93.33%	93.75%	1 (3.13%)
LOGISTIC REGRESSION TOP-MODEL								
CVE - Descriptive Performance (Training)	14.20%	79.27%	91.49%	79.27%	83.50%	89.04%	87.31%	
CVE - Predictive Performance (Validation)	13.07%	79.27%	93.62%	79.27%	83.81%	91.55%	88.44%	
CVE - Real Performance (External Test)	31.25%	60.00%	76.47%	60.00%	68.42%	69.23%	72.22%	
IVE - Descriptive Performance (Training)	16.64%	79.27%	92.55%	79.27%	83.65%	90.28%	87.88%	
IVE - Real Performance (External Test)	31.25%	60.00%	76.47%	60.00%	68.42%	69.23%	72.22%	
RANDOM FOREST TOP-MODEL								
CVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
CVE - Predictive Performance (Validation)	12.50%	79.27%	94.68%	79.27%	83.96%	92.86%	89.00%	
CVE - Real Performance (External Test)	12.50%	86.67%	88.24%	86.67%	88.24%	86.67%	88.24%	
IVE - Descriptive Performance (Training)	0.57%	98.78%	100.00%	98.78%	98.95%	100.00%	99.47%	
IVE - Real Performance (External Test)	15.63%	80.00%	88.24%	80.00%	83.33%	85.71%	85.72%	
BOOSTED TREES TOP-MODEL								
CVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
CVE - Predictive Performance (Validation)	5.11%	93.90%	95.74%	93.90%	84.74%	95.06%	95.24%	
CVE - Real Performance (External Test)	12.50%	86.67%	88.24%	86.67%	88.24%	86.67%	88.24%	
IVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
IVE - Real Performance (External Test)	18.75%	76.47%	76.47%	86.67%	86.67%	76.47%	81.25%	
NEURAL NETWORK TOP-MODEL								
CVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
CVE - Predictive Performance (Validation)	6.25%	90.24%	96.81%	90.24%	91.92%	96.10%	94.30%	
CVE - Real Performance (External Test)	12.50%	86.67%	88.24%	86.67%	88.24%	86.67%	88.24%	
IVE - Descriptive Performance (Training)	1.14%	97.56%	100.00%	97.56%	97.92%	100.00%	98.95%	
IVE - Real Performance (External Test)	12.50%	86.67%	88.24%	86.67%	88.24%	86.67%	88.24%	